



Published in final edited form as:

Sci Stud Read. 2012 ; 16(3): 187–217. doi:10.1080/10888438.2010.538780.

Retell as an Indicator of Reading Comprehension

Deborah K. Reed and

Educational Psychology and Special Services, The University of Texas at El Paso

Sharon Vaughn

The Meadows Center for Preventing Educational Risk, The University of Texas at Austin

Abstract

The purpose of this narrative synthesis is to determine the reliability and validity of retell protocols for assessing reading comprehension of students in grades K–12. Fifty-four studies were systematically coded for data related to the administration protocol, scoring procedures, and technical adequacy of the retell component. Retell was moderately correlated with standardized measures of reading comprehension and, with older students, had a lower correlation with decoding and fluency. Literal information was retold more frequently than inferential, and students with learning disabilities or reading difficulties needed more supports to demonstrate adequate recall. Great variability was shown in the prompting procedures, but scoring methods were more consistent across studies. The influences of genre, background knowledge, and organizational features were often specific to particular content, texts, or students. Overall, retell has not yet demonstrated adequacy as a progress monitoring instrument.

Studies investigating the skill deficits of those who struggle with reading indicate that word identification, fluency, and comprehension are often distinct categories of ability among upper elementary children and adolescents (Catts, Adlof, & Weismer, 2006; Fletcher, Lyon, Fuchs, & Barnes, 2007; Valencia & Buly, 2004). Students may exhibit difficulty in only one domain (identified by Catts et al., 2006, as *specific deficit in word reading* or *specific comprehension deficit*), or they may struggle with a combination of skills (referred to as *mixed deficit*). Regardless of the number or type of reading abilities concerned, all affected students will demonstrate poor understanding of text.

Given the number of component skills involved in an individual's "reading competence," determining whether students need assistance in one or more areas can be challenging. An instrument designed to measure only one type of ability (e.g., word identification or vocabulary knowledge) might fail to identify those students whose reading difficulty rests largely in another domain. Similarly, instruments of overall comprehension are problematic in that they do not measure equivalent cognitive processes (Cutting & Scarborough, 2006; Spooner, Baddeley, & Gathercole, 2004). Yet, obtaining accurate data, particularly from assessments that can be easily and frequently administered, is viewed as critical to planning effective instruction and preventing reading failure (Coyne, Kame'enui, Simmons, & Harn, 2004; Stecker & Fuchs, 2000).

RATIONALE AND RESEARCH QUESTIONS

It has been suggested that a retell prompt might be added to an oral reading fluency (ORF) measure as a means of improving the validity of the assessment without diminishing its efficiency (Roberts, Good, & Corcoran, 2005). In comprehension research, retelling, recalling, summarizing, and paraphrasing are considered distinct skills that require differing levels of complex thought and different degrees of telling or transforming knowledge (Kintsch & van Dijk, 1978; Scardimalia & Bereiter, 1987). Within studies examining retell as a measurement tool, however, these skills are treated almost interchangeably (Duffelmeyer & Duffelmeyer, 1987). Depending upon the instrument or study, “retell” and “recall” could be used to elicit main ideas, summaries of the content, or a thorough restatement of the passage. In the most common approach, students are asked to read a passage, either silently or orally, and are then prompted to tell or write about the passage in their own words without referring back to the text.

Retells are among the more popular elements of reading comprehension assessment (Nilsson, 2008; Talbott, Lloyd, & Tankersley, 1994), but they have several limitations. Notably, students with learning disabilities (LD) reportedly perform more poorly on retell tasks than students without LD, even after controlling for topical vocabulary and passage comprehension (Carlisle, 1999). Hence, it is possible that retell could not accurately convey a student’s comprehension. There are several possible explanations for this. To retell a passage verbally or in writing, the student must be able to recall information, organize it in a meaningful way, and possibly draw conclusions about the relationships among the ideas (Klingner, 2004). Producing the retell is highly dependent upon the student’s productive language abilities (Johnston, 1981). In fact, oral retell performance reliably differentiates adults with and without aphasia, an impairment in the ability to produce or comprehend language resulting from brain injury (Ferstl, Walther, Guthke, & Yves von Cramon, 2005; McNeil, Doyle, Fossett, Park, & Goda, 2001; Nicholas & Brookshire, 1993).

Moreover, the quality, accuracy, and completeness of students’ written retells are related to their transcription fluency, or the number of letters the students can write in 1 min (Olive & Kellogg, 2002; Peeverly et al., 2007). Some have suggested that assessing comprehension with open-ended questions, such as a retell prompt, makes it difficult to distinguish among difficulties at the level of input, retrieval, expression, or some combination thereof (Johnston, 1981; Spooner et al., 2004). Others have cautioned that socioeconomic status and cultural-linguistic differences might influence student performance on comprehension tasks that require oral language processing (Snyder, Caccamise, & Wise, 2005). Unfortunately, no known studies have explored this with respect to students’ retell performance or the teachers’ judgments of students’ retell ability.

In addition, there are concerns about the psychometric properties of retells. Reportedly, there is no uniform scoring procedure across instruments (Nilsson, 2008), and the interrater reliabilities are often weak (Klesius & Homan, 1985). Yet retell tasks remain an appealing compliment to ORF measures due to their efficiency, equivalency of format across passages, reliance on active reconstruction of text, and relevancy to comprehension instruction (Hansen, 1978; Roberts et al., 2005). In addition, ORF measures may be more dependent upon students’ decoding ability, a skill that is believed to have a diminishing contribution to reading competence as students become older (Schatschneider et al., 2004; Wiley & Deno, 2005) and better able to rely upon compensatory strategies such as context clues (Savage, 2006). A retell component, therefore, has the potential to detect other instructional areas of need that might be missed by the ORF measure alone. This information would be highly useful in planning reading interventions.

However, no systematic review of the practice has been conducted to determine if a retell component contributes unique, valid, and reliable information about students' reading comprehension. Therefore, this descriptive synthesis seeks to address the following question: How well do retell measures contribute unique, valid, and reliable information about students' reading comprehension?

METHOD

To identify studies of the reliability and validity of retell measures, the Academic Search Complete, PsycINFO, ERIC, and Medline electronic databases were initially searched using the following descriptors: *recall* OR *retell** AND *read** *comprehend** AND *reliab** OR *valid** OR *factor analysis*. No limitation was set on the initial date of publication because there was no reason to believe that the age of the study would be relevant to ascertaining the technical adequacy of a retell protocol. However, this netted only 67 studies and omitted many articles the authors knew to be available. Therefore, a second, wider search was conducted using the descriptors: *recall* OR *retell** AND *read** *comprehend**. Again, date of publication was not used as a delimiter, so articles published through 2009 were included in this review. This identified 921 abstracts, and only 2 articles from the first search did not appear in the second. In addition, an ancestral search of reference lists and technical reports was conducted to ensure no other pertinent studies were missing.

All identified abstracts were evaluated on the basis of the following criteria:

1. Article was published in a peer-reviewed journal. Dissertations and conference papers were excluded due to difficulties in reliably obtaining such manuscripts, particularly those dating back more than 10 years.
2. Participating students were in grades K–12. Studies with older or younger students were included if they also sampled students within the target (school-age) range.
3. Participants were not identified on the basis of sensory impairments.
4. Students in all conditions were assessed using connected text (as opposed to graphic displays, wordless picture books, rebuses or other symbolic representations) that the students had to read. Studies in which the text was read to students (e.g., Moss, 1997; Wright & Newhoff, 2001) were omitted to avoid confounding with listening comprehension.
5. Retell was used as an indicator of reading comprehension, not verbal/oral/listening comprehension, general memory, or a measure of language processing (e.g., Burke & Hagan-Burke, 2007). Studies that involved recall of isolated words or grammatical elements were omitted, as were studies that had students produce pictures, graphics, or role plays for the recall.
6. The study was not intended to manipulate retell ability through the introduction of instruction, study aides, or other treatment (e.g., Gambrell, Pfeiffer, & Wilson, 1985; Popplewell & Doty, 2001).
7. Results reported sufficient data on the retell measure suitable for determining the reliability, validity, and/or utility of retell as an indicator of reading comprehension.

The abstracts were sorted into piles to denote those that clearly violated one or more criteria, those that clearly met all criteria, and those that possibly met the criteria. The lead author read all articles from the latter two categories in their entirety and narrowed the set to the 54 studies included in the synthesis.

Data Analysis

Coding procedures—Studies of retell measures were coded for elements pertinent to this descriptive review. The code sheet included the grade level(s) and characteristics of participants; whether the passages were read orally or silently; the purpose of including a retell measure in the study; whether the retell was provided verbally or in writing; the initial prompt given to students as well as any follow-up prompting; the scoring procedure; and findings related to the reliability, validity, and/or utility of the retell measure. The information from all code sheets was organized in the Table of Retell Study Characteristics to summarize the studies. To facilitate the interpretation of the findings, studies were grouped according to the purpose for which the retell measure was included (e.g., validation study, relationship between retell and decoding ability, genre and organizational features of text, etc.). The code sheet and the table for the studies are available via the Web at <http://www.meadowscenter.org/files/CodesheetA.pdf> and <http://www.meadowscenter.org/files/RetellTables.pdf>.

RESULTS

Retell Study Features

The 54 studies that met the selection criteria (summarized in the Table of Retell Study Characteristics) were fairly evenly distributed over a 34-year period from 1975 to 2009, with a slight increase in more recent publications. Fourteen studies were published in the decade spanning 1975 to 1984, 15 were published from 1985 to 1994, 13 were published from 1995 to 2004, and 12 were published in the 5-year period from 2005 to 2009.

Sample characteristics—Although a combined total of 6,404 students participated in studies of retell measures, this sum is inflated by a single study that included 1,518 students (Riedel, 2007). Excluding that study, participant counts ranged from 5 (Cohen, Krustedt, & May, 2009) to 510 students (Keenan, Betjemann, & Olson, 2008). The majority of studies ($n = 37$) had fewer than 100 participants. One study (Tenenbaum, 1977) did not report the sample size. A variety of ability levels were represented in the aggregate data. However, individual studies might have focused solely on students with disabilities (e.g., Fuchs & Fuchs, 1992; Fuchs, Fuchs, & Maxwell, 1988) or students considered average or above-average readers (e.g., Aulls, 1975; Hagtvet, 2003; Irwin, 1979; Loyd & Steele, 1986; Rasinski, 1990; Riedel, 2007; Vieiro & Garcia-Madruga, 1997).

The selection criteria for this review allowed for students in kindergarten through Grade 12, and only kindergarten was not included in any of the studies identified. Grades 4 to 6 were included more often across studies; however, first graders represent the single largest population in the aggregate data due to the large sample in the Riedel (2007) study. Twenty-nine studies targeted multiple grade levels, so the data depicted in Figure 1 reflect overlapping studies.

Most studies included comparable numbers of male and female participants, with the exception of two studies that included only boys (Fuchs et al., 1988; Pretorius, 1996). Other student characteristics were less consistently reported across studies. Most ($n = 38$) did not report the ethnic composition of participants. Of the 16 studies providing, at least, some information on students' racial and ethnic backgrounds, 5 studies reported predominately (85% or greater) Caucasian samples (Applegate, Applegate, & Modla, 2009; Kucer, 2009; Rasinski, 1990; Shinn, Good, Knutson, Tilly, & Collins, 1992; Walczyk & Raska, 1992). Three studies had samples between 30 and 50% varied with respect to ethnicity (Carlisle, 1999; Fuchs et al., 1988; Pearman, 2008), and 3 studies referred more generally to students being of diverse backgrounds (Richgels, McGee, Lomax, & Sheard, 1987; Williams, 1991;

Zinar, 1990). Five studies reported the majority of participants (60–100%) were not European American/Caucasian/White (Best, Floyd, & McNamara, 2008; Cervetti, Bravo, Hiebert, Pearson, & Jaynes, 2009; Marcotte & Hintze, 2009; Riedel, 2007; Roberts et al., 2005). Five studies (Cervetti et al., 2009; Cohen et al., 2009; Marcotte & Hintze, 2009; Pretorius, 1996; Riedel, 2007) referred to the inclusion of second language learners, whereas two studies (Keenan et al., 2008; Pearman, 2008) reported excluding second language learners.

Purpose for including retell in the study—As the grouping and labels in the Table of Retell Study Characteristics indicate, 10 of the 54 studies sought to determine the validity of a retell measure. An additional 10 studies examined the relationship between retell and decoding, language ability, and/or fluency. These relationships contribute information to the discriminant validity of retell as a measure of reading comprehension (rather than a measure more indicative of other skills) as well as the reliability of the instruments when assessing students at different age or developmental levels. Similarly, 5 studies addressed the difference in retell performance among students with and without LD. The remaining studies ($n = 29$) included in this review focused on the potential measurement errors associated with retell protocols, including the potential influence of the prompting condition ($n = 2$), the genre and organizational features of text ($n = 12$), reading from print versus electronic texts ($n = 3$), delivering an oral or written retell ($n = 1$), and aspects of text difficulty ($n = 11$). As might be expected, many of the studies did not fit neatly into one category. Although they are listed in the Table of Retell Study Characteristics by the primary information about retell each was judged to have contributed, they are discussed in the results section in all applicable categories.

Retell measure format—The format of the retell measures employed in the studies differed in three primary ways. First, students could have been reading orally ($n = 13$), silently ($n = 21$), or in combination ($n = 5$). In nearly one third of the studies ($n = 15$), the author(s) did not identify the type of reading conducted prior to the administration of the retell. The second variation in the format concerned the type of text read prior to the administration of the retell. Passages could have been expository ($n = 26$), narrative ($n = 19$), or both ($n = 4$). Five of the studies did not specify the text genre. Finally, students could have been asked to provide their retell orally ($n = 34$), in writing ($n = 15$), or both ($n = 3$). In 2 of the studies, it was not clear whether students were providing oral or written retells. The format of the retell measures is included in each study's information on the Table of Retell Study Characteristics. In addition, Table 1 provides a matrix of these format variations to better depict the types of studies conducted.¹

Findings From Retell Studies

This section begins by addressing one aspect of retell instruments' technical adequacy: the interrater reliabilities reported in studies of different measures. Then, the major categories used to group studies in the Table of Retell Study Characteristics are each explored in depth. As previously mentioned, those studies that provided information pertinent to more than their primary category are discussed in all applicable subtopics.

Interrater reliability—A total of 32 studies provided interrater reliabilities with an overall range of 59 to 100% agreement. The two studies that compared oral and written retells reported higher agreements for scoring written retells (Fuchs et al., 1988; Marcotte & Hintze, 2009). However, data from across studies do not indicate a clear advantage for one

¹Table 1 is also available at <http://www.meadowcenter.org/files/RetellTables.pdf>. At this Web site, the Table of Retell Study Characteristics is identified as Table 1, and the current Table 1 is identified as Table 2.

response format over the other. The mean interrater reliability for all reported written retell scoring was 91%, the same as the mean for all reported oral retell scoring.

Just 1 of the 54 studies in this review (Rasinski, 1990) did not explicitly describe how the retells were scored, and merely 4 studies reported exclusively using a qualitative approach—all of which relied upon an analysis of narrative story grammar elements (Applegate et al., 2009; Doty, Popplewell, & Byers, 2001; Hagtvet, 2003; Pearman, 2008). Among those latter 3 studies, only Pearman (2008) provided interrater reliability information (84%), which was below the mean but within the range of the interrater reliability data from quantitatively scored retells. Seven other studies utilized a mixture of qualitative and quantitative approaches to scoring retells. These combined a count or proportion of predetermined idea units or text-based sentences (interrater agreement ranging 78–99%) with a scale score of how well the retell organization matched the text's structure (interrater agreement 88–91%; Richgels et al., 1987; Yochum, 1991), a holistic judgment of the student's "communicative effectiveness" (agreement not specified; Tindal & Parker, 1989), a scaled level of coherence of the response (agreement from 72% to potentially 90%, if the latter referred to qualitative analysis alone; Cote, Goldman, & Saul, 1998; Loyd & Steele, 1986), or a narrative schemata analysis of story grammar (unclear whether agreement of 96% applied to quantitative or qualitative analysis; Vieiro & Garcia-Madruga, 1997).

The clear preference in the design of studies of retell was to use quantitative approaches to scoring, typically involving numerical counts of words, idea units, propositions, or story elements. These approaches were used exclusively in 41 of the 54 total studies, and in 24 of the 32 studies reporting interrater reliability. Despite nuanced variations in the quantitative methods as described in the Table of Retell Study Characteristics, Fuchs et al. (1988) found no significant differences among scoring by number of words, percent of content words matching original text, or percentage of predetermined idea units. This is particularly noteworthy because the study employed both written and oral retells after both oral and silent reading. However, the students were allowed 10 min to respond with repeated prompting if they paused for 30 s. This yielded a longer period that involved more examiner cuing than was reported in other studies of oral retell.

What was not addressed in the studies was interpretation of the numerical counts. In many cases, the counts were converted into a proportion of idea units recalled (e.g., Best et al., 2008; Curran, Kintsch, & Hedberg, 1996; Hansen, 1978; McGee, 1982; Miller & Keenan, 2009; Pflaum, 1980; Richgels et al., 1987; van den Broek, Tzeng, Risden, Trabasso, & Basche, 2001; Zinar, 1990). However, little guidance was provided for making conclusions about what a desirable percentage of recalled idea units might be, or what percentage might indicate comprehension difficulty. Hansen (1978) noted that even on-grade-level readers recalled only about one third of the idea units. In comparing the average performance of third and fifth graders, McGee (1982) found that students recalled a low of less than 20% of the main ideas and less than 30% of the details to a high of about 50% of the main ideas and less than 40% of the details. The most striking results were found in a study of students in Grades 6, 8, and 11 who only retold about 10% of the predetermined idea units (Tindal & Parker, 1989).

These relatively low percentages were common across studies, suggesting the proportions recalled are not equivalent to percentages for traditional classroom grades (e.g., 90% or better indicating an A grade, 80–89% indicating a B grade, etc.). In all studies with quantitative scoring techniques, interrater reliability was based on the count itself, not on a translation of the tally or proportion to categories of "better" or "weaker" reading comprehension skill. No studies using either quantitative or holistic approaches to scoring

examined teacher or student factors that might influence the scoring and/or interpretation of results.

Aspects of validity—Thirteen studies provided information on retell's validity as a reading comprehension measure. Of primary importance is whether retell assesses some observable skill indicative of comprehension. In modeling the factor structure of reading, Shinn and colleagues (1992) found that the number of recognizable words recalled in writing loaded onto the single factor of overall reading competence for Grade 3 (.53) and onto the comprehension construct in a two-factor model of reading at Grade 5 (.74). In another study, data from participants at ages 8 to 18 revealed that oral retell scores (.62) loaded more strongly onto the comprehension factor than Woodcock–Johnson passage comprehension scores (.43) in a two-factor model distinguishing decoding from comprehension (Keenan et al., 2008).

Although not designed to examine the latent constructs of reading competence, six studies sought to determine the relationship between retell performance and other standardized measures of reading comprehension. The strength of the correlations discussed in this and later sections will be judged conservatively using the following scale of absolute correlation coefficient values (Williams, 1968): (a) 0.00 to 0.30: weak, almost negligible relationship; (b) 0.30 to 0.70: moderate correlation, substantial relationship; and (c) 0.70 to 1.00: high/strong correlation, marked to perfect relationship.

For the large sample of first-grade participants (Riedel, 2007), oral retell results were moderately correlated ($r = .39-.69$) to the Vocabulary and Comprehension subtests of two standardized measures of reading, GRADE and TerraNova. This is consistent with the results of Marcotte and Hintze (2009), who found that fourth graders' oral retells were moderately correlated to GRADE and other informal measures of comprehension ($r = .45-.49$). Moreover, written retells had slightly stronger correlations ($r = .49-.59$).

Similarly, the written retells of students with LD and/or emotional/behavior disorders (EBD) in Grades 4 and 8 were moderately correlated ($r = .59-.82$) to the SAT-7 Reading Comprehension subtest (Fuchs et al., 1988). In another study of upper-middle grades students with and without LD (Hansen, 1978), Spearman's rank correlation coefficients revealed the proportion of idea units recalled by fifth through eighth graders were moderately to highly correlated with performance on open-ended, factual comprehension questions ($\rho = .46-.77$). Fourth graders' oral retell scores in the study conducted by Walczyk and Raska (1992) were also moderately correlated to the Iowa Tests of Basic Skills (ITBS) Reading subtest ($r = .52$), as were the oral retells of 10- and 11-year-old participants ($r = .638$) in Pflaum's (1980) research.

However, Walczyk and Raska (1992) found no significant relationship of retell to the ITBS among the scores of second and sixth graders, despite significant improvements in retell performance at each successive grade level (i.e., Grades 2, 4, 6). The only other potentially disconfirming evidence to the pattern of moderate correlations between retell and other measures of reading comprehension across grades was the weak to moderate correlations ($r = .28-.41$) between the written retells of students in Grades 11 and 12 and the SRA reading comprehension sub-test (Loyd & Steele, 1986). The relationship was even less robust ($r = .11-.26$) when retells were scored by level of coherence rather than by the sum of predetermined idea units weighted for importance. This was interpreted by the authors as an indication that free recall was tapping skills similar to the standardized measure of reading comprehension as well as unique skills associated with the underlying construct.

Four studies (Fuchs & Fuchs, 1992, Studies 1 and 2; Malicky & Norman, 1988; Tindal & Parker, 1989) further explored the utility of retell instruments in monitoring students' progress and in discriminating among students at different levels of reading comprehension. These generally revealed the limitations of retell instruments. Written retells, measured by the total number of words written, produced instable scores over 15-week periods in the two studies conducted by Fuchs and Fuchs (1992). For fourth- and fifth-grade participants with LD and/or EBD in the first of the Fuchs and Fuchs studies, the size of the slope was small in relation to the average standard errors of estimate. For participants at an unspecified age/grade in the second of the Fuchs and Fuchs studies, the standard errors of estimate were lower, but the slopes decreased proportionally, resulting in considerable bounce relative to growth.

No other study in which retells were administered as frequent progress monitoring measures were identified for inclusion in this review, so it is unknown whether these findings would hold true for upper grade levels, students without disabilities, or different retell instruments. The remaining two studies of the four in this category, however, examined issues of concurrent or discriminant validity. Total word count had a negative relationship to the inclusion of passage content information when comparing written summary retells of sixth, eighth, and eleventh graders (Tindal & Parker, 1989). Although there were no significant differences in the performance of students at Grades 6 and 8, students in Grade 11 included significantly less content in their summaries than either younger grade despite writing more total words. The cluster of scores representing total words, number of predetermined idea units, and ratings of communicative effectiveness was progressively looser in each successive grade, indicating the skills associated with retell were increasingly differentiated. However, the standard errors of measurement (SEM) associated with retell scores standardized by the standard deviation were barely able to discriminate eighth graders at the 70th and 90th percentiles. SEM could not differentiate students at the 10th and 50th percentiles.

In another study, retells of students in Grades 1 to 9 with perceived reading problems were able to distinguish two types of readers with low ability to integrate information from text (Malicky & Norman, 1988). At the lowest level of integration skill, student retells included a high percentage of erroneous information and little verbatim recall. The authors characterized these readers as prioritizing word reading accuracy over making meaning from text. At the next level of integration ability, students were better able combine units of text-based information in their retells but were not yet making many inferences, which was interpreted as a function of students' reading proficiency.

The relationship between retell and word-level or language skills is explored more thoroughly in the next section. To summarize the findings on the validity of retell measures, available data indicate oral and written retells load onto the construct of reading comprehension and are, overall, moderately correlated with other measures of comprehension. Nevertheless, retell scores derived through different quantitative methods have not yet demonstrated they function well in monitoring students' reading progress or in determining their understanding of narrative and expository texts across the full range of readers.

The relationship between retell and decoding ability—Five studies examined the retell performance of students categorized by different decoding levels. Among second graders, IQ and decoding ability made significant contributions to oral retell scores based on predetermined elements of story structure (Hagtvet, 2003). Students classified as good, average, and poor decoders had significantly different retell performance even when IQ was covaried, thus indicating an interdependence between decoding and retell. However,

vocabulary knowledge was the best predictor of retell and made a significant contribution to retell performance over and above that of IQ and decoding ability.

The study of students between the ages of 8 and 18 conducted by Keenan and colleagues (2008) looked at the role of decoding developmentally rather than by IQ or ability level. Results suggested that the contribution of decoding skill to oral retell was higher for younger students. In the aggregate data, the amount of variance accounted for by the interaction of chronological age, decoding, and retell (.033) as well as reading age, decoding, and retell (.034) was minimal, particularly when compared with the variance the interactions account for in the Woodcock–Johnson Passage Comprehension subtest (.270 and .255, respectively).

It is possible that the performance of students classified as average or poor comprehenders could be influenced by the degree of prior knowledge they bring to bear on the passage reading. Fourth and fifth graders considered poor decoders orally recalled significantly fewer central ideas as compared to peripheral ideas when they lacked prior knowledge of the passage content (Miller & Keenan, 2009). With prior knowledge, however, they recalled similar amounts of central and peripheral ideas. These findings held when adjusted for IQ and listening comprehension.

To summarize the findings with respect to the relationship between retell and decoding or language ability, results are somewhat inconclusive. There is evidence that decoding and language skills make less of a contribution to the retell performance of older students and those with more prior knowledge, but well-developed word-level abilities seem to free up some cognitive resources for students at all ages to store and retrieve meaningful information from text as measured by retelling.

The relationship between retell and fluency—Six studies provided information on the relationship between students' fluency and their retell performance, and an additional four studies compared the contributions of retell and fluency to students' reading comprehension. This line of research helps establish whether retell instruments are capable of identifying students who are dysfluent readers but adequate comprehenders, or vice versa. Often, the identified studies attempted to parcel the components of fluency. For example, oral retell in Grades 3 and 5 was not correlated with phrasing ability but was weakly correlated ($r = .38-.52$) with miscue and reading rate (Rasinski, 2007).

Other research has further subdivided miscues to determine what types of errors in word reading accuracy improve or decrease recall. Among 10- and 11-year-olds with LD, high phonic cue use was significantly associated with adequate oral retell performance ($r = .419$), but meaning-change miscues and self corrections were not (Pflaum, 1980). In addition, none of those oral reading behaviors were related to the retell performance of student participants without LD. Similar classifications of errors in word reading accuracy were used in a study of fourth graders (Kucer, 2009). Students who orally retold the most information read faster, made fewer meaning-change miscues, and corrected more miscues than those who recalled the least amount of information. However, reading level, the number of total miscues, and miscues that did not change the meaning of the text were not related to the amount of information students recalled. In fact, students were more likely to recall information on which they made an acceptable miscue consistent with the meaning of the text than if they made no errors in word reading accuracy at all.

Just as miscues were not necessarily indicative of decreased recall, reading rate among third graders was not always associated with better oral retell (Cohen et al., 2009). Although the fastest word-readers retold the most information, less proficient readers retold more dialogue clauses that were read slowly as compared to narrative clauses read more quickly. At

comparable holistic ratings of oral reading fluency (including overall quality, intonation, punctuation use, and pacing), 2nd- through 10th-grade students could be grouped into three levels of oral retell ability (Applegate et al., 2009). The lack of a relationship between fluency and a qualitative evaluation of the story grammar and personal response elements in students' retells was true across the grade levels.

Not all identified research confirms these findings, however. Fuchs and colleagues (1988) found the written retell scores of fourth- through eighth-graders with LD and/or EBD were highly correlated to an ORF measure ($M r = .75$). With respect to measuring the construct of reading comprehension, other studies have found that ORF accounts for more variance in performance on standardized assessments of comprehension than retell. Counts of the number of words orally retold only improved the predictive accuracy of first graders' reading comprehension by about 1% (Riedel, 2007; Roberts et al., 2005). Among fourth graders in the Marcotte and Hintze (2009) study, the amount of unique variance in comprehension accounted for by written retell was slightly increased (3–8%).

Comparing the retell performance of students with and without LD—All five studies that examined whether students with LD demonstrated differential performance on retell measures found that students without LD recalled significantly more information (Carlisle, 1999; Curran et al., 1996; Hansen, 1978; Hess, 1982; Williams, 1991). Even after covarying the content vocabulary knowledge of sixth and eighth graders, students without LD still had better constructed and elaborated oral retells in which a greater proportion of the information was attributable to main ideas rather than subordinate details (Carlisle, 1999).

Similar results were reported by Hansen (1978), who found that students in fifth and sixth grade without LD had more partially correct propositions and recalled significantly more superordinate propositions. However, all students included similar amounts of subordinate details and intrusions (information not explicitly provided in the passage) in their oral retells. The kinds of intrusions made by fourth and sixth graders were more closely analyzed by Hess (1982), who found an age as well as an ability difference in oral retell performance. Students in sixth grade made significantly fewer unacceptable intrusions that were not consistent with the passage's theme did than fourth graders. Within each grade level, students without LD made significantly fewer unacceptable intrusions than those with LD. By contrast, there were no significant differences by ability level in the number of acceptable, thematically consistent intrusions made, despite a significant grade-level difference.

In addition to interactions by age, retell performance of students with LD has also demonstrated interactions with socioeconomic status (SES). Students in Grades 8 and 9 with LD and considered to be at a low SES orally recalled significantly fewer relevant ideas from text than students at a low SES without LD (Williams, 1991). Although there was no difference in retell performance among students at a low SES without LD and students at a high SES with LD, the latter students recalled significantly less than students a high SES without LD. In other words, having LD or a low SES was associated with lower retell performance but having both was associated with the worst retell performance.

Across all studies reviewed, students without LD had more complete retells that included more text-based information. However, all students recalled explicit details or subordinate ideas at a lower frequency. It is interesting to note that every study specifically exploring differential performance of students with LD employed oral retells, so it is not possible to determine whether the pattern of performance on written retells would be consistent with the findings reported here. Moreover, all five studies included follow-up prompting, usually to

encourage students to tell more, but this did not ameliorate the disparity in the amount of information recalled by students with LD. By contrast, third and fifth graders identified as having low comprehension ability (but not LD) did increase the amount of information retold when specifically probed to a total more comparable to that of their peers with better comprehension ability (Bridge & Tierney, 1981; Zinar, 1990).

Prompting condition—Although only one study addressed the influence of a stated reading goal or focus on students' retells (Gagne, Bing, & Bing, 1977), there was remarkable variety in the prompts and other cuing provided to students across all studies. Only about 15% ($n = 8$) of the studies provided complete verbatim accounts of the initial prompt and any follow-up prompting or cuing used to elicit the retell from students (Best et al., 2008; Cervetti et al., 2009; Harris, Mandias, Terwogt, & Tjinjelaar, 1980; Hess, 1991; Riedel, 2007; Yochum, 1991; Zinar, 1990). In fact, it was more common within the corpus of studies to not report the prompt at all ($n = 12$). From what could be determined, some version of any of the following types of directions could have been delivered to students: compose a summary (Fuchs & Fuchs, 1992; Tindal & Parker, 1989; Vieiro & Garcia-Madruga, 1997), retell or write as many details/everything they can remember about what they read (Bridge & Tierney, 1981; Carlisle, 1999; Cohen et al., 2009; Doty et al., 2001; Hagtvet, 2003; Kerr & Symons, 2006; Malicky & Norman, 1988; Marcotte & Hintze, 2009; McGee, 1982; McNamara, Kintsch, Songer, & Kintsch, 1996; Penning & Raphael, 1991; Pretorius, 1996; Rasinski, 1990; Richgels et al., 1987; Riedel, 2007; Tenenbaum, 1977; Walczyk & Raska, 1992; Zinar, 1990), paraphrase or retell the passage in their own words (Hansen, 1978; Hess, 1982; Kucer, 2009; Loyd & Steele, 1986; Shinn et al., 1992; van den Broek et al., 2001), retell or write the story/passage as if telling it to someone who had never read it (Curran et al., 1996), produce a report for their peers (Cote et al., 1998), list 10 facts from the passage (Gagne et al., 1977), or retell the story/passage word for word (Aulls, 1975; Freebody & Anderson, 1983). One study even allowed the examiner to select among different prompts (Pearman, 2008).

Despite some indication that the term *retell* has been used more frequently by researchers when referring to application with narrative text and *recall* more often with expository, the terms were inconsistently applied in the identified studies when prompting students. When considered in light of how students' responses were evaluated (see the Table of Retell Study Characteristics and the section on interrater reliability), the distinctions among *retelling*, *recalling*, *summarizing*, *paraphrasing*, and *identifying the main ideas* are even less distinct.

The numbers and types of allowable follow-up prompts across studies further increase the variability among the procedures employed, including 27 studies without descriptions of follow-up prompting. Based on the information reported, students might have been asked scripted questions based on the reading (Cervetti et al., 2009; Doty et al., 2001; Hess, 1982; Kerr & Symons, 2006; Penning & Raphael, 1991), cued to the major headings or sections in the passage (Best et al., 2008; Rasinski, 1990), asked to write a summary (Freebody & Anderson, 1983; Richgels et al., 1987), encouraged to tell more (Carlisle, 1999; Curran et al., 1996; Fuchs et al., 1988; Marcotte & Hintze, 2009; McGee, 1982; Pearman, 2008; Shinn et al., 1992), asked to tell or write anything else they left out (Cote et al., 1998; Yochum, 1991), probed for elaboration and clarification (Bridge & Tierney, 1981; Kucer, 2009), asked to complete a graphic organizer on story elements (Cohen et al., 2009), or specifically probed about predetermined propositions not freely recalled (Malicky & Norman, 1988; Marr & Gormley, 1982; Zinar, 1990). In some studies, students were both encouraged to tell more and asked scripted follow-up questions (e.g., Hansen, 1978).

The number of combinations of initial and follow-up prompts could not be accurately determined given the lack of specific information in, at least, half the articles. If the

instructions provided to students prior to reading the passages were also included in this analysis, the variation would be even greater. The results from Gagne et al. (1977) suggest this inconsistency can significantly influence retells. Students in Grades 10 to 12 who were provided different reading goals in their preliminary instructions, but the same retell prompt (i.e., write down the first 10 facts that can be remembered from the passage), produced the same amount of information but with qualitatively different content. Students told to read an expository text for discrete and sequential facts about a single topic almost exclusively recalled explicit facts on the topic. Whereas students told to read the same expository text for two to three nonsequential, descriptive attributes of a topic almost exclusively recalled attributes.

When prompting occurs it may also influence students' retell performance. When comparing the performance of students in Grades 4, 7, 10, and undergraduates in college, the school-age students generally included significantly less information in written retells when provided questions during and after reading (van den Broek et al., 2001). The youngest students showed the most severe impairment in recall with questions used during reading. In contrast, the college students benefited from the inclusion of causal relation questions and recalled significantly more information when provided the questions during reading. Students of all ages included in their recalls significantly more story propositions that were also needed to answer the questions, so memory of and attention to information was universally heightened by the nature of the questions asked during or after reading.

The influence of text structures, such as the causal relation probed by van den Broek and colleagues (2001), is addressed in the next section. To summarize the information on retell prompting conditions, the extant literature demonstrates a high variation in the ways in which student recalls are elicited. This could contribute to inconsistencies in student performance across studies or, depending on the flexibility of individual retell protocols, the reliability of the measure within a given study. In addition, no studies explored teacher or student characteristics that might influence the delivery of, or response to, retell prompts.

Genre and organizational features of text—Fourteen studies provided information on how retells might be influenced by narrative or expository text as well as by explicit or implicit structural cues. Because retell protocols utilize texts of different genres and organizational styles, information on students' performance when reading these different texts contributes to an understanding of the measurement errors that might threaten test–retest, parallel-forms, and/or internal consistency reliability of retell instruments.

In comparing recall of expository texts with that of narratives, studies of third graders found that students orally recalled significantly more predetermined propositions in narratives than in expository text (Best et al., 2008; Bridge & Tierney, 1981). With neither genre did students in the Best and colleagues (2008) study include many inferences (1–3%). In comparison, Bridge and Tierney (1981) found that good and poor comprehenders recalled similar ratios of explicit and inferred information in narratives, but good readers generated significantly more inferences than poor comprehenders when reading expository text.

Findings from other studies indicate that narrative text is not always associated with better retell performance. For example, fourth graders orally expressed twice as many misconceptions when reading about a science topic in a narrative version as opposed to a content-equivalent informational version of the text (Cervetti et al., 2009). Because many retell scoring mechanisms take the accuracy of students' statements into account, instruments that include narrative passages on science topics may be vulnerable to systematic error that would reduce the retell protocol's reliability. This concern over measurement error was echoed in another study of eighth and tenth graders who indicated

narratives with the setting explicitly stated in the title were easier for them to understand and recall than those with no explicitly stated setting or the setting provided as part of the retell prompt (Harris et al., 1980). Although there was no significant difference in their oral retell performance when provided the setting in the prompt versus no setting at all, students performed significantly better with an explicit setting in either the title or the prompt as compared to having no explicitly stated setting at all.

This is somewhat consistent with the results of Aulls's (1975) study of sixth graders. Expository texts with an explicit main idea and/or topic significantly improved written recall, and the greatest amount of improvement was associated with the main idea. Yet not all explicitly stated information has demonstrated usefulness. Students in Grades 9 and 11 showed no differences in retell performance when presented newspaper articles with or without headlines and with or without summary paragraphs (Leon, 1997). In fact, the high school students were reported to show no preference for relevant information but tried, instead, to retain all possible information. This could be a function of their sensitivity, or lack thereof, to the linkages among the ideas presented in the text, as explored by Zinar (1990). Fifth graders who were considered strong comprehenders were more likely to retell causal relationships when they were explicitly rather than implicitly stated in expository text, whereas explicit or implicit causal relations did not produce significant differences in the overall retell performance of students identified as having comprehension difficulties. The latter students did not include any causal information in their free recalls, but they included comparable amounts of causal information as their higher ability counterparts when probed.

These results differed slightly from those of Bridge and Tierney (1981) with younger students. Neither good nor poor comprehenders in Grade 3 orally recalled many linkages among the text-based propositions, but when they did, the linkages were more common in free rather than probed recall. However, the inclusion of causal and conditional information distinguished the performance of good comprehenders and resulted in more coherent retells. Poor comprehenders tended to make more global inferences, independent of the relational information, which resulted in less specific recall.

Organizing retells in ways that match the structure of the text seems related not only to ability, but also to developmental levels. McGee (1982) found that Grade 3 on-level readers did not match the structure of the passage, Grade 5 students reading on a third-grade level partially matched the structure of the passage, and Grade 5 on-level readers fully matched the structure.

In addition to qualitative differences in retells, linkages are associated with quantitative differences. Strong correlations ($r = .92-.96$) were found between the number and density of causal relations recalled and the overall length of written retells among 10th and 11th graders (Pretorius, 1996). Recall of the hierarchical links among causal units was also strongly associated with overall length of students' retells ($r = .73$). Shorter retells included causal relationships but had fewer of them and lacked the hierarchical linkages. Therefore, the writing appeared somewhat random and was interpreted as reflecting a less coherent mental representation of the text. Similarly, 5th graders in Irwin's (1979) study included more causal and temporal connectives in their oral recalls when the connectives were explicitly stated in the passage. Students had significantly better recall of propositions when a temporal connective was used to link the ideas, but students struggled more with causal structures even when provided an explicit connective.

Other studies also found the particular type of organizational structure used in the text-influenced student recall. At a more simplistic level, passages that represented the

relationship between main ideas and subtopics were related to better written retells of sixth graders than paragraphs organized in a list like fashion (Aulls, 1975). However, the opposite was true among written retells of 12th graders: Students had significantly better performance with list versus hierarchal organization (Tenenbaum, 1977).

In a more complex analysis of four expository text structures (collection, comparison–contrast, causation, problem-solution), Richgels and colleagues (1987) found causation to be the most challenging organizational pattern for Grade 6 students of all abilities to detect and apply, whereas comparison–contrast was the most recognized and conveyed structure. The more aware students were of a text’s structure, the more likely they were to understand and remember that text as reflected in their written recalls.

When disaggregating results of fifth graders by ability, students at higher achievement levels had better written recall of the attribution (description) text structure, and students at lower levels of achievement had better recall of the comparison structure (Yochum, 1991). Neither group emulated the structure of the text very often. Ability differences were also reported by Penning and Raphael (1991), who found Grade 6 good comprehenders had significantly better oral recall than poor comprehenders of cause–effect, narrative, and problem-solution structures. However, there were no significant differences between groups on description, listlike, or comparison–contrast structures.

The role of text difficulty in retell is addressed more thoroughly in a later section. To summarize the influence of text genre and organization, narrative passages appeared to improve explicit recall but could lead to more misconceptions if the purpose is to convey new content or unfamiliar content to students. The presence of explicit information (e.g., setting, main idea, topic, linkages among ideas) tended to improve retells. Older students and those of higher ability were better able to detect and emulate the organization of the text, and across grade-levels, more coherent and longer retells included more linkages. Certain structures appeared to be harder for students to understand and recall, but inconsistencies in the findings suggest this may be related to particular texts and not the structures themselves. However, the cause–effect structure was more consistently difficult and associated with distinctions among student ability levels.

Presentation of print versus electronic text—Only three studies addressing measurement artifacts associated with the medium of the text were identified for inclusion in this review. Results of studies with second graders indicated that simply changing the presentation of static text from print to an electronic format did not appreciably alter oral retells (Doty et al., 2001), particularly among students identified as high- or medium-proficiency readers (Pearman, 2008). However, lower proficiency students who could access hypertextual supports (i.e., labels, vocabulary definitions, and pronunciations of words or segments of text) in the electronic versions of a story did have significantly better holistic ratings of story elements (Pearman, 2008). The results of these two studies are interesting because they both relied on narrative storybooks and oral retells scored with the same 10-point analysis. This degree of similarity in procedures employed by different researchers was uncommon within the corpus of studies.

How the retells are scored may be related to differences in the results found by Kerr and Symons (2006) with a sample of fifth graders. When considering the number of ideas orally recalled from electronic- or print-based text, students had significantly better performance when reading from the computer. However, there were no reliable differences in inferential cued recall or in free or cued recall efficiency (the product of multiplying the accuracy of recalled ideas by students’ reading rate). Inferential comprehension efficiency, on the other hand, was significantly higher when reading text on paper. It should be noted that students

read the paper text more slowly, which suggests that retelling a text at more than a superficial level may be associated with increased processing time. All that can be determined from the available studies, however, is that electronic texts, particularly those with hypertextual supports, tend to support students' literal recall.

Oral versus written retell—Although students predominately were required to produce retells orally ($n = 34$ studies), only 3 studies included in this review specifically addressed performance differences by retell format. Written retells produced consistently higher correlations with standardized measures of reading comprehension than oral retells in studies conducted with fourth graders (Marcotte & Hintze, 2009) and students in Grades 4 to 8 (Fuchs et al., 1988). Although Fuchs and colleagues (1988) found no differences when scoring by number of words, percentage of content words matching the text, or percentage of predetermined idea units, Vieiro and Garcia-Madruga (1997) more closely analyzed the content of the words students produced and found that written retells only increased the literal recall of third and fifth graders. Inferences and generalizations were significantly greater in oral retells. Therefore, no clear advantage for written retells can be concluded.

Text difficulty—Fifteen studies provided information on how prior knowledge, text coherence/complexity, vocabulary difficulty, and interest influence retell performance. As with genre and organizational features, any differences in student performance when reading texts within and across retell protocols of varying difficulty might indicate a possible threat to the test–retest, parallel-forms, and/or internal consistency reliability.

The most common area of study ($n = 9$) was whether or not background knowledge was important to students' recall of text. Although most studies found that prior knowledge or topic familiarity played some role in retell, such as in increasing the amount of information recalled by students in Grades 6 (Aulls, 1975) and 12 (Tenenbaum, 1977), there were variations among the results. For example, Yochum (1991) found that fifth graders with high levels of prior knowledge included proportionally more text-based information in written retells than those with low prior knowledge, but recall performance was often specific to categories of information within particular texts. Similarly, Malicky and Norman (1988) reported that the use of background knowledge was one factor of information processing ability that distinguished the retell performance of students in Grades 1 through 9, but they were also prone to misusing background knowledge in making inferences that were not consistent with the text.

Some differences in the influence of background knowledge appear to be related to an interaction with students' ability levels. Typically achieving students in Grades 4 and 5 showed no effect of topic knowledge on the amount of central versus peripheral ideas orally recalled, but poor decoders recalled significantly fewer central ideas when they lacked prior knowledge of the text's topic (Miller & Keenan, 2009). With similar age groups, Taylor (1979) found no differences in the oral retells of on-level readers in Grade 3 and fifth graders matched by reading ability when reading unfamiliar information that was not difficult to decode. Both groups recalled significantly less than on-level readers in Grade 5. This changed when the passage content was familiar to students. Then, both fifth-grade groups performed similarly and recalled significantly more than the students in Grade 3. The difference in the quantity of information recalled from familiar and unfamiliar texts was greatest among the fifth graders reading below grade level.

Examining results by literal recall was more common in the corpus of studies focused on background knowledge and was evident in the one related study that examined student interest in the passage topic (Naceur & Schiefele, 2005). Interest among students in Grades 8 to 10 was weakly to moderately correlated to retell scored by number of words ($r = .183$ –

376) and number of propositions ($r = .213-.377$). Only two identified studies addressed the role of background knowledge in inference ability. Fourth graders with prior knowledge of text content included more text-based information in their oral recalls but also inferred more information in cued prompting questions (Marr & Gormley, 1982). Cote and colleagues (1998) attributed the production of coherent, well-integrated oral retells with fourth and sixth graders' ability to use prior knowledge to paraphrase, connect, and reconstruct information.

It is possible that features of the text itself are associated with differences in the applicability of background knowledge. Organizing a text to represent the relationship between the main ideas and subtopics improved written recall only if the passage could be related to sixth graders' prior experience (Aulls, 1975). This is consistent with the findings of McNamara and colleagues (1996). In a more coherent text that specified the relationships among the propositions, students in Grades 7 to 10 exhibited adequate oral retell regardless of whether they had high levels of background knowledge.

Aspects of coherence and text complexity were further explored in five studies, two of which are reported by McNamara et al. (1996). The first compared the oral recall performance of students in Grades 7 to 9 with three different versions of a text: the original, one revised for coherence by explicitly identifying the major subtopics and providing the interconnections between the content words, and one expanded to include additional content about the main topic with a decrease in the irrelevant details. Oral recall was significantly better on the revised version as compared to the expanded and original text, but main ideas were recalled significantly better in both the revised and expanded versions.

The second study by McNamara and colleagues (1996) found students in Grades 7 to 10 orally recalled macropropositions significantly better when the text included specific cues to that information. Likewise, micropropositions were recalled significantly better in text that specifically cued that information. Without any signals, students recalled similar amounts of both macro- and micropropositions. Readers' attention to select pieces of information was also noted in a study of 10th and 11th graders (Otero & Kintsch, 1992). Those who attended more evenly across the text detected an intentionally placed contradiction in information and reportedly mentioned that contradiction more frequently in their written retells than would be expected. Students who did not detect the contradiction recalled significantly less information from the beginning of the passage to the end.

Other studies of text complexity have manipulated the perceived difficulty of the passage. Cote and colleagues (1998) created easier and harder versions of passages based on calculated readability levels, ratings by experts, and the responses of students. In two studies conducted with students in Grades 4 and 6, significantly more information was orally recalled from the easier passages. In the second study, there were no reliable effects of passage difficulty on the coherence of fourth graders' retells, which were mostly fragmentary responses. Sixth graders did demonstrate significant differences by passage difficulty and had a higher occurrence of fragmentary responses on the harder version.

The difficulty of text was also manipulated through the inclusion of low frequency vocabulary words in two studies conducted by Freebody and Anderson (1983). In the first, the difficulty level of the vocabulary only made a significant difference in sixth graders' written retell performance when the scores from the free recall, follow-up summary, and sentence verification tasks were combined. Therefore, the particular placement of the low-frequency vocabulary was examined in the second study. When included in important propositions, written recall performance of the sixth graders significantly decreased. When

difficult vocabulary was used in trivial propositions, the retells were more sophisticated and focused on the important ideas.

Overall, the studies examining facets of text difficulty indicate that the influences of background knowledge, text coherence/complexity, and vocabulary frequency are complex. Attempts to manipulate one element by placement, explicit cuing, or revision may assist some students with some information in some passages but could also result in decreasing the recall of other students, information, or passages. Components of text difficulty are interwoven with each other and with organizational elements. As McGee (1982) noted, differences in performance thought to be related to text structure awareness could also be associated with the degree of difficulty the passage presented to students. In her study, fifth-grade better readers not only found the text (written on a third-grade level) easier but were also more likely to have the requisite background knowledge and experience with expository text. Hence, although this review of studies of retell reported results by categories relevant to reliability and validity, the research reveals a great deal of overlap and interaction among the elements influencing students' retell performance.

DISCUSSION

This descriptive review sought to determine how retell measures contribute unique, valid, and reliable information about students' reading comprehension. Results from the 54 studies reviewed are discussed by the major categories used to group them in the Table of Retell Study Characteristics. Where appropriate, some categories are addressed collectively.

Interrater Reliability

It was most common for researchers to report the percentage agreement between different scorers of student retells than to offer other information on the technical adequacy of the instruments. Across all scoring methods, there was a considerable range in the reported inter-rater reliabilities (59–100% agreement), which makes it difficult to draw conclusions about the consistency with which retells can be scored. What was clear was that quantitative approaches were preferred over qualitative or hybrid scoring methods. In 24 of the 32 studies providing interrater reliability, quantitative approaches were used exclusively.

Aspects of Validity

Results of the studies reviewed indicated that retells contribute unique variance in the comprehension of younger students (Hagtvet, 2003) but tend to load more strongly onto a comprehension construct for older students (Keenan et al., 2008; Shinn et al., 1992). Across grade levels, retell scores exhibit moderate correlations with standardized measures of reading comprehension (Fuchs et al., 1988; Hansen, 1978; Marcotte & Hintze, 2009; Riedel, 2007). However, only two studies conducted by Fuchs and Fuchs (1992) have examined the adequacy of retell as a progress monitoring instrument, with findings that provide little support for the practice.

Relationship Between Retell and Decoding and Between Retell and Fluency

Studies indicated that foundational reading skills make less of a contribution to retell performance of older students than measures that address prior knowledge (Applegate et al., 2009; Keenan et al., 2008). Moreover, measures of fluency and retell seem to be assessing different skills among older students. By Grade 5, ORF is less of a factor in students' reading comprehension (Shinn et al., 1992). Above this age, ORF as a measure of reading progress begins to asymptote (Fuchs, Fuchs, Hops, & Jenkins, 2001; Stage & Jacobsen, 2001) and the correlation between ORF and standardized measures of reading

comprehension emerges as less robust than for younger students (Schatschneider et al., 2004; Wiley & Deno, 2005).

Consistent with Perfetti's (1985) verbal efficiency theory, however, better developed word-level abilities and more efficient processing still free up some cognitive resources for students at all ages to store and retrieve meaningful information from text. Due to reported instability in retell scores among fourth and fifth graders with LD and/or EBD (Fuchs & Fuchs, 1992) and a lack of ability to distinguish the performance of eighth graders by SEM (Tindal & Parker, 1989), it is not yet clear whether retell can serve as a valid and reliable means of monitoring the progress of older students or discriminating among comprehension skill levels.

Comparing the Retell Performance of Students With and Without LD

Defining what reading comprehension information retells can provide might improve the development of retell protocols. Students of all ability levels only provided somewhere between less than 20% and about 50% of the text-based propositions (Hansen, 1978; McGee, 1982; Tindal & Parker, 1989). Students were more likely to recall main ideas than details and were not likely to spontaneously include inferences or implicit information in their recalls (Best et al., 2008; Zinar, 1990). Follow-up prompting was necessary to improve the recall of targeted information among students identified as poor comprehenders (Best et al., 2008; Bridge & Tierney, 1981; Zinar 1990), particularly when using expository texts. Students with LD recalled less information (Curran et al., 1996; Hansen, 1978; Hess, 1982; Williams, 1991) and produced more poorly constructed and less well elaborated recalls than students without disabilities (Carlisle, 1999).

Prompting Condition

Other limitations on the interpretation and use of retell data are derived from the lack of consistency in and influence of the retell prompt. Prompts and the expectations for student responses interchangeably apply the terms *retell*, *recall*, *summarize*, and *paraphrase*. However, these do not measure equivalent cognitive processes (Cutting & Scarborough, 2006; Keenan et al., 2008; Kintsch & van Dijk, 1978; Scardimalia & Bereiter, 1987). Greater attention was paid to reporting the scoring procedures employed in the studies than the procedures for obtaining the gist of what was scored. Although existing retell measures reportedly lack of uniform scoring procedures (Nilsson, 2008) and demonstrate weak interrater reliabilities (Klesius & Homan, 1985), the studies of retell in this review were found to have more commonality in scoring and consistency in interrater reliabilities than for retell prompts.

Findings suggest that variations in the wording of a question or prompt (Fuchs & Fuchs, 1992; Gagne et al., 1997; Seifert, 1994) or in the administration procedures surrounding the prompt (Cordon & Day, 1996; Gambrell & Jawitz, 1993; van den Broek et al., 2001) can substantively alter both the quantity and the quality of participants' responses. As indexes of quantity and quality are the means by which retells are scored, insufficient reporting of the prompts employed and a paucity of data on the outcomes associated with different prompts reduces confidence in interpretations.

Genre, Organizational Features, and Print Versus Electronic Formats

As with the threat to reliability introduced by the retell prompt, researchers also investigated whether aspects of the texts that students are asked to recall might affect the test–retest, parallel forms, and/or internal consistency reliability of retell protocols. When able to access hypertextual supports, computer-based passages may assist students who struggle with reading to recall more literal information (Pearman, 2008), but electronic formats might also

decrease the generation of inferences (Kerr & Symons, 2006). Similarly, manipulating other aspects of text also produces inconsistent results. Some students benefitted from having background knowledge of some text context (Cote et al., 1998; Marr & Gormley, 1982; Miller & Keenan, 2009; Taylor, 1979). However, prior knowledge was not associated with improvements in recall of all content (Yochum, 1991), and students could also misuse prior knowledge, particularly when making inferences (Malicky & Norman, 1988). Attempts to explicitly cue students to particular content in a text were generally successful at increasing their recall of that information (Aulls, 1975; McNamara et al., 1996), but explicitly stated cues or information did not always overcome students' inability to attend to information evenly across a text (Otero & Kintsch, 1992) or detect more relevant information (Leon, 1997).

Oral Versus Written Retells

Some evidence indicates that written retells might be more adequate indicators of reading comprehension than oral retells (Fuchs et al., 1988; Marcotte & Hintze, 2009). However, the correlation of the written retell score to ORF and standardized measures of reading was still within the moderate range, and inferences occurred more frequently in oral retells (Vieiro & Garcia-Madruga, 1997). Moreover, scoring oral retells through a quantitative analysis of predetermined propositions or idea units (as opposed to holistic ratings) produced interrater reliabilities comparable to those for written retells (Fuchs et al., 1988). Hence, it may not be of practical significance to require written responses from students, especially given the time efficiency of scoring oral responses at the moment they are elicited.

Text Difficulty

Texts at lower readability levels (Cote et al., 1998) or with fewer difficult vocabulary words in important propositions (Freebody & Anderson, 1983) tended to increase the amount of information students recalled but did not necessarily improve the quality or coherence of student responses. Several researchers associated the use of linkages or relational information with better reading and retell ability (Bridge & Tierney, 1981; Irwin, 1979; Pretorius, 1996; Zinar, 1990), but particular text structures (i.e., causal relations) appeared to be more challenging for students (Richgels et al., 1987; Yochum, 1991). Even with greater awareness of text structures, students still recalled less than 55% of the main ideas in short passages specially written to present a consistent and recognizable organizational pattern (Richgels et al., 1987). Because texts of different complexity may be used within and across retell protocols, the test–retest, alternate-form, and/or internal consistency reliability of the instruments may be threatened.

Limitations and Directions for Future Research

The available data from the studies included in this review were often from a limited subset with a particular focus and employing unique protocols. Many studies addressed multiple elements of technical adequacy and demonstrated how interwoven and complex the influences on retell are. By presenting information across studies in distinct categories, some of the reciprocal nature of these influences is lost.

In addition, the findings reported are limited by the type and quality of the other measures of reading comprehension employed by the researchers. As noted in the introduction, different instruments of overall comprehension do not measure equivalent cognitive processes (Cutting & Scarborough, 2006; Spooner et al., 2004).

Hence, the generalizations made about retell measures are tenuous. Much more research is needed to provide a convergence of evidence on the reliability and validity of retell measures. The conclusions and recommendations provided in this review can only be

considered preliminary. To advance the field, future studies should address the optimal wording of the initial prompt administered in a retell protocol. To the extent that variations in how and when a question is asked (Fuchs & Fuchs, 1992; Gagne et al., 1997; Seifert, 1994; van den Broek et al., 2001) or how instructions are provided (Cordon & Day, 1996; Gambrell & Jawitz, 1993) can substantively alter both the quantity and the quality of participants' responses, retell scores can confound students' comprehension with the influence of the prompt.

Future research might determine the number or proportion of idea units that are associated with "better" or "weaker" comprehension in order to guide teachers in making instructional decisions. Finally, studies might be conducted to compare performance in oral and silent reading and to explore the influence that teacher or student characteristics might have on the assessment of retell performance.

A well-defined line of research on retell measures would explicate their role in assessing students' reading comprehension. If retells are less sensitive to decoding ability and are able to distinguish comprehension skills from other component skills, retell protocols could become valuable tools in schoolwide approaches to reading intervention that rely upon cost-effective and time-efficient data gathered at multiple times throughout the year.

Acknowledgments

This research was supported by grant P50 HD052117 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Eunice Kennedy Shriver National Institute of Child Health and Human Development or the National Institutes of Health.

References

- *References marked with an asterisk indicate articles or assessments synthesized for the current article.
- *. Applegate MD, Applegate AJ, Modla VB. "She's my best reader; she just can't comprehend": Studying the relationship between fluency and comprehension. *The Reading Teacher*. 2009; 62:512–521.
 - *. Aulls MW. Expository paragraph properties that influence literal recall. *Journal of Reading Behavior*. 1975; 7:391–400.
 - *. Best RM, Floyd RG, McNamara DS. Differential competencies contributing to children's comprehension of narrative and expository texts. *Reading Psychology*. 2008; 29:137–164.
 - *. Bridge CA, Tierney RJ. The inferential operations of children across text with narrative and expository tendencies. *Journal of Reading Behavior*. 1981; 13:201–214.
 - Burke MD, Hagan-Burke S. Concurrent criterion-related validity of early literacy indicators for middle of first grade. *Assessment for Effective Intervention*. 2007; 32(2):66–77.
 - *. Carlisle JF. Free recall as a test of reading comprehension for students with learning disabilities. *Learning Disability Quarterly*. 1999; 22(1):11–22.
 - Catts HW, Adlof SM, Weismer SE. Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language, and Hearing Research*. 2006; 49:278–294.
 - *. Cervetti GN, Bravo MA, Hiebert EH, Pearson PD, Jaynes CA. Text genre and science content: Ease of reading, comprehension, and reader preference. *Reading Psychology*. 2009; 30:487–511.
 - *. Cohen L, Krustedt RL, May M. Fluency, text structure, and retelling: A complex relationship. *Reading Horizons*. 2009; 49:101–124.
 - Cordon LA, Day JD. Strategy use on standardized reading comprehension tests. *Journal of Educational Psychology*. 1996; 88:288–295.

- *. Cote N, Goldman SR, Saul EU. Students making sense of informational text: Relations between processing and representation. *Discourse Processes*. 1998; 25(1):1–53.
- Coyne MK, Kame'enui EJ, Simmons DC, Harn B. Beginning reading intervention as inoculation or insulin: First-grade reading performance of strong responders to kindergarten intervention. *Journal of Learning Disabilities*. 2004; 37:90–105. [PubMed: 15493232]
- *. Curran CE, Kintsch E, Hedberg N. Learning-disabled adolescents' comprehension of naturalistic narratives. *Journal of Educational Psychology*. 1996; 88:494–507.
- Cutting LE, Scarborough HS. Prediction of reading comprehension: Relative contribution of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading*. 2006; 10:277–299.
- *. Doty DE, Popplewell SR, Byers GO. Interactive CD-ROMS storybooks and young readers' reading comprehension. *Journal of Research on Computing in Education*. 2001; 33:374–384.
- Duffelmeyer FA, Duffelmeyer BB. Main idea questions on informal reading inventories. *The Reading Teacher*. 1987; 41:162–166.
- Ferstl EC, Walther K, Guthke T, Yves von Cramon D. Assessment of story comprehension deficits after brain damage. *Journal of Clinical and Experimental Neuropsychology*. 2005; 27:367–384. [PubMed: 15969358]
- Fletcher, JM.; Lyon, GR.; Fuchs, LS.; Barnes, MA. *Learning disabilities: From identification to intervention*. New York, NY: Guilford; 2007.
- *. 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, Fuchs D, Hosp MK, Jenkins JR. Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*. 2001; 5:239–256.
- *. Fuchs LS, Fuchs D, Maxwell L. The validity of informal reading comprehension measures. *Remedial and Special Education*. 1988; 9(2):20–28.
- *. 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; 69:428–431.
- Gambrell LB, Jawitz PB. Mental imagery, text illustrations, and children's story comprehension and recall. *Reading Research Quarterly*. 1993; 28:265–276.
- Gambrell LB, Pfeiffer WR, Wilson RM. The effects of retelling upon reading comprehension and recall of text information. *Journal of Educational Research*. 1985; 78:216–220.
- *. 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.
- *. 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.
- *. Harris PL, Mandias F, Terwogt MM, Tjintelaar J. The influence of context on story recall and feelings of comprehension. *International Journal of Behavioral Development*. 1980; 3:159–172.
- *. Hess AM. An analysis of the cognitive processes underlying problems in reading comprehension. *Journal of Reading Behavior*. 1982; 14:313–333.
- *. Irwin JW. Fifth-grade readers' comprehension of explicit and implicit connective propositions. *Journal of Reading Behavior*. 1979; 11:261–271.
- ED 201987. Johnston, P. Implications of basic research for the assessment of reading comprehension (Tech. Rep. No. 206). Washington, DC: National Institute of Education; 1981.
- *. Keenan JM, Betjemann RS, Olson RK. Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*. 2008; 12:281–300.
- *. Kerr MA, Symons SE. Computerized presentation of text: Effects on children's reading of informational material. *Reading and Writing*. 2006; 19(1):1–19.

- Kintsch W, van Dijk TA. Towards a model of text comprehension and production. *Psychological Review*. 1978; 85:363–394.
- 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]
- Klingner JK. Assessing reading comprehension. *Assessment for Effective Intervention*. 2004; 29(4): 59–70.
- * Kucer SB. Examining the relationship between text processing and text comprehension in fourth grade readers. *Reading Psychology*. 2009; 30:340–358.
- * Leon JA. The effects of headlines and summaries on news comprehension and recall. *Reading and Writing: An Interdisciplinary Journal*. 1997; 9:85–106.
- * Loyd BH, Steele JL. Assessment of reading comprehension: A comparison of constructs. *Reading Psychology: An International Quarterly*. 1986; 7:1–10.
- * Malicky GV, Norman CA. Reading processes subgroups in a clinical population. *The Alberta Journal of Educational Research*. 1988; 34:344–354.
- * 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]
- * Marr MB, Gormley K. Children's recall of familiar and unfamiliar text. *Reading Research Quarterly*. 1982; 18(1):89–104.
- * McGee LM. Awareness of text structure: Effects on children's recall of expository text. *Reading Research Quarterly*. 1982; 17:581–590.
- * McNamara DA, Kintsch E, Songer NB, Kintsch W. Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*. 1996; 14(1):1–43.
- McNeil MR, Doyle PJ, Fossett TRD, Park GH, Goda AJ. Reliability and concurrent validity of the information unit scoring metric for the story retelling procedure. *Aphasiology*. 2001; 15:991–1006.
- * Miller AC, Keenan JM. How word decoding skill impacts text memory: The centrality deficit and how domain knowledge can compensate. *Annals of Dyslexia*. 2009; 59:99–113. [PubMed: 19475514]
- Moss B. A qualitative assessment of first graders' retelling of expository text. *Reading Research and Instruction*. 1997; 37:1–13.
- * Naceur A, Schiefele U. Motivation and learning—The role of interest in construction of representation of text and long-term retention: Inter- and intraindividual analyses. *European Journal of Psychology of Education*. 2005; 20:155–170.
- Nicholas LE, Brookshire RH. A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech and Hearing Research*. 1993; 36:338–350. [PubMed: 8487525]
- Nilsson NL. A critical analysis of eight informal reading inventories. *The Reading Teacher*. 2008; 61:526–536.
- Olive T, Kellogg RT. Concurrent activation of high- and low-level production processes in written composition. *Memory & Cognition*. 2002; 30:594–600.
- * Otero J, Kintsch W. Failures to detect contradictions in a text: What readers believe versus what they read. *Psychological Science*. 1992; 3:229–235.
- * Pearman CJ. Independent reading of CD-Rom storybooks: Measuring comprehension with oral retellings. *The Reading Teacher*. 2008; 61:594–602.
- * Penning MJ, Raphael TE. The impact of language ability and text variables on sixth-grade students' comprehension. *Applied Psycholinguistics*. 1991; 12:397–417.
- Perfetti, CA. *Reading ability*. New York, NY: Oxford University Press; 1985.
- Peverly ST, Ramaswamy V, Brown C, Sumowski J, Alidoost M, Garner J. What predicts skill in lecture note taking? *Journal of Educational Psychology*. 2007; 99(1):167–180.
- * Pflaum SW. The predictability of oral reading behaviors on comprehension in learning disabled and normal readers. *Journal of Reading Behavior*. 1980; 12:231–236.

- Popplewell SR, Doty DE. Classroom instruction and reading comprehension: A comparison of one basal reader approach and the four-blocks framework. *Reading Psychology*. 2001; 22:83–94.
- *. Pretorius EJ. A profile of causal development amongst ten-year-old: Implications for reading and writing. *Reading and Writing: An Interdisciplinary Journal*. 1996; 8:385–406.
 - *. Rasinski TV. Investigating measures of reading fluency. *Educational Research Quarterly*. 1990; 14(3):36–44.
 - *. Richgels DJ, McGee LM, Lomax RG, Sheard C. Awareness of text structures: Effects on recall of expository text. *Reading Research Quarterly*. 1987; 22:177–196.
 - *. Riedel BW. The relation between DIBELS, reading comprehension, and vocabulary in urban first-grade students. *Reading Research Quarterly*. 2007; 42:546–567.
 - *. Roberts G, Good R, Corcoran S. Story retell: A fluency-based indicator of reading comprehension. *School Psychology Quarterly*. 2005; 20:304–317.
- Savage R. Reading comprehension is not always the product of nonsense word decoding and linguistic comprehension: Evidence from teenagers who are extremely poor readers. *Scientific Studies of Reading*. 2006; 10:143–164.
- Scardamalia, M.; Bereiter, C. Knowledge telling and knowledge transforming in written composition. In: Rosenberg, S., editor. *Advances in applied psycholinguistics, Vol. 1: Disorders of first-language development*. New York, NY: Cambridge University Press; 1987. p. 142–175.
- Schatschneider, C.; Buck, J.; Wagner, R.; Hassler, L.; Hecht, S.; Powell-Smith, K. A multivariate study of individual differences in performance on the reading portion of the Florida Comprehensive Assessment Test: A brief report (Tech. Rep. No. 5). Tallahassee, FL: Florida Center for Reading Research; 2004.
- Seifert TL. Brief research report: Enhancing memory for main ideas using elaborative interrogation. *Contemporary Educational Psychology*. 1994; 19:360–366.
- *. Shinn M, Good R, Knutson N, Tilly W, Collins V. Curriculum-based measurement reading fluency: A confirmatory factor analysis of its relation to reading. *School Psychology Review*. 1992; 21:459–479.
- Snyder L, Caccamise D, Wise B. The assessment of reading comprehension: Considerations and cautions. *Topics in Language Disorders*. 2005; 25(1):33–50.
- Spooner ALR, Baddeley AD, Gathercole SE. Can reading accuracy and comprehension be separated in the Neale Analysis of Reading Ability? *British Journal of Educational Psychology*. 2004; 74:187–204. [PubMed: 15130187]
- Stage SA, Jacobsen MD. Predicting student success on a state-mandated performance-based assessment using oral reading fluency. *School Psychology Review*. 2001; 30:407–419.
- Stecker PM, Fuchs LS. Effecting superior achievement using curriculum-based measurement: The importance of individual progress monitoring. *Learning Disability Research and Practice*. 2000; 15:128–134.
- Talbott E, Lloyd JW, Tankersley M. Effects of reading comprehension interventions for students with learning disabilities. *Learning Disability Quarterly*. 1994; 17:223–232.
- *. Taylor BM. Good and poor readers' recall of familiar and unfamiliar text. *Journal of Reading Behavior*. 1979; 11:375–380.
 - *. Tenenbaum AB. Task-dependent effects of organization and context upon comprehension of prose. *Journal of Educational Psychology*. 1977; 69:528–536.
 - *. Tindal G, Parker R. Development of written retell as a curriculum-based measure in secondary programs. *School Psychology Review*. 1989; 18:328–343.
- Valencia SW, Buly MR. Behind test scores: What struggling readers really need. *Reading Teacher*. 2004; 57:520–531.
- *. van den Broek P, Tzeng Y, Risdien K, Trabasso T, Basche P. Inferential questioning: Effects on comprehension of narrative texts as a function of grade and timing. *Journal of Educational Psychology*. 2001; 93:521–592.
 - *. 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.
- Wiley HI, Deno SL. Oral reading and maze measures as predictors of success for English learners on a state standards assessment. *Remedial and Special Education*. 2005; 26:207–214.
- Williams, F. Reasoning with statistics. New York, NY: Holt, Rinehart, and Winston; 1968.
- *. Williams JP. Comprehension by learning-disabled and nondisabled adolescents of personal/social problems presented in text. *American Journal of Psychology*. 1991; 104:563–586. [PubMed: 1793127]
- 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; 10:308–319.
- *. Yochum N. Children's learning from informational text: The relationship between prior knowledge and text structure. *Journal of Reading Behavior*. 1991; 23(1):87–103.
- *. Zinar S. Fifth-graders' recall of propositional content and causal relationships from expository prose. *Journal of Reading Behavior*. 1990; 22:181–199.

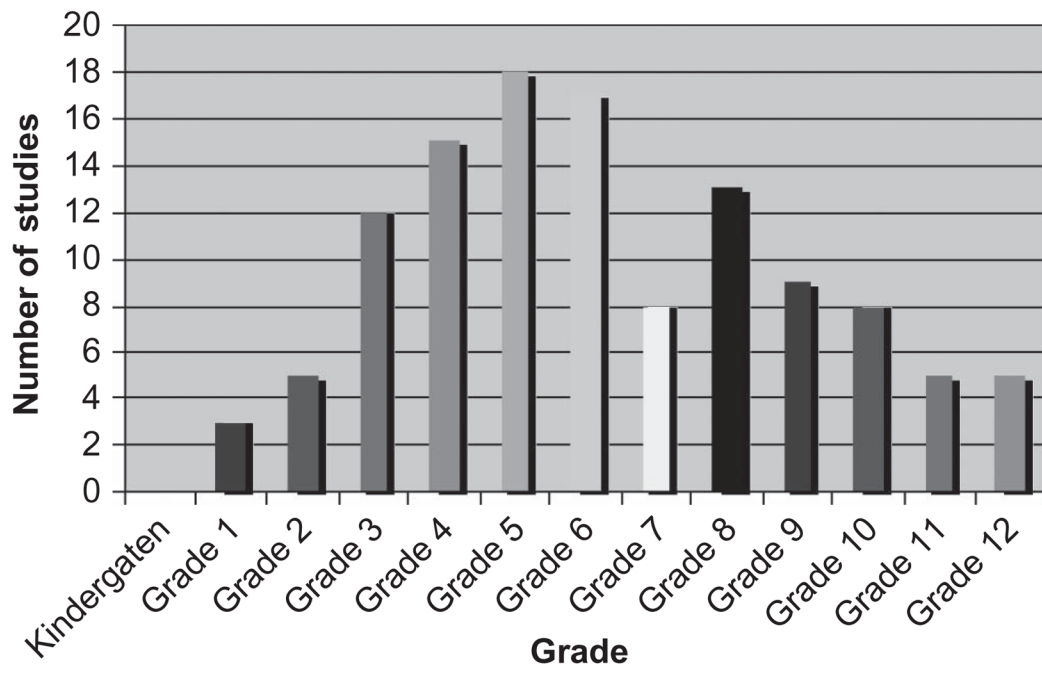


FIGURE 1.
 Number of studies in which each grade level was included.

TABLE 1
 Number of Studies Employing the Specified Combination of Text Type, Reading Format, and Retell Format

	Oral Reading	Silent Reading	Oral and Silent Reading	Oral or Silent Reading	Unknown
Oral Retell					
	Narrative text	6	3	1	1
	Expository text	4	6		4
	Both	2	1	1	
	Unknown	1			
Written Retell					
	Narrative text		2		
	Expository text		7		3
	Both				
	Unknown		2	1	
Both Oral and Written Retell				1	1
	Narrative text				
	Expository text				
	Both				
	Unknown				
Unknown if Retell Oral or Written				1	
	Narrative text				
	Expository text				2
	Both				
	Unknown				