The Validity of Retell Coding Procedures in Elementary School Students with Dyslexia

Senior Honors Thesis

Rachel Waters

December 2016

Honors Thesis Committee Members:

Beth Meisinger, Ph.D. (Chair)

Randy Floyd, Ph.D.

Abstract

This study examined the psychometric properties of two common frameworks for scoring retell data, those based on clauses, and those based on idea units, using a longitudinal sample of 36 third- and fourth-grade students with dyslexia. Results from the clause-based and idea unit-based scoring frameworks were compared to two standardized, norm-referenced measures of reading comprehension (i.e., the GORT-5 and IOWA Assessments). The two scoring methods were found to correlate robustly to one another and both were sensitive to growth across the year. Both scoring methods of the retell measures also moderately correlated to the criterion measures of reading comprehension. However, retell may not accurately classify struggling readers. Lastly, trends in the retell data across reading modality (oral versus silent reading) suggest that additional research is warranted.

Table of Contents

Introduction

Reading comprehension refers to the construction of meaning from text (Sweet & Snow, 2003) and is the most essential outcome of the reading process. However, because reading comprehension is a construct and not a directly observable phenomenon, the degree to which measurement of reading comprehension accurately captures this latent construct varies across measures (Keenan, Betjemann, & Olson, 2008). Given the importance of reading comprehension for subsequent academic and occupational success (Hakkarainen, Holopainen, & Savolainen, 2013), the degree to which measures of reading comprehension accurately capture this construct remains an important question. Multiple approaches to measuring reading comprehension, including retell measures, have been developed for use in the classroom.

Retell is a common approach to assessing reading comprehension where students read a passage orally or silently and then are asked to tell the examiner everything they can about what was read without looking back at the passage. Extant research suggests that retell scores correlate moderately to highly (.52-.82) with other reading comprehension measures (Reed & Vaughn, 2012). A variety of different protocols exist for scoring retells, but many of these approaches lack appropriate evidence regarding their psychometric properties (Reed, 2011). Even less research has investigated the validity of retell as an assessment of reading comprehension in populations with specific learning disabilities (SLD) in reading, including dyslexia.

Dyslexia refers to students who experience difficulties with accurate or fluent word recognition and have poor decoding skills, dysfluent text reading, and poor spelling abilities (American Psychiatric Association, 2013; International Dyslexia Association, 2002). It is widely accepted that difficulties with phonological processing represent a core deficit for students with dyslexia (Melby-Lervåg, Lyster, & Hulme, 2012). Reading comprehension problems in this population may be viewed as a secondary consequence of word-level difficulties and reduced

reading experience (International Dyslexia Association, 2002). Deficits in phonological processing may impede the development of fast and accurate word reading skills resulting in slow and dysfluent reading. Cognitive and attentional resources that would otherwise be available for comprehension processes are devoted to these lower level tasks (LaBerge & Samuels, 1974). Further, struggling readers may also become discouraged and tend to spend less time engaged in reading, which may impede growth in vocabulary and background knowledge and thereby contribute to subsequent comprehension difficulties (Stanovich, 1986).

Understanding how dyslexia affects reading comprehension is an important topic in reading research. Given the evidence base for linguistic deficits within a dyslexic population (Shaywitz, 2003), it could be argued that an assessment of comprehension that relies heavily on the production of speech, as required by story retell tasks, is a less than ideal approach. Individuals with dyslexia not only have trouble decoding words when reading, but may experience difficulty with word retrieval during speech production (Shaywitz, 2003). For example, a word with a similar sound but dissimilar meaning may be substituted (i.e., recession for reception) when speaking. Given the dearth of research regarding retell procedures used with dyslexic students, it is not clear if this concern is merited; however, the current research will investigate whether retell is a valid tool for assessing reading comprehension when used with dyslexic elementary school students.

Retell Scoring Frameworks

Two frameworks for scoring retell data are commonly used in the literature, those based on clauses and those based on idea units. A clause is defined as "any verb and the elements that 'cluster' with it" (Gee, 1999, p. 99), whereas an idea unit is the smallest unit of information with meaning associated with the passage. For example, the sentence "Although the bunny was cute

and furry, it was very mean." has two clauses: "Although the bunny was cute and furry" and "it was very mean." However, the sentence has three possible idea units: "the bunny was cute," "the bunny was furry," and "the bunny was mean." Scoring retell on a clausal basis allows readers to make inferences and exhibit their knowledge of the text with as many phrases as necessary. In this way, the participant is scored not only on content of the retell, but also in their ability to accurately construct phrases and sentences. This method may be problematic in a dyslexic population due to the linguistic deficits in producing speech, as discussed earlier. The other scoring framework, idea unit-based, focuses solely on the content of the retell without regards to grammatical aspects of speech. This method is more accommodating for participants with language deficits, but may not accurately represent comprehension. For instance, a student who retells random words or phrases verbatim from the text may have a high retell score, but only because of their ability to remember such words or phrases without having to put meaning to them.

Some researchers argue that retell data should only be scored in terms of clauses (Kucer, 2010; Merritt & Liles, 1987). In one such study, 34 fourth-graders were asked to read the first chapter of a narrative book and then retell what they read (Kucer, 2010). Researchers then queried the students about certain aspects of the retell without introducing new topics to the student. The number of spoken clauses were then counted and matched to the story for accuracy. Notably, results suggested that retell scored by clauses might not accurately represent comprehension of the passage, due to the considerable gap between content of retell and understanding of the text as measured by other comprehension assessments. In another study, Merritt and Liles (1987) analyzed narratives told by 40 fourth-grade individuals, of whom 20 were diagnosed with language disabilities. The data were scored by number of main and

subordinate clauses, as well as basic grammar components. Results showed that participants with language disabilities had lower means of clauses retold and lower levels of comprehension of the text, showing comparability between the measure and the construct, as compared to students without a language disability. The results of both of these studies are important to consider in light of the current study in order to examine the difference in retell measures when dealing with a dyslexic population.

In contrast, other studies have scored retell data based on percentage of idea units, the most basic form of information, including the subject, verb, and a direct object, retold from a passage (Marr & Gormley, 1982; McNamara, Ozuru, & Floyd, 2011; Roberts, Good, & Corcoran, 2005). Using this methodology, retell data were scored based on the percentage of non-repetitive propositions, or idea units, retold. Roberts, Good, and Corcoran (2005) presented retell as an accurate tool for teachers to differentiate students with reading disabilities using this approach. In the study, 86 first-grade students read texts developed by researchers and the Letter-Word Identification, Word Attack, and Passage Comprehension subtests from the Woodcock Diagnostic Reading Battery subtests (WDRB; Woodcock & Johnson, 1997), along with other oral reading probes. Results suggested that retell data, when scored on an idea unit basis, accurately predicted reading fluency and may lead to a better understanding of the gap between reading fluency and comprehension.

In a similar study, McNamara et al. (2011) investigated both cued and free retell (cued retells having prompts or questions related to the story, free retells involving a general telling about the story following a generic prompt) and its relationship to reading skills, including reading comprehension, in a sample of 65 fourth-grade students. In this study, students read eight texts (a mixture of expository and narrative) of varying text-cohesion. Retell data were scored

based on idea units, or propositions, and then compared to multiple-choice questions associated with the same passages. Results of the study suggested that when retelling, students were more likely to make inferences about the text rather than retell solely the information within the text, thus creating a better understanding of the text as a whole. However, McNamara et al. (2011) found that the type of text (high or low cohesion, narrative or expository) also influenced students' performance on reading comprehension assessments.

It is important to consider these results when developing the framework of the current study dealing with a dyslexic population. Previous research suggests that there is a significant difference in reading comprehension of expository and narrative passages (Dickens & Meisinger, 2015; Best, Floyd, & McNamara, 2008), and that retellings of expository texts are generally shorter and less correct than narrative texts (Kucer, 2010). Expository texts are particularly difficult to both read and understand for students with learning disabilities (Reed & Vaughn, 2012). In light of these findings, the present study utilized only narrative texts that were administered in a counterbalanced manner for modality to control for text difficulty level.

Marr and Gormley (1982) examined reading comprehension of both familiar and unfamiliar texts based on free and probed retell in a sample of 33 fourth-graders. To assess reading comprehension, retell data were scored based on propositions (i.e., idea units) retold and scores were then compared to multiple-choice questions about the same passage. Results suggested that retell measures accurately differentiated between familiar or unfamiliar texts, thus demonstrating a relationship to reading comprehension. Marr and Gormley (1982) also found that once a participant was probed rather than asked for a general free recall, the individual tended to revert to prior knowledge rather than reflect on what was read in the text. This finding highlights the importance of prior knowledge in supporting text comprehension, and suggests

that free retell may be a better measure of comprehension for particularly difficult texts, as the individual is forced to reflect specifically on the information within the text rather than previous knowledge or amount of phrases stated.

Clause-based and idea unit-based methods of scoring retell data each have their benefits, but both methods also their drawbacks. Although scoring retell data based on idea units is more straightforward than the clause approach, it places a greater emphasis on literal rather than amount of information. Information explicitly stated in the text may be scored as correct, but the student's knowledge of the meaning of the text is not necessarily exhibited. Although the clause-based approach requires the student to make complete phrases or sentences to show knowledge, this may not be an appropriate method for students with language disabilities. Given the use of equivocal methods throughout the literature, the data generated by these rather different approaches need to be compared to criterion measures of reading comprehension with established psychometric properties in order to understand which may be more representative of knowledge as well as appropriate for students who may have language difficulties.

Current Study

The purpose of this study was to examine the two most commonly used methods of retell coding, clause-based and idea unit-based scoring, to determine which approach yielded the most robust measure of reading comprehension in a dyslexic population of elementary school students. Based on the current literature, it is not clear which approach for scoring retell data yields the best information regarding comprehension for typical students (Reed, 2011), and virtually no research has been conducted who have students with specific learning disabilities in reading, including dyslexia. To this end, the concurrent and predictive relations for the two retell scoring techniques were examined by comparing their scores to one another and to other

measures of reading comprehension including the Comprehension subtest from the Gray Oral Reading Test, Fifth Edition (GORT-5; Wiederholt & Bryant, 2012) and the Reading subtest from the IOWA Assessments, form E (Dunbar & Welch, 2011). The following research questions guided this study:

1. How closely related are the retell scores generated from the clause-based and idea unit-based protocols to one another when measured concurrently and across time (fall to spring)?
2. How closely do the retell scores generated by the clause-based and idea unit-based protocols correlate with the criterion measure of reading comprehension (i.e., the GORT-5 Comprehension subtest and Iowa Assessments Reading subtest)?
3. Are the retell scores sensitive to growth across the school year (fall to spring)?
4. What is the classification accuracy of the retell data in identifying students struggling with regard to reading comprehension?

Method

Participants

This study utilized a subgroup of students participating in a larger longitudinal study examining the development of reading fluency in students with dyslexia. Participants were third- and fourth-grade (N = 36) students enrolled in a private school in the Southeastern United States that provides an intensive day-treatment program for students with dyslexia. All students were diagnosed with a specific learning disorder in reading using the criteria outlined in the Diagnostic and Statistical Manual, Fifth Edition (DSM-5; American Psychiatric Association,

2013) prior to admission to the school. Students were approximately 89% European American, 6% Asian American, 3% African American, and 2% other; 55.6% of the students were male.

Measures

Oral and silent retell. Two grade-level narrative AIMSWeb benchmark assessment passages were selected for use in this study (Shinn & Shinn, 2002), and then were shortened to prevent fatigue effects. The third-grade passages were 137 and 141 words in length, and the fourth-grade passages were 170 and 179 words in length. To ensure that the newly shortened passages were equal across measures, readability was calculated using an average of scores from the Dale-Chall, Spache, and Flesch-Kincaid (F-K grade level) formulas. The average readability across the three formulas was 3.73 and 4.06 for the third-grade passages, and 4.23 and 4.53 for the fourth-grade passages. Students read one grade-level passage aloud and another silently using the following directions:

Silent Reading Directions

> I am going to give you a reading passage. When I say begin, I want you to read the passage silently, or in your head, as carefully and as quickly as you can. If you come to a word that you don't know, do your best and keep going. Only read the passage through once. When you have finished reading the passage, look up at me, and say, "Done." I will take up the passage and ask you questions about what you just read. I cannot give you any hints or help. Do you have any questions? Ok, here is the passage (*examiner places the passage in front of the student*). Remember to read the passage silently. Begin reading here (*examiner points to the first word of the passage and start the timer*).

Oral Reading Directions

> I am going to give you a reading passage. When I say begin, I want you to read the passage out loud as carefully and as quickly as you can. If you come to a word that you don't know, do your best and keep going. When you have finished reading, I will take up the passage and ask you questions about what you just read. I cannot give you any hints or help. Do you have any questions? Ok, here is the passage (*examiner places the passage in front of the student*). Remember to read the passage *out loud.* Begin reading here (*examiner points to the first word of the passage*). *If the child begins reading silently, provide prompt*, "Remember to read out loud," *and restart timer.*

After removing the passage following the oral or silent reading, examiners read the following prompt taken from the DIBELS 6th ed. (Good & Kaminski, 2002, p. 34), "Please tell me about what you just read. Try to tell me everything you can. Begin." Students had 1 minute to retell the passage. If students paused for 3 seconds they were prompted to "Try to tell me everything you can." The retell was ended if a subsequent pause exceeded 5 seconds.

Audio recordings of the retells were used to transcribe each student's response for scoring purposes. This procedure is in accordance with studies demonstrating that scoring in-lab results were more accurate in comparison to real-time scoring (Bellinger & DuPerna, 2011). Each individual's response was then coded for both clauses and non-repeated idea units for comparison to the original text for correctness, summarization, and inferential reasoning. Similar to previous research, retell data were scored on a scale from 0-1 based on correctness: 0 for information not recalled, 0.5 for information either insufficient or partially incorrect, and 1 for a complete, correct recall (McNamara, Ozuru, & Floyd, 2011). The percentage of idea units and

clauses (number of idea units or clauses in retell out of number of idea units in original text that could potentially be recalled) were also recorded for each participant.

The four passages used in the study were independently reviewed by two grammar experts, both professors at a southeastern university with Master's degrees in English, in order to accurately separate the texts by clauses and to ensure similar clause count in each text. Researchers then coded the retell transcriptions and compared retells to the original text model given by these grammar experts. To score data for idea units, the main ideas, words, or phrases in each original text were identified by researchers using an example from the Qualitative Reading Inventory, Fifth Edition (QRI-5; Leslie & Caldwell, 2010). Using this procedure, an idea unit was identified as any information pertinent to the story. In accordance with the QRI-5, researchers marked every mention of an idea unit in student transcriptions, but no points were given to repetitive idea units or phrases unrelated to the passage (e.g., "That's all I remember" Or "I didn't like that story"). Once a retell transcription was scored, the percentage correct was calculated by dividing the number of points received by the number of possible points (total number of clauses or idea units in the passage). The percentage correct was used in analyses in order to account for the varying total numbers of clauses and idea units per passage. A sample retell scoring sheet used to score the clause and idea unit information is provided in Appendix A.

Oral reading comprehension. The Gray Oral Reading Test, Fifth Edition (GORT-5; Wiederholt & Bryant, 2012) is a test of oral reading fluency and reading comprehension consisting of a series of successive passages that increase in difficulty level. Passages are read aloud, and the reading time and number of reading errors are recorded. Students are then asked five open-ended comprehension questions tied to the passage. The GORT-5 yields scaled scores (M = 10, SD = 3) for the Accuracy, Rate, Fluency, and Comprehension subtests. Test-retest

reliability estimates range from .82 to .90, internal consistency coefficients range from .91 to .97, and alternate form reliability coefficients range from .91 to .96 for 7- to 9-year-old students. Validity estimates with other measures of reading achievement had a median correlation of .72 (Widerholt & Bryant, 2012).

Silent reading comprehension. The Reading subtest of the IOWA Assessments, form E (Dunbar, Welch, Hoover, & Frisbie 2011a, 2011b) was used to assess silent reading comprehension. Level 9 and 10 were administered to the third- and fourth-graders, respectively. Students silently read a series of fiction and nonfiction passages and answered multiple-choice questions about each passage within a 30-minute time period. Questions are designed to address key ideas of the passage, implicit and explicit meaning, author's craft, and vocabulary. The Reading subtest provides standard scores, means, and standard deviations that vary based on the grade level and the test form that is given. For form E, the manual reported mean standard score for the Reading subtest was 187.8 (*SD* = 24.5) for third-grade and 202.6 (*SD*=28.7) for fourth-grade. Due to the difference in means and standard deviations across grade levels, the z-score of each value of the present sample was calculated in order to collapse grade levels for subsequent analyses. The test manual reported internal consistency reliability estimates exceeding .90 for Reading subtest (levels 9 & 10) and strong correlations with other measures of reading achievement (.75 to .85; Dunbar, Welch, Hoover, & Frisbie, 2014).

Procedure

Following the approval from the University of Memphis Institutional Review Board, data were collected in August 2015 and May 2016. Written parental consent and student assent were required for participation in the study. Examiners obtained a minimum of 95% agreement on practice test administration prior to data collection. Sessions were audio recorded to allow for

subsequent review and scoring. The IOWA Assessments Reading subtest was administered by school personnel in the spring only, a week prior to other assessments, and scores were collected from school records. The two grade-level AIMSweb probes and the GORT-5 were administered by researchers at each time point (fall and spring). With regard to the GORT-5, form A was administered in the fall and form B in the spring. The administration of the two grade-level AIMSWeb probes were counterbalanced across reading modality for the fall and spring time points, such that each passage was read once aloud and once silently. The GORT-5 and AIMSWeb passages were individually administered in a counterbalanced order in a quiet location at the school. Using audio recordings of testing sessions, 20% of the GORT-5 data were randomly selected for blind review by an independent reviewer. Discrepancies were rare (< 1%) and were resolved via discussion..

Two raters, the first author and a graduate student in School Psychology, used audio recordings to transcribe each passage retell. Each rater transcribed half of the retells, and all transcriptions were reviewed by the other rater for accuracy. There were no discrepancies within the transcriptions, resulting in 100% agreement between raters. The retell transcriptions were subsequently used for both coding schemes (clause-based and idea unit-based scoring methods). Following a preliminary review of the transcription data, it was obvious that a scoring manual, in addition to the original scoring rubric, was necessary to establish clear guidelines in the case of ambiguous retells for scoring the retell responses. Fifty percent of transcriptions were then randomly selected (Research Randomizer, n.d.) to be independently scored by both raters using the coding manual. The percent agreement was calculated by number of agreements between the raters divided by the total number of scores resulting in 91.4% agreement (within 1-scoring unit) for clause-based scoring as well as idea unit-based scoring. The remaining transcriptions were

then jointly scored by both raters. Sample excerpts from the retell coding manual used to score the clause and idea unit information are provided in Appendix B.

Results

Data Screening and Processing

Prior to analyses, data were screened for missing data points, out of range or unexpected values, outliers, and normality. No data points were identified as outliers (z-scores < 3.29; Tabachnik & Fidell, 2012) and skewness and kurtosis statistics also fell within acceptable limits (values < |2.0|; Tabachnik & Fidell, 2012). Scores of zero on any retell measure were reviewed to determine whether these scores likely represented poor comprehension rather than "fake reading" behavior (i.e., not reading the passage). A retell score of zero was counted as a missing data point if any of the following criteria were met: a) the examiner noted on the student's file that student clearly was not reading (i.e., looking around the room instead of at the passage), b) the student had an implausible reading time compared to other documented reading rates, or c) student gave a response unrelated to the story (i.e., student made up a story noticeably different to the one read or student claimed to not remember any information about the passage). If it was evident that the student tried to read the passage but did not understand enough to produce a score-able retell, then the data were kept as an accurate representation of the student's reading comprehension. Using these criteria, four data points across the 36 students were dropped and then estimated in SPSS version 21.0 for windows using missing value analysis (estimation maximization algorithm).

Descriptive Statistics and Correlations

Mean and standard deviations for all variables are reported in Table 2. As would be expected for a sample of students identified as having a reading disability, the average scores on

both standardized, norm-referenced measures of reading skill fell in the below average range. Mean scores for retell coding methods are based on the percentage of correct scores for each scoring method. Strength of correlations was determined as follows: .00-.19 negligible, .20-.39 weak, .40-.69 moderate, .70-.89 strong, and .90-1.0 very strong (Floyd, Clark, & Shaddish, 2008).

The first research question addressed the strength of the concurrent and predictive relations between retell scores generated from the clause-based and idea unit-based protocols. To this end, Pearson product-moment correlations were first conducted between the two retell methods for oral and silent reading for each time point separately (see Table 1). In the fall, very strong concurrent correlations were found between clause and idea unit scores whether passages were read aloud, r =.955, p <.01, or silently, r = .982, p < .01. Similarly, very strong concurrent correlations were found between these two methods in the spring for passages read aloud, r = .936, p < .01, and strong concurrent correlations were observed when passages were read silently, r = .890, p < .01. In sum, strong concurrent relations were observed between the clause-based and idea unit-based scoring methods across both time points and regardless of reading modality.

Next, the predictive correlations (from fall to spring) across the retell scoring methods were examined for each modality separately. Correlations between the fall idea unit scores and the spring idea unit scores, r = .342, p < .05, and spring clause scores, r = .346, p < .05, fell in the weak range when the student read orally. When reading silently, moderate predictive correlations were found for fall idea unit scores and spring idea unit scores, r = .439, p < .01, and weak correlations were found for the fall idea unit scores and spring clause scores, r = .360, p < .01. The predictive correlations between fall clause scores and spring clause scores, r = .248, p > .05,

and spring idea unit scores, $r = .228$, $p > .05$, were not statistically significant when the student read orally. When the student read silently, weak to moderate predictive correlations between fall clause scores and spring clause scores, $r = .372$, $p < .05$, and spring idea unit scores, $r = .449$, $p < .01$. In sum, although strong concurrent relations were found for the two scoring methods at each time point, the temporal stability of these relations generally fell in the weak to moderate range.

Table 1
Correlations for Clause-Based and Idea Unit-Based Retell Scores

Assessment	1	2	3	4
1. Fall Idea Unit	--	.955**	.342*	.346*
2. Fall Clause	.982**	--	.228	.248
3. Spring Idea Unit	.439**	.449**	--	.936**
4. Spring Clause	.360**	.372*	.890**	--

Note. *p<.05; **p<.01; Oral scores are reported above the diagonal and silent scores are reported below the diagonal

The second research question addressed the association between the clause and idea unit retell scores and the criterion measure of reading comprehension. To this end, Pearson product-moment correlations for each of the retell scoring methods and other measures of reading comprehension (i.e., the GORT-5 Comprehension subtest and IOWA Assessments Reading subtest) were examined (see Table 3). A restricted range was evident for the GORT-5 Comprehension subtest scores for both the fall and spring ($M = 6.53$, $SD = 1.58$ and $M = 7.56$, $SD = 1.96$, respectively, compared to a standard $M = 10$, $SD = 3$), which is not surprising given that the sample consisted of students identified as having reading disability. Therefore, it was

necessary to correct the correlations between the GORT-5 Comprehension subtest and retell data for this restricted range with a formula utilizing the known population parameters (P. Cohen, J. Cohen, West, & Aiken, 2003; Gulliksen, 1987). Because standard scores vary across grade levels for the IOWA Assessments Reading subtest, these correlations between the retell scores were not corrected although a restricted range was found. The corrected and uncorrected values are reported in Table 3, but only the corrected correlations were interpreted.

Table 2
Descriptive Statistics for Retell Scores and Criterion Measures of Reading Comprehension

	Mean	Standard Deviation
Fall Oral Idea Unit	17.51	9.88
Fall Oral Clause	17.57	9.96
Fall Silent Idea Unit	16.76	13.67
Fall Silent Clause	17.06	13.83
Spring Oral Idea Unit	28.76	15.01
Spring Oral Clause	28.39	15.81
Spring Silent Idea Unit	22.27	14.36
Spring Silent Clause	24.37	15.57
Fall GORT-5 Comprehension	6.53	1.58
Spring GORT-5 Comprehension	7.56	1.96
Spring IOWA Reading (third grade)	192.70	18.00
Spring IOWA Reading (fourth grade)	205.75	25.84

Note. Retell Means and Standard Deviations are based on percent correct values; IOWA Assessments Reading subtest Means and Standard Deviations are provided separately by grade because these values change based on the grade and form used.

Moderate correlations were found between the fall GORT-5 Comprehension subtest and both the clause-based and idea unit-based retell scoring methods in the fall as well as the spring when the student read orally, $r = .427\text{-}.547$, $p < .01$. When the student read silently, fall retell scores also correlated moderately with the fall GORT-5 comprehension subtest, $r = .546$, $p < .01$ for clause-based scoring and $r = .549$, $p < .01$ for idea unit-based scoring. Interestingly, a strong correlation was found between the fall GORT-5 Comprehension subtest and both retell scoring methods when the student read silently in the spring, $r = .739$, $p < .01$ for clauses and $r = .720$, $p < .01$ for idea units. When comparing retell scoring methods with the spring GORT-5 Comprehension subtest, similar results were found. Moderate correlations were found across all scoring methods when the student read orally in the fall and spring, as well as silently in the fall, $r = .586\text{-}.695$, $p < .01$, and once again a strong correlation was found in both scoring methods when the student read silently in the spring, $r = .806$, $p < .01$ for clause-based scoring and $r = .842$, $p < .01$ for idea unit-based scoring.

In contrast, the IOWA Assessment Reading subtest scores (administered in the spring only) demonstrated less consistent relations with the fall and spring retell scores. The IOWA Assessments standard scores were converted to z scores to that they could be collapsed across grade levels for analyses. The correlations between the spring IOWA Assessment reading subtest and the fall idea unit or clause based scores were not statistically significant when passages were read orally, $r = .205$, $p = .229$ and $r = .327$, $p = .052$, respectively. However, moderate and statistically significant relations were observed for both scoring methods when passages were read silently in the fall, $r = .403$, $p = .015$ for both clause-based and idea unit-based scoring methods. With regard to the concurrent relations among the spring retell scores with the IOWA Assessment, a weak correlation was observed from clause-based retells following oral reading, *r*

= .398, p = .016, but none of the other correlations reached significance, r = .314-.324, p > .05. In sum, retell methods within each time point demonstrated moderate to strong concurrent relations with the GORT-5 measure of reading comprehension, but correlations were not as robust when comparing retell scores and the IOWA Assessments reading measure.

Sensitivity to Growth across the School Year

The third research question addressed whether the retell methods were sensitive to detecting growth in reading comprehension across the school year (from fall to spring). To examine this, a series of paired-samples t-tests were used to compare the fall and spring retells scores for the idea unit-based and clause-based methods. For both retell scoring methods and across reading modalities (i.e., oral and silent), significant growth was observed across the school year. Specifically, scores from retells following oral reading were higher in the spring compared to the fall for both clause-based, $t(1, 35) = -3.945$, $p < .001$, $d = .819$ and idea unit based methods, $t(1, 35) = -4.536$, $p < .001$, $d = .885$. Similar growth, but smaller effect sizes, was found across the fall and spring scores for the silent reading retell data for clause-based, $t(1, 35) = -2.658$, $p = .012$, $d = .496$, and idea unit-based scoring methods, $t(1, 35) = -2.225$, $p = .033$, $d = .393$. In comparison to a standardized measure of reading comprehension, the GORT-5 Comprehension subtest also showed significant growth from the fall to the spring, $t(1, 35) = -3.631$, $p = .001$, $d = .579$.

Table 3
Correlations among Retell Scores and Criterion Reading Comprehension Measures

	Fall GORT-5	Spring GORT-5	Spring IOWA
Fall Oral Idea Unit	.305 (.521**)	.535** (.695**)	.205
Fall Oral Clause	.325 (.547**)	.513** (.674**)	.327
Fall Silent Idea Unit	.326 (.549**)	.441** (.600**)	.403*
Fall Silent Clause	.324 (.546**)	.428** (.586**)	.403*
Spring Oral Idea Unit	.241 (.427**)	.437** (.596**)	.314
Spring Oral Clause	.293 (.504**)	.465** (.626**)	.398*
Spring Silent Idea Unit	.479** (.720**)	.714** (.842**)	.324
Spring Silent Clause	.499** (.739**)	.665** (.806**)	.322

Note. *p<.05; **p<.01; Correlations in parenthesis were corrected for restricted range; IOWA Assessments correlations were based on z-scores to account for grade-level differences in the standard scores.

ROC Analyses

Receiver Operating Characteristic (ROC) curve analyses were conducted to examine the classification accuracy of spring retell scores using the spring GORT-5 Comprehension subtest and IOWA Assessments Reading subtest as criterion measures. Because the IOWA Assessments Reading subtest criterion measure was only administered in the spring, only spring retell and GORT-5 comprehension subtest data were used to conduct these analyses. Area under the curve (AUC) values represent the likelihood of a test to accurately differentiate participants with specific disorders from typically developing participants (Youngstrom, 2013), and are commonly interpreted using the following criteria: ≥ .90 excellent, ≥ .80 good, ≥ .70 fair, < .70 poor (Swets, Dawes, & Mohanan, 2000). When the spring GORT-5 comprehension subtest was used as the

criterion measure, AUC values were poor for spring oral retell data, AUC = .638 for idea units, AUC = .647 for clauses, but excellent for spring silent retell data, AUC = .850 for idea units, AUC = .844 for clauses (see Figure 1). However, when the Iowa Assessments Reading subtest was used as the criterion measure, the resulting AUC values were poor across spring idea unit-based and clause-based retell scoring methods for passages read orally, AUC = .626 for idea units and AUC = .648 for clauses, as well as silently, AUC = .587 for idea units and AUC = .568 for clauses (see Figure 2).

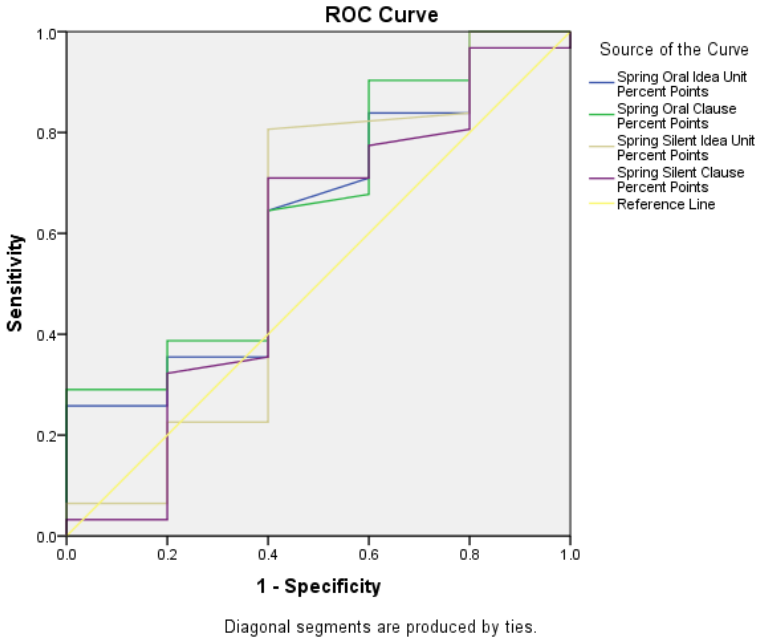

Figure 1. ROC curve analysis with the GORT-5 Reading Comprehension subtest as a criterion measure.

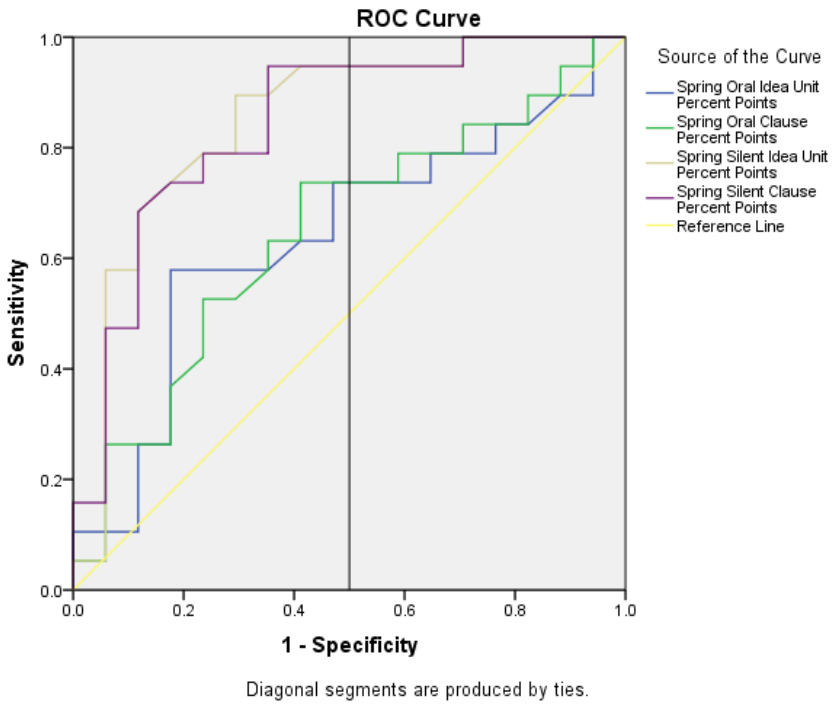

Figure 2. ROC curve analysis with the IOWA Assessments Reading subtest as a criterion measure.

Discussion

Reading comprehension is a difficult construct to measure and skill estimates often vary across assessment measures (Keenan, Betjemann, & Olson, 2008). However, research is always looking to expand the community's knowledge about reading comprehension and how it can be appropriately measured. While retell is commonly used measure of reading comprehension in the classroom, it is uncertain if this measure can accommodate readers with reading or language disabilities such as dyslexia. Even more, there is no set way to score retell data (based on clauses or idea units) that is agreed upon in the literature. This study aimed to compare both scoring methods to determine the validity of retell assessments in measuring reading comprehension by comparing retell data to two standardized, norm-referenced measures of reading comprehension (i.e., the GORT-5 Comprehension subtest and the IOWA Assessments Reading subtest).

Four research questions guided this study; the first addressed the correlation between the retell scores generated from the clause-based and idea unit-based protocols to one another. When comparing both scoring methods, there were significant, strong concurrent correlations across all time points and modalities. These results alone suggest that there is tremendous overlap across these two retell scoring methods, showing minimal differences between the two. This could partly be due to the fact that although both methods are evaluating different content structures in the retell (clause-based scoring assesses phrase/sentence structure while idea unit-based scoring assesses pure information without regards to grammatical organization), ultimately the same retell with the same content is being evaluated.

When comparing predictive relations, correlations were less robust. For passages that were read orally, fall idea unit-based scores correlated weakly with spring idea unit-based and clause-based scores. Correlations were also weak for both spring scoring methods when the student read orally and retells were scored based on clauses. When reading silently, however, correlations were moderate between both scoring methods and the idea unit-based scores in the spring, but were weak when comparing both fall scoring methods to the clause-based scores in the spring. These results suggest that although retell methods demonstrate strong concurrent correlations, these associations weaken when looking at correlations predictively. Nevertheless, lower predictive correlations can be expected when the time lapse between the assessment extend to many months (i.e., approximately 8 months). It is interesting to note that predictive correlations were moderate only when the student read silently, demonstrating an accurate depiction of increasing silent reading skill, which is not emphasized in education until the third or fourth grade (Chall, 1983; Chall, 1996; Prior & Welling, 2001). However, there is not a significant difference between retell scoring methods in predictive correlations when the students

read orally, again exemplifying a negligible difference between clause-based and idea unit-based scoring.

The second question that guided this study addressed the strength of the relations between retell scores generated by the clause and idea unit based protocols with the criterion measure of reading comprehension (i.e., the GORT-5 Comprehension subtest and Iowa Assessments Reading subtest). When comparing both retell scoring methods to other measures of reading comprehension, retell scores correlated moderately to strongly to the GORT-5 Comprehension subtest, lending support to the notion that retell assessments are accurate measures of reading comprehension (Keenan, Betjemann, & Olson, 2008). Keenan et al. report that a typical correlation among reading comprehension assessments are moderate and may range from $p = .31$ to $p = .70$. These results are also consistent with previous literature showing a moderate correlation between retell assessments and other measures of reading comprehension (Reed & Vaughn, 2012).

It is important to note that the significant correlations between the GORT-5 Comprehension subtest (for both fall and spring time points) and retell scores were strongest for spring silent reading sessions with virtually no difference between method of scoring. This is perhaps surprising given that the GORT-5 is an oral reading measure and would be expected to correlate more strongly with oral reading retells for either scoring method. In terms of the IOWA Assessments Reading subtest, correlations with the retell scores were mostly weak and not significant, with the exception of a significantly moderate correlation with both fall silent scoring methods and with the spring oral clause-based scoring method. It is curious that a spring-administered assessment correlated more strongly with fall data than with spring data, and that

once again, a correlation between a silent measure (The IOWA Assessments Reading subtest) and an oral measure (the retell passages) are shown.

The third question guiding this study addressed the sensitivity to growth across the school year (fall to spring) of the two scoring methods. Results show that each scoring method was significantly sensitive to growth across the school year for both reading modalities. In terms of the magnitude of the observed effect sizes for these results, some difference was noted between the silent and oral reading modalities. A large effect size was found for retells following oral reading while a small effect size was shown for silently read passage retells across both clause-based and idea unit-based scoring methods. In comparison, the GORT-5 Comprehension subtest had a medium effect size, demonstrating the standard expectation for growth across the school year for this population. Once again, different patterns are observed based on the modality of reading while there is no differentiability between clause-based and idea unit-based scoring. A larger effect size from orally read passages suggests that students with dyslexia improve oral reading skills at a higher rate than silent reading skills. This finding is consistent with extant literature utilizing a sample of typical readers at the same grade-level (Dickens et al., 2015; Elgart, 1978; Hale et al., 2007).

The last question guiding this study addressed the classification accuracy of the retell data in identifying students struggling with regard to reading comprehension. The ROC curve analyses in this study suggest that while measuring reading comprehension using a retell assessment may moderately correlate to other measures of reading comprehension, retell may not be used accurately to distinguish students with reading or language disabilities such as dyslexia from those without. The only exception to this can be seen when evaluating the silent modalities in the spring for both retell scoring methods when using the GORT-5 comprehension subtest as a

criterion measure. It is unclear as to why this specific modality is an excellent classifier for reading disabilities.

A pattern regarding reading modality was revealed in these results that is not consistent with previous research. By third- and fourth-grade, the curriculum places a greater emphasis on silent as opposed to oral reading and it is widely assumed that students are proficient in both modalities (Prior & Welling, 2001). However, oral reading is still shown to elicit higher reading comprehension scores at this grade level (Dickens et al, 2015; Hale et al., 2007).Therefore, predictive correlations that show that there is a stronger relation between both silent reading retell scores in the fall and silent reading idea unit-based scoring in the spring is surprising. This would suggest that in the fall, both scoring methods might accurately predict spring silent reading comprehension scored when the retell is scored by idea units.

Secondly, the strongest correlations between retell scoring methods and the GORT-5 Comprehension subtest for both fall and spring were found when the student read silently in the spring. When comparing to the IOWA Assessments Reading subtest, correlations were weak except for a moderate correlation for fall silent retell scores. Once again, the modality of the retell passage was found to have more variability than the method of retell scoring. These results are again surprising considering the idea that students at these grade levels typically have stronger reading comprehension skills when reading orally (Dickens et al, 2015; Hale et al., 2007). A possible explanation for this could be reading apprehension. As the sample consisted of struggling readers, when reading aloud students may have focused on accuracy while de-emphasizing comprehension skills.

Finally, the effect size was most robust when examining growth across the school year for orally read retell measures. This would suggest that, although retell measures are most related

to other measures of reading comprehension when the student read silently, growth across the school year is more observable when the student reads orally. In other words, retell measures most accurately measure reading comprehension for silent modalities, but are most sensitive to growth across the school year when the passage is read orally.

Limitations and Future Research

Several limitations of this work warrant our attention. The main limitation in this study was the sample size available for analyses. This study was able to access a clinical sample of students attending a small private school specializing in remediating dyslexia, and the potential pool of participants across the third- and fourth-grade was rather small. However, the sample size in this study is comparable to the average sample sizes of typical clinical samples for a study studying dyslexia (Melby-Larvåg, Lyster, & Hulme, 2012). In addition, the sample was not diverse in terms of the students' racial and/or ethnic group membership, which limits the generalizability of these results to more representative populations. Another limitation to the study, and to using retell methods in general, was the subjectivity of both scoring methods. In other words, almost every retell transcription had ambiguous words or phrases that could have been arguably either correct, partially correct, or incorrect. Because each student had some sort of special exception or case needing discussion, it is important to continue comparing retell measures to standardized measures of reading comprehension to accurately monitor the validity of such measures.

Future research should seek to utilize larger samples to replicate results of the current study and perhaps venture to obtain data from both dyslexic and typical students to allow for direct comparisons of the sensitivity of retell as a reading comprehension measure across both groups. Furthermore, it would be interesting to further examine the influence of reading

modalities and the resulting retell data. The differences between reading modality and not retell scoring methods was unexpected. Future investigations may elucidate how the modality influences reading comprehension, and how retell as a measure of reading comprehension may capture these differences.

Conclusion

In conclusion, results from this study suggest that there is no superior way to code or score retell measures when assessing reading comprehension. Although clause-based scoring is more focused on sentence structure and grammar of a retell and idea unit-based scoring focuses solely on the content of the retell, both are highly correlated to each other and, therefore, are both acceptable for use in research environments. In regards to retell as an accurate measure of reading comprehension as a whole, moderate correlations to other measures of reading comprehension such as the GORT-5 Comprehension subtest and the IOWA Assessments Reading subtest suggest that retell is valid in its use as a measure of reading comprehension. Each retell scoring method is sensitive to displaying growth across the school year, although difference in effect size suggests a difference in growth rates across reading modality. Finally, while retell measures accurately assess reading comprehension, they may not be able to distinguish between students with and without reading or language disabilities such as dyslexia.

While there may be no distinguishable difference between the two scoring methods, using idea unit-based scoring for retell measures may be the efficient choice for researchers and practitioners. In making a scoring rubric and manual, researchers were able to distinguish idea units by referral to the QRI-5, whereas clarification was sought by grammar experts in order to ensure accurate identification of clauses within the passages and retell transcriptions. Also, because inferences about the passage were not made as often as expected, retells mostly focused

on pure information rather than amount of phrases spoken in each retell. However, it is important to consider the modality of reading when using retell as a measure of reading comprehension, regardless of scoring method, when utilizing an SLD sample.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (*5th ed.). Arlington, VA: American Psychiatric Association.

Bellinger, J. M., & DiPerna, J. C. (2011). Is fluency-based story retell a good indicator of reading comprehension? *Psychology in the Schools, 48(4),* 416-426.

Best, R. M., Floyd, R. G., & McNamara, D. S. (2008). Differential competencies contribution to children's comprehension of narrative and expository texts. *Reading Psychology, 29(2),* 137-164.

Chall, J. S. (1983). *Stages of reading development.* New York: McGraw-Hill.

Chall, J. S. (1996). *Learning to read: The great debate (revised).* New York: McGraw-Hill.

Cohen, P., Cohen, J., West, S., & Aiken, L. (2003). *Applied Multiple Regression/Correlation for the Behavioral Sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

International Dyslexia Association. (2002). Retrieved April 11, 2016, from http://eida.org/definition-of-dyslexia/

Dickens, R. H., & Meisinger, E. B. (2015). Examining the effects of skill level and reading modality on reading comprehension. *Reading Psychology,* 1-20.

Dunbar, S. B., Welch, C. J., Hoover, H. D., & Frisbie, D. A. (2011). *Iowa Assessments, Form E.* Orlando, FL: Houghton Mifflin Harcourt.

Dunbar, S. B., Welch, C. J., Hoover, H. D., & Frisbie, D. A. (2014). *Iowa Assessments, Form E and F, Research and Development Guide.* Orlando, FL: Houghton Mifflin Harcourt.

Elgart, D. B. (1978). Oral reading, silent reading, and listening comprehension: A comparative study. *Journal of Reading Behavior, 10,* 203-207.

Floyd, R. G., Clark, M. H., & Shaddish, W. R. (2008). The exchangeability of IQs: Implications for professional psychology. *Professional Psychology: Research and Practice, 39,* 414-423.

Gee, J. P. (1999). *An introduction to discourse analysis: Theory and method.* New York, NY: Routledge.

Good, R.H., & Kaminski, R. A. (Eds.) (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement. Retrieved from http://dibels.uoregon.edu

Gulliksen, H. (1987). *Theory of mental tests.* Hillsdale, NJ: Erlbaum.

Hakkarainen, A., Holopainen, L., & Savolainen, H. (2013) Mathematical and reading difficulties as predictors of school achievement and transition to secondary education. *Scandinavian Journal of Educational Research, 57(5),* 488-506.

Hale, A. D., Skinner, C. H., Williams, J., Hawkins, R. O., Neddenriep, C. E., & Dizer, J. (2007). Comparing comprehension following silent and aloud reading across elementary and secondary students: Implications for curriculum-based measurement. *The Behavior Analyst Today, 8,* 9-23

Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading, 12,* 281-300.

Kucer, S. B. (2010). Readers' tellings: narrators, settings, flashbacks and comprehension. *Journal of Research in Reading, 33(3),* 320-331.

Kucer, S. B. (2011). Going beyond the author: what retellings tell us about comprehending narrative and expository texts. *Literacy, 45(2),* 62-69.

LaBerge, D., & Samuels, S. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology,* 6, 293-323.

Leslie, L., & Caldwell, J. (2010). *Qualitative Reading Inventory-5 (QRI-5).* Boston, M.A.: Allyn & Bacon.

Marr, M. B., & Gormley, K. (1982). Children's recall of familiar and unfamiliar text. *Reading Research Quarterly, 18(1),* 89-104.

McNamara, D. S., Ozuru, Y., & Floyd, R.G. (2011). Comprehension challenges in the fourth grade: The roles of text cohesion, text genre, and readers' prior knowledge. *International Electronic Journal of Elementary Education, 4(1),* 229-257.

Melby-Lervåg, M., Lyster, S. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin, 138,* 322-352.

Merritt, D. D., & Liles, B. Z. (1987). Story grammar ability in children with and without language disorders. *Journal of Speech, Language, and Hearing Research,* 30, 539-552.

Prior, S. M., & Welling, K. A. (2001). "Read in your head:" A Vygotskian analysis of the transition from oral to silent reading. *Reading Psychology, 22(1),* 1-15.

Reed, D. K. (2011). A review of the psychometric properties of retell instruments. *Educational Assessment, 16(3),* 123-144.

Reed, D. K., & Vaughn, S. (2012). Retell as an indicator of reading comprehension. *Scientific Studies of Reading, 16(3),* 187-217. Research Randomizer. (n.d.). Retrieved July 10, 2016, from https://www.randomizer.org/

Roberts, G., Good, R., & Corcoran, S. (2005). Story retell: A fluency-based indicator of reading comprehension. *School Psychology Quarterly, 20(3),* 304-317.

Shaywitz, S. (2003). *Overcoming dyslexia: A new and complete science-based program for reading problems at any level*. New York, NY: Vintage Books.

Shinn, M. M. & Shinn, M. R. (2002). *AIMSweb Training Workbook: Administration and Scoring of Reading Curriculum-Based Measurement (R-CBM) for Use in General Outcome Measurement*. Retrieved from http://www.aimsweb.pearson.com.

Snow, C. E., & Sweet, A.P. (2003). Reading for comprehension. In A.P. Sweet & C.E. Snow (Eds.), *Rethinking reading comprehension* (pp. 1-11). New York, NY: Guilford Press.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, XXI,* 360-407.

Tabachnik, B. G., & Fidell, L. S. (2012). *Using Multivariate Statistics* (6th ed.). New York, NY: Pearson

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest, 1*, 1-26.

Tabachnik, B. G., & Fidell, L. S. (2012). *Using Multivariate Statistics* (6th ed.). New York, NY: Pearson

Wiederholt, J. L., & Bryant, B. R. (2012). *Gray Oral Reading Tests* (5th ed.). Austin, TX: Pro-Ed.

Woodcock, R. W., & Johnson, M. B. (1997). *Woodcock Diagnostic Reading Battery (WDRB).* Allen, TX: DLM Teaching Resources.

Youngstrom, E. A., (2014). A Primer on Receiver Operating Characteristic Analysis and Diagnostic Efficiency Statistics for Pediatric Psychology: We Are Ready to ROC. *Journal of Pediatric Psychology, 39(2),* 204-221.

Appendix A

Retell Scoring Rubric Sheet

Student #_____
Fall 2015 ____ or Spring 2016___
Oral ____ or Silent____

1 point: Completely correct phrase or inference. Matches or is synonymous to clause or idea unit listed below

.5 points: Insufficient or partially correct. Lacks detail or enough information to be a full clause or idea unit. May be a summarization

0 points: Incorrect. Clause or idea unit does not match what could be learned or inferred from the text. Also any omissions

Correct Clauses: Fall Day (grade 3)

Clauses	Points Awarded
It rained all day long	0_____ .5_____ 1_____
The wind and rain knocked the remaining leaves to the ground	0_____ .5_____ 1_____
Where they were swept into the street	0_____ .5_____ 1_____
Today was a typical fall day	0_____ .5_____ 1_____
Just two days ago, the sun was out	0_____ .5_____ 1_____
And the temperatures were very pleasant	0_____ .5_____ 1_____
Raking leaves into large playful piles was very relaxing	0_____ .5_____ 1_____
The family worked together gathering the maple leaves into piles	0_____ .5_____ 1_____
Kids will be kids	0_____ .5_____ 1_____
And they loved jumping and hiding in the leaves	0_____ .5_____ 1_____
Even their dog liked to romp around in the leaves	0_____ .5_____ 1_____
It was fun for everyone	0_____ .5_____ 1_____
The next day, the weather changed slightly	0_____ .5_____ 1_____
Clouds began to roll into the area and darken the sky	0_____ .5_____ 1_____
It did not rain then	0_____ .5_____ 1_____
But it was clear	0_____ .5_____ 1_____
That winter was near	0_____ .5_____ 1_____
The family thought	0_____ .5_____ 1_____
That the ground would be covered in no time	0_____ .5_____ 1_____
Winter was approaching fast	0_____ .5_____ 1_____
	Total: _____

Retell Scoring Rubric Sheet

Student #_____
Fall 2015 ____ or Spring 2016___
Oral ____ or Silent____

1 point: Completely correct phrase or inference. Matches or is synonymous to clause or idea unit listed below

.5 points: Insufficient or partially correct. Lacks detail or enough information to be a full clause or idea unit. May be a summarization

0 points: Incorrect. Clause or idea unit does not match what could be learned or inferred from the text. Also any omissions

Correct Idea Units: Fall Day (grade 3)

Idea Units	Points Awarded
Rained all day	0_____ .5_____ 1_____
Leaves knocked to ground	0_____ .5_____ 1_____
By wind/by rain	0_____ .5_____ 1_____
Leaves swept into street	0_____ .5_____ 1_____
Typical fall day	0_____ .5_____ 1_____
Two days ago	0_____ .5_____ 1_____
Sun out	0_____ .5_____ 1_____
Temperatures pleasant	0_____ .5_____ 1_____
Raking leaves was relaxing	0_____ .5_____ 1_____
Maple leaves	0_____ .5_____ 1_____
Family worked together	0_____ .5_____ 1_____
Kids will be kids	0_____ .5_____ 1_____
Loved jumping in leaves	0_____ .5_____ 1_____
Loved hiding in leaves	0_____ .5_____ 1_____
Dog liked to romp in leaves	0_____ .5_____ 1_____
Fun for everyone	0_____ .5_____ 1_____
Next day, weather changed	0_____ .5_____ 1_____
Clouds roll into area	0_____ .5_____ 1_____
Clouds darken sky	0_____ .5_____ 1_____
Did not rain	0_____ .5_____ 1_____
Clear winter near	0_____ .5_____ 1_____
Family thought ground would be covered in no time	0_____ .5_____ 1_____
Winter approaching fast	0_____ .5_____ 1_____
	Total: _____

Appendix B

Clauses: Fall Day (Grade 3) Coding Manual

1. It rained all day long
    a. It rained counts as half credit
2. The wind and rain knocked the remaining leaves to the ground
    a. Must mention both wind and rain for full credit
    b. Leaves were knocked to ground is half credit
3. Where they were swept into the street
4. Today was a typical fall day
    a. It was fall is half credit
5. Just two days ago, the sun was out
    a. Mention of just one of these is half credit
    b. Few days ago is also acceptable
6. Raking leaves into large playful piles was very relaxing
    a. Mention of raking leaves is half credit
7. The family worked together gathering the maple leaves into piles
    a. Worked together can be substituted with raking
8. Kids will be kids
    a. Kids are kids also counts
9. And they loved jumping and hiding in the leaves
    a. Loved=liked
    b. Played outside is half credit
10. Even their dog liked to romp around in the leaves
    a. Roam does not count for credit
    b. Play and jump can count for romp
11. It was fun for everyone
12. The next day, the weather changed slightly
    a. Mention of weather is half credit, mention of the next day is half credit
13. Clouds began to roll into the area and darken the sky
    a. Must mention both for full credit
14. It did not rain then
15. But it was clear
16. That winter was near
    a. Near=coming
17. The family thought
    a. The family knew or could tell is half credit
18. That the ground would be covered in no time
    a. Must say no time
19. Winter was approaching fast
    a. Snow is acceptable instead of winter
    b. Must be a repetition, otherwise credit goes to “winter was near”
    c. Doesn’t have to have fast for full credit

Idea Units: Fall Day (Grade 3) Coding Manual

1. Rained all day
2. Leaves knocked to ground
3. By wind/by rain
    a. Must say both wind and rain for full credit
4. Leaves swept into street
5. Typical fall day
    a. Fall is half credit
6. Two days ago
    a. Few days ago is correct
7. Sun out
    a. Mention of sunny weather acceptable
8. Temperatures pleasant
9. Raking leaves was relaxing
10. Maple leaves
11. Family worked together
    a. Family raked together
12. Kids will be kids
    a. Kids are kids also counts
13. Loved jumping in leaves
    a. Loved=liked
    b. Played is half credit
14. Loved hiding in leaves
    a. Loved=liked
15. Dog liked to romp in leaves
    a. Romp=play=jump
    b. Roam is incorrect
16. Fun for everyone
17. Next day, weather changed
    a. Mention of weather is half credit
18. Clouds roll into area
    a. Cloudy is partial credit, clouds came is partial credit
19. Clouds darken sky
20. Did not rain
21. Clear winter near
    a. Near=coming
22. Family thought ground would be covered in no time
    a. Must say “no time”
23. Winter approaching fast
    a. Must be repetition of winter near