



Language Matters: Measuring Reading Comprehension in Children with Oral Language Impairment



L'importance du langage : mesurer la compréhension de lecture chez les enfants ayant un trouble du langage oral

KEY WORDS
LANGUAGE DISORDERS
LITERACY
ASSESSMENT
CHILDREN

Melissa J. Skoczylas
 Phyllis Schneider
 Salima Suleman

Melissa J. Skoczylas,
 Faculty of Rehabilitation Medicine,
 University of Alberta,
 Edmonton, AB
 CANADA

Phyllis Schneider,
 Department of Communication
 Sciences and Disorders,
 University of Alberta,
 Edmonton, AB
 CANADA

Salima Suleman,
 Faculty of Rehabilitation Medicine
 University of Alberta,
 Edmonton, AB
 CANADA

Abstract

Purpose: This article presents a literature review focused on the association between oral language skills and performance on standardized reading comprehension. The investigators were particularly interested in the impact of atypical language skills among children with oral language impairment (LI) on test results and interpretation.

Method: A scoping review was undertaken to identify research focused on performance of school-aged children who have LI on standardized reading comprehension tests. Following initial searches, no literature was found to directly address the research question. The question was revised to include research comparing standardized reading comprehension measures, which included measurement of oral language skills. A search of 5 online databases was conducted, identifying 7 studies that compared reading comprehension tests based on participants' language and reading skills.

Conclusions: A gap in knowledge was identified regarding the use of standardized reading comprehension tests with children with LI. Three main conclusions were reached: (1) Further research is required documenting the relationship between the reading comprehension test scores and oral language skills in children with LI; (2) Selection of reading comprehension tests and oral language measures in research should be based on evaluation of participant characteristics and the purposes of assessment. The selection rationale should be stated in reports; (3) Reading comprehension is best represented by a profile of component skills, including various language skills, rather than by a single test score.

Abrégé

Objectif : Cet article présente une revue de la littérature portant sur l'association entre les habiletés de langage oral et la performance à des tests standardisés de compréhension de lecture. Les chercheurs se sont particulièrement intéressés à l'impact des habiletés langagières atypiques présentes chez les enfants avec un trouble du langage oral (TL) sur les résultats et l'interprétation des tests.

Méthodologie : Une revue exploratoire de la littérature a été effectuée afin d'identifier les articles portant sur la performance des enfants d'âge scolaire avec un TL à des tests standardisés de compréhension de lecture. À la suite d'une première recherche, aucune littérature répondant directement à la question de recherche n'a été trouvée. La question a donc été révisée pour inclure des études qui comparent des mesures standardisées de compréhension de lecture et qui incluent une mesure des habiletés langagières orales. La recherche a été effectuée dans 5 bases de données en ligne. Sept études comparant des tests de compréhension de lecture et s'appuyant sur les habiletés langagières orales et de lecture des participants ont été identifiées.

Conclusion : Un manque de connaissances a été identifié en ce qui concerne l'utilisation des tests standardisés de compréhension de lecture auprès des enfants avec un TL. Trois conclusions ont été formulées : (1) Des recherches supplémentaires sont nécessaires pour documenter la relation entre les résultats aux tests de compréhension de lecture et les habiletés langagières orales des enfants avec un TL; (2) Le choix des tests de compréhension de lecture et des mesures de langage oral utilisés en recherche devrait être basé sur l'évaluation des caractéristiques des participants ainsi que sur les buts de l'évaluation. La justification des choix devrait être énoncée dans les articles; (3) La meilleure manière de décrire la compréhension de lecture est de fournir un profil de plusieurs habiletés, ce qui inclut les diverses habiletés langagières, au lieu de présenter le résultat d'un seul test.

The development of strong literacy skills among all students is a key goal in our school systems today. Consequently, the ability to accurately identify and plan intervention for children at risk of academic difficulties is of interest to educators and speech-language pathologists working in the schools. Children with developmental oral language impairment (LI) frequently experience difficulties with literacy skills; early and ongoing language difficulties place children at risk for long-term reading difficulties (Nation & Norbury, 2005). Both decoding and comprehension issues appear to be common, though not universal, with LI. Given that children with LI are at increased risk of reading difficulties, it is important to have reading tests that accurately characterize the skills of such children. The measurement of reading comprehension is part of a complete reading assessment. It may be that commonly used standardized reading comprehension tests are less appropriate for children with developmental oral language impairment, due to the impact of the child's oral language skills on the tasks being measured. It may also be the case that differences in how they respond affect how their test results should be interpreted.

In this paper, the results of a scoping review regarding the impact of oral language skills on scores from standardized reading comprehension tests are presented. The review summarizes studies that compared results from various reading comprehension tests and that linked the test results to measures of oral language skills. Implications and recommendations for measuring reading comprehension for children with LI are discussed.

Exploring the Relationship between Language Impairment and Reading Comprehension

Characteristics of Language Impairment

Developmental oral language impairment can come in a variety of forms, including when the language difficulties are concomitant with conditions such as ADHD. The term "specific language impairment" (SLI) refers to developmental oral language difficulties with no known etiology (Leonard, 2000). A diagnosis of SLI frequently requires nonverbal IQ scores within the expected range for the child's age. In this paper, the term "language impairment" will be used to include children with SLI, as well as those with lower nonverbal IQ scores, and who are not identified as globally delayed. This latter group is sometimes described as "nonspecific language impairment". Cognitive referencing in labelling developmental language difficulties is currently thought to have limited usefulness, as IQ measures are inconsistently correlated with language skills (Dethorne & Watkins, 2006). Therefore the terms "specific"

and "nonspecific" are avoided here. The label "language impairment" is selected to describe children with language difficulties of unknown cause (see Bishop, 2014 for a full discussion of these issues).

Children with LI are a heterogeneous group. Difficulties may affect receptive language, expressive language, or both. Typically, the more language functions are affected, the higher the likelihood of the child experiencing reading difficulties (Simkin & Conti-Ramsden, 2006). Within receptive and expressive language, these children may have difficulty with phonology, semantics (including vocabulary), grammar, and/or usage (pragmatics) (Paul & Norbury, 2012). The language skills of individual children can be expected to impact the development of their reading skills, including reading comprehension.

Theories of Reading and Implications for Reading Comprehension Assessment

Theories of reading have varied in their consideration of language skills. A number of theories and models exist to describe the reading process, generally characterizing reading as the recognition of printed words and creation of a mental representation of what is read. A selection of models is briefly presented here to highlight how language abilities have been treated among reading models. The Spear-Sternberg Model of Reading Disability (Spear-Swerling & Sternberg, 1994) focuses largely on the development of automatic decoding; language skills beyond phonology are not specifically invoked until later stages. The model posits the normal path of reading acquisition as a single stepwise developmental sequence, working towards swift word recognition and eventually to the use of reading strategies. The model allows for categorization of reading disabilities based upon the stage at which a child has stalled in acquiring reading skills. In the Construction-Integration Model of Reading (Kintsch, 1998), basic language skills are primarily involved in deriving propositions from the text. Fluent adult reading is viewed as a set of processes that result in a mental representation of the text. This model recognizes the role of the reader's background knowledge and experience in comprehension, the ability to make inferences, and decoding abilities. The Lexical Quality Hypothesis (Perfetti, 2007) emphasizes the role of lexical knowledge, suggesting that the primary determinant of reading skill is the rapid and correct recognition of words, including "well-specified orthographic, phonological, and semantic-syntactic information" (p.211). Individual differences in reading skill, including comprehension, are attributed to differences in the quality of stored mental representations of words and the ability to activate them

quickly while reading. Based on this premise, reading difficulties would be expected in children with LI because the phonological and semantic-syntactic representations relate closely to the types of language skills typically affected in this population. The Integrated Model of Reading Comprehension (IMREC) (van den Broek and Espin, 2012) is presented as an effort to amalgamate current models of reading comprehension, with an eye toward developing more effective assessment tools. The work highlights the need to consider both automatic and strategic processing tasks and cognitive resources in the individual including such factors as language, working memory, ability to make inferences, background knowledge, and so on.

In general then, it appears to be accepted that language skills are involved in reading; however, not all models or theories clearly articulate how oral language skills are involved. When language is included in studies of reading, it is sometimes treated as only receptive language or comprehension. For example, much current research is founded in the Simple View of Reading (Gough & Tunmer, 1986), which has been validated by numerous studies. In short, this theory proposes that “reading equals the product of decoding and comprehension” (p. 7). Reading comprehension is essentially equated with listening comprehension when decoding is well-developed. This statement has important implications for research; the measurement tools typically used in oral language testing can then be used to evaluate comprehension. It also implies that expressive language need not be sampled in a reading assessment. It is possible that such measurement approaches may inadequately represent the skills of children with language impairment, particularly if expressive language is relatively more affected than receptive language.

No model currently considers all the subcomponents of language and their potential independent contribution to reading comprehension (but recall that IMREC does amalgamate current models of reading comprehension (van den Broek & Espin, 2012)). The specific model selected by a test developer or researcher typically determines whether or not oral language skills are examined and how they are measured.

The Role of Oral Language in Reading Comprehension

As noted in the previous section, oral language skills are typically assumed to function as part of reading comprehension processes. Current evidence suggests that both receptive and expressive language skills appear to be involved.

Receptive language. The role of receptive language in reading comprehension is not fully defined in the literature. The Simple View of Reading specifies “listening comprehension” as a key component of reading skills, yet this could be defined very differently depending on the researcher and/or test developer. Not only are various terms used to refer to receptive language skills (e.g. oral comprehension, listening comprehension), but how these constructs are operationalized can be quite variable (Keenan, Betjemann, & Olson, 2008; Ricketts, 2011). The term “receptive language” may construe a broader approach to the construct, whereas “listening comprehension” is often roughly equivalent to the understanding of spoken sentences and texts. It would be useful to know how these constructs are defined in research and whether there is disparity between studies.

A variety of schemes have been suggested to capture receptive language in the context of reading assessment. Carlisle (1991) laid out a method of assessing listening and reading comprehension and recommended screening measures such as “oral receptive ... vocabulary, syntax, and verbal memory for sentences and stories” (p. 36). However, Ekins and Schneider (2006) found that comprehension measures were understudied in the context of predicting reading skills. Among six studies reporting on reading comprehension measures, they found some evidence that receptive measures of semantics and syntax are predictive of reading comprehension scores. Ricketts (2011) reviewed the reading comprehension literature for children with LI using the Simple View as a framework. Ricketts recommended that the definition of “language comprehension” should be expanded to include specific sub-skills such as vocabulary.

Expressive language. Expressive language skills are typically not specifically invoked in reading theories. An interesting situation arises regarding children with LI when the Simple View of Reading is applied: since only receptive skills are considered in the model, there is no logical reason to measure expressive skills. If one applies this reasoning to children with receptive skills within the expected range but with expressive language difficulties, such children would not be considered to be at risk for reading disability (assuming adequate decoding skills). In contrast, a number of sources suggest the relevance of expressive language to reading. A study reported by Catts, Fey, Zhang, and Tomblin (1999) found that a broad range of oral language skills, including expressive language, was needed to account for variance in reading comprehension scores. Snyder, Caccamise, and Wise (2005) detailed the key points to evaluate when selecting a reading comprehension

instrument. They noted that expressive vocabulary and syntax are useful measures to include in reading comprehension screenings for older children. As with receptive language however, Ekins and Schneider (2006) found expressive language measures were understudied in the prediction of reading comprehension: only 5 of 13 studies examined such measures. From these studies, they found that the ability to provide definitions, narrative skills, and “standardized measures” of expressive language (p. 32) predicted reading comprehension.

The Impact of Language Impairment on Reading Comprehension Assessment

There are many psychometrically sound standardized tests of reading comprehension available, with various methods of measuring reading comprehension. It is possible that presenting the same material with different response tasks could result in different performances for children with LI. In other words, not only are language skills at play during reading comprehension processes, but atypical language skills may impact how children can respond to reading comprehension tests themselves. Researchers and educators need to know how test results may differ for children with LI relative to children with well-developed oral language abilities in order to best select and interpret reading comprehension tests.

Carlisle (1991) provided an overview of reading comprehension assessment tasks, comparing them for validity. Among five response tasks (multiple-choice, free recall, cloze, sentence verification, and picture identification), Carlisle noted various difficulties in task design including points relevant to the student with oral language weakness. She asserted that expressive language skills should be considered and noted the impact of the test questions on measurement results. For example, free recall requires a child to remember and repeat detailed information; as such it is likely to underestimate comprehension for children with expressive language deficits. Carlisle recommends the sentence verification task as a fair measurement for children with varying language skills, and one that can be used for listening or reading. She concludes her review by recommending that the task be carefully matched to the student. Similar conclusions were reported by Cain and Oakhill (2006) in a review focused on the *Neale Analysis of Reading Ability (NARA)* (Neale, 1989). These authors evaluated five response tasks (cloze, true/false sentence recognition, sentence verification, multiple choice, and open-ended questions) based on the literature. They commented on potential concerns with each, particularly noting the need to avoid ceiling effects. The authors pointed out that these tasks

may alter comprehension as they necessarily provide some information in the question. As an example, to measure *ability to make inferences* with true/false sentence recognition, the correct answer must be presented. The authors point out that the task is now to recognize an inference, not to make an inference. Similar concerns were presented for the sentence verification and multiple choice tasks. In this review, cloze tasks were noted to rely heavily on decoding skills. They noted that cloze tasks may also be more related to sentence- rather than text-level skills, unless properly designed. Other concerns with response tasks included high processing demands (e.g. multiple choice) and expressive language demands (open-ended questions). The authors concluded that these concerns could result in the misrepresentation of actual reading comprehension skills.

The impact of expressive deficits on the *NARA* was investigated by Spooner, Baddeley, and Gathercole (2004). The researchers compared a group of children who had age-appropriate decoding skills and weak reading comprehension on the *NARA* to a group who had both decoding and reading comprehension within the expected range. When the response task was switched from question response to forced choice, they found that the apparent reading comprehension differences between the groups disappeared. They suggested that the differences were reflective of tasks demands; children with reduced expressive ability were underestimated in terms of the amount they had actually comprehended, the effect predicted by Carlisle (1991) as noted above.

It seems likely that other test characteristics, such as text type and length, may affect the performance of children with LI differently than children with typically-developing language skills. Knowledge of such effects would be important in matching test to student in practice and research.

Research Questions

In the preceding discussion of the literature, a number of research questions were identified that we will attempt to answer in the following review of the literature:

1. Results for children with developmental oral language impairment: What is known regarding the appropriateness of standardized reading comprehension tests for children with developmental oral language impairment?
2. Comparability among reading comprehension tests: Do different tests yield comparable results?
3. Impact of response task on performance: What is

the impact of assessment task on the student with oral language impairment?

4. Definition and measurement of oral language skills:
 - a. How is "listening comprehension" defined and operationalized in studies of reading comprehension measures?
 - b. What is known about the role of expressive language skills in the measurement of reading comprehension?
 - c. Do researchers provide clear selection rationale for oral language tasks?
5. Aspects of tests to consider: What are the characteristics of reading comprehension tests that may affect the performance of children with LI?

Methods

The scoping review methodology employed was described by Arksey and O'Malley (2005). The scoping format suited the purpose of the review: to identify existing literature on this topic. Five phases were followed: "identifying the research question"; "identifying relevant studies"; "study selection"; "charting the data"; "collating, summarizing", and "reporting the results" (Arksey & O'Malley, 2005, pp. 8-9). The review proceeded in an iterative process. Initially, the broad research question was: What is known regarding the appropriateness of standardized reading comprehension tests for children with developmental oral language impairment? This was refined into the current set of questions as relevant papers were identified. At each step, the accumulated literature was reviewed to search for answers to the questions that arose. The review was complete when no further questions arose and all existing questions were answered based on the literature that met the inclusion criteria. A researcher who was not involved in the initial review was recruited to review the inclusions list (Levac, Colquhoun, & O'Brien, 2010). Full agreement on inclusions was reached with discussion. A third researcher performed a reliability check on the results for two of the articles. This researcher answered the stated research questions relative to those two articles and the results were compared with the original findings; full agreement on the answers was achieved.

Since the purpose of the review was to inform future research, grey literature (i.e., material not published in journals or other commercial sources (Dijkers, Murphy & Krellman, 2012)) was not included in the search as would be typical for a scoping review. In addition, although scoping reviews do not typically include critical evaluation of articles,

this process was employed in the interests of planning future studies.

A university librarian was consulted to determine the best databases and search terms to identify the literature of interest. Five databases were searched: Scopus, Web of Science, ComDisDome, ERIC, and ProQuest Dissertations and Theses. The search terms used were: "reading comprehension"; measur* OR assess* OR diagnos* OR identif*; "language impair*" OR "impairments in language" OR "language disorder*" OR "language delay*" OR "language deficit*"; child* OR school-age* OR school age* OR elementary NOT deaf* OR aphasia. The search of the 5 databases returned a combined list of 113 papers once duplicates were removed. At this time, no papers were found that directly addressed the primary research question. Inclusion criteria were amended to include literature that could shed light on the impact of oral language skills on reading comprehension test scores. Criteria were revised as follows: papers included measures of oral language and either directly compared measures of reading comprehension or evaluated the use of reading comprehension tests relative to oral language performance. Other inclusion criteria were unchanged: only papers including school-age participants were considered; if papers specified a disorder type, only those evaluating developmental oral language impairment were considered. Only English language papers were considered. The review continued with these criteria in place. An additional 17 papers for consideration were found by hand searching the references of relevant papers from the database search. This total list of 130 resulted in 6 inclusions. A Web of Science citation search of the list of 6 papers provided an additional 12 papers for consideration; this resulted in 1 additional inclusion for a total of 7 studies that met criteria. The literature search was completed by October, 2013. A total of six papers and one book chapter were identified as meeting the criteria of this review; each compared reading comprehension measures and included measures of oral language.

Results

The following table summarizes the seven studies that met inclusion criteria for this review. Next, the studies will be discussed in five sections that follow the research questions: results for children with LI; comparability among reading comprehension measures; impact of task on performance; definition and measurement of oral language skills; and aspects of tests to consider. Key details pertaining to each research question are elaborated in the section specific to that question.

Table 1. Summary of studies meeting inclusion criteria

Authors	Year and Location	Age Range	Number of participants	Number of children with LI included	Reading comprehension tests studied
Nation & Snowling	1997 UK	7-10 years	184	Not reported	<i>Neale Analysis of Reading Ability (NARA)</i> ; <i>Suffolk Reading Scale</i>
	Main results	Using hierarchical regression, more score variance in the NARA results was accounted for by listening comprehension scores than by decoding; the <i>Suffolk</i> results were better predicted by decoding skills than by listening comprehension.			
Nation & Snowling	1997 UK	7-10 years	Good comprehenders (N=17); poor comprehenders (N=17)	Not reported; expected overlap with "poor comprehenders"	<i>Neale Analysis of Reading Ability</i> ; <i>Suffolk Reading Scale</i>
	Main results	As task complexity increased from single word to text reading, poor comprehenders did increasingly poorly compared to the good comprehenders. This effect was eliminated on the <i>Suffolk</i> results when single-word reading was used as a control, but not for the <i>NARA</i> .			
Spear-Swerling	2004 USA	Mean age 9 years 8 months (4 th graders)	95	Not reported	<i>Connecticut Mastery Test -Degrees of Reading Power (DRP)</i> ; <i>Connecticut Mastery Test Reading Comprehension (RC)</i>
	Main results	In a hierarchical regression analysis, decoding skill explained more variance for cloze (DRP) than for question response (RC). Listening comprehension (measured as oral cloze) was nearly equally predictive for both tests.			
Francis, Fletcher, Catts & Tomblin (analysis 1)	2005 USA	Grade 1 and 2	945	Not reported	<i>Woodcock-Johnson Revised (WJRPC)</i> , <i>Formal Reading Inventory</i>
	Main results	Results of these tests correlated differently to measures of oral and written vocabulary and to decoding measures. Although all correlations were significant, correlations with all measures were higher overall with the <i>WJRPC</i> .			
Francis, Fletcher, Catts & Tomblin (analysis 2)	2005 USA	Grade 2 and 4	570	Not reported	<i>Woodcock Reading Mastery Test Passage Comprehension (WRMTPC)</i> ; <i>Diagnostic Assessment Battery</i> , <i>Gray Oral Reading Test</i>
	Main results	Correlations between receptive language and reading comprehension scores were similar for the 3 measures but correlations with decoding varied significantly. For all 3 measures, correlations with decoding were stronger in Grade 2 than in Grade 4; in both grades, correlations were highest to the <i>WRMTPC</i> .			

Cutting & Scarborough	2006 USA	7 to 15 years	97	Not reported	<i>Gates–MacGinitie Reading Test–Revised (G–M); Gray Oral Reading Test–Third Edition (GORT-3); Wechsler Individual Achievement Test</i>
	Main results	In a hierarchical multiple regression analysis, the maximum variance accounted for by oral language skill and decoding together was for the <i>G–M</i> . The <i>GORT-3</i> had the lowest amount of variance accounted for. The instruments differentially tapped decoding skills and oral language skills, with both being useful predictors in regression models. The model for the <i>G–M</i> included both lexical and sentence processing composites; the <i>WIAT</i> was only predicted by the sentence composite.			
Keenan, Betjemann & Olson	2008 USA	8 to 18 years	510	Not reported; requirement for full scale IQ >85 suggests some children with LI could be excluded	<i>Woodcock-Johnson Passage Comprehension (WJ-PC); Qualitative Reading Inventory (QRI); Gray Oral Reading Test (GORT); Peabody Individual Achievement Test (PIAT-RC)</i>
	Main results	Via hierarchical regression, decoding skill explained more variance for the <i>PIAT-RC</i> and <i>WJ-PC</i> than the other two reading comprehension measures in the study. Conversely, listening comprehension was more strongly related to the scores from the <i>GORT</i> and the <i>QRI</i> .			
Kendeou, Papadopoulos & Spanoudis	2012 Cyprus	Mean age 6 years, 6 months; 7 years, 7 months	286	Children with history of speech, language or hearing problems were excluded	<i>Woodcock-Johnson Passage Comprehension (WJIII-PC), Recall task, CBM-Maze</i>
	Main results	The <i>WJIII-PC</i> was most linked to “orthographic processing and working memory skills” (p. 363) as was the recall task with the addition of phonological skills. The CBM maze task depended most upon vocabulary and fluency.			
Eason, Goldberg, Young, Geist & Cutting	2012 USA	10 to 14 years	126	Not reported; included children with reading disabilities	Stanford Diagnostic Reading Test –4 th Ed.
	Main results	Repeated measures analysis of variance revealed an interaction between question type and text type. Functional text yielded similar results for the 3 questions types. Narrative and expository texts were similar in percent correct answers across question types; however they differed when specific question types were considered. The critical analysis and process strategies questions were most difficult. Regression analyses revealed that word recognition and receptive vocabulary significantly predicted reading comprehension for all 3 text types. Comprehension of expository texts was also predicted by inferencing.			

Results for Children with Oral Language Impairment: What is known regarding the appropriateness of standardized reading comprehension tests for children with developmental oral language impairment?

As can be seen from the preceding table, none of the included studies specifically evaluated results for children with identified LI. Note that in some cases, children with LI were included in the study samples. Implications for this population that may be extended from the included studies are explored in the discussion.

Comparability among Reading Comprehension Tests: Do different tests yield comparable results?

This relatively small body of literature is strikingly consistent in the primary findings of the studies: reading comprehension measures are not interchangeable. As will be further explored in the following sections, each research

team took a different approach in comparing reading comprehension measures. Yet in each case, they found significant differences among the instruments. For both research and academic use, instrument selection must be an intentional process; for purposes of comparing between studies, researchers should clearly articulate the reasons for the instrument selections made.

Impact of Response Task on Performance: What is the impact of assessment task on the student with oral language impairment?

Response task was noted as one significant source of variation between instruments. The tests examined in the studies reviewed here are summarized in Table 2 in terms of the tests' characteristics and the type of response task for each; in some cases categories were unclear due to limited information in the reports.

Table 2. Summary of Reading Comprehension Tests Used in Comparison Studies

Test	Passage Length	Oral/Silent Reading	Genre of Passage	Response Task
<i>Suffolk Reading Scale</i>	Sentences	Not reported	Not reported	Cloze
<i>Woodcock Johnson Passage Comprehension</i>	Sentences, 2-3 lines	Silent	Not reported	
<i>Woodcock Reading Mastery Test Passage Comprehension</i>	Not reported	Not reported	Not reported	
<i>Connecticut Mastery Test --cloze</i>	Not reported	Not reported	Not reported	
<i>Connecticut Mastery Test --question response</i>	Not reported	Not reported	Not reported	Question response
<i>Formal Reading Inventory: Silent Reading Score</i>	"graded passages"	Silent	Not reported	
<i>Diagnostic Assessment Battery--2</i>	Not reported	Silent	Not reported	
<i>Neale Analysis of Reading Ability</i>	"short"	Aloud	Narrative	
<i>Wechsler Individual Achievement Test</i>	2-3 sentences	Silent	Narrative and expository	
<i>Qualitative Reading Inventory</i>	250-785 words	Aloud	Narrative and expository	<i>QRI also has recall</i>

Researcher-created recall task (Kendeou et al., 2012)	177 words	Not reported	Narrative	Recall
<i>Gates-McGinitie Reading Test</i>	3-15 sentences	Silent	Narrative and expository	Written multiple choice
<i>Gray Oral Reading Test</i>	6-7 sentences	Not reported	Narrative and expository	Spoken multiple choice
<i>Stanford Diagnostic Reading Test—4th Edition</i>	Not reported (timed)	Not reported	Narrative, expository and functional	Multiple choice (not reported as written/spoken)
<i>Peabody Individual Achievement Test—Reading comprehension</i>	Not reported	Silent	Not reported	Picture pointing
<i>CBM Maze</i> (Kendeou et al., 2012)	155-183 words (time limited)	Not reported	Narrative	Maze

Cloze and question-response were the most common response formats evaluated. In general, cloze formats were found to be more related to decoding-level skills than were open-ended questions (Keenan et al., 2008; Kendeou, Papadopoulos, & Spanoudis, 2012; Nation & Snowling, 1997; Spear-Swerling, 2004). The interpretation of task results is complicated by the lack of a consistent passage length (Keenan et al., 2008). For example, Spear-Swerling (2004) reported results that were slightly at odds with Nation and Snowling (1997) regarding the relevance of listening comprehension to cloze tasks. The cloze task in Spear-Swerling's study involved passages, whereas the *Suffolk* cloze task in Nation and Snowling (1997) used single sentences. It is possible that the additional information in the passages allowed for higher-level processes to come into play, whereas the sentences comprehension was more constrained to information gained from lower-level processing, including decoding (Spear-Swerling, 2004). Similarly, a picture-pointing task found on the *PIAT*, on which correct answers depend upon recognition of key vocabulary words in short texts, was found to be linked to decoding ability (Keenan et al., 2008).

Cutting and Scarborough (2006) evaluated three tests with various question-response formats (*GORT-3*, *WIAT*, and *G-M*). They found that the three tests they evaluated were differentially related to the language composites they created, measuring "lexical" and "sentence processing" skills, as well as to decoding skills. The results from the *WIAT* were the most closely related to decoding skills; the *G-M* was most related to the oral language composites used. The

WIAT is a question-response task, whereas the *G-M* uses a written multiple choice format. The authors note that even when the measures compared are constrained to those using question-response tasks as opposed to including measures using cloze procedures, significant differences between measures were found, implicating other test characteristics in affecting results.

The study by Eason, Goldberg, Young, Geist, and Cutting (2012) directly examined the impact of task and text on reading comprehension accuracy. This team used a single reading comprehension instrument with one response type (multiple choice) but with 3 levels of complexity in the questions and three different text types. The specific results of their analyses are summarized in Table 1; in short, both question type and text type had significant impact on reading comprehension accuracy for the participants.

Definition and Measurement of Oral Language Skills

How is "listening comprehension" defined and operationalized in studies of reading comprehension measures? There was significant variability between studies in the definition and measurement of oral language skills. Oral language comprehension or "listening comprehension" was the construct of interest in four studies (Francis, Fletcher, Catts, & Tomblin, 2005; Keenan et al., 2008; Nation & Snowling, 1997; Spear-Swerling, 2004). Measures of oral language comprehension ranged from vocabulary measures to inferential questions based on passages. Cutting and Scarborough (2006) took a different approach; rather than focus on listening comprehension, they used a

variety of measures to develop two language composites. This was interesting given their reference to the Simple View of Reading, since their composites included a number of expressive tasks. As discussed earlier, the Simple View refers directly to comprehension tasks only. The language tasks employed by Eason et al. (2012) were receptive in nature, tapping vocabulary, morpho-syntactic awareness, and ability to make inferences. The cognitive processing measures used by Kendeou and colleagues (2012) included several language-loaded tasks that tapped receptive vocabulary and sentence comprehension including syntax. The definition and measurement of oral language in the studies are summarized in Table 3.

What is known about the role of expressive language skills in the measurement of reading comprehension?

Expressive language was not focused upon in any of the studies. As noted, the measures employed by Cutting and Scarborough (2006) did include expressive scores, but since composites were reported, no specific conclusions regarding expressive skills could be made.

Do researchers provide clear selection rationale for oral language tasks? Again there was variability among the studies. Francis et al. (2005), Keenan et al. (2008), and Nation and Snowling (1997) stated the need to measure oral language comprehension without further elaboration. Spear-Swerling (2004) provided discussion regarding the necessity of sampling vocabulary as well as language comprehension. Nation and Snowling (1997) provided a brief description of their listening task without information on the development of the stories or questions. Cutting and Scarborough (2006) presented research evidence for the importance of vocabulary and syntax in reading comprehension as rationale for their variable selection. Eason et al. (2012) explained their study motivation in terms of fusing the developmental perspective of the Simple View of reading and the Construction-Integration Model of Reading to lead to the inclusion of basic oral language and inferencing variables. Kendeou et al. (2012) laid out in detail the theoretical bases and research evidence underlying their test and variable selections.

Aspects of Tests to Consider

The impact of response task on reading comprehension test results has been considered, but other test characteristics may impact results. Eason et al., (2012) directly investigated the impact of text genre. This group found an interaction between text type and question type, with expository text, not narrative, tapping inferencing ability. Numerous researchers commented on other aspects of tests that could impact results such as reading

tasks. Keenan and colleagues (2008) purposely selected their assessments to vary in terms of reading task, including oral/silent reading, and passage length. They concluded that passage length was significant in explaining their results; shorter passages appeared to measure decoding more than comprehension.

Discussion

This review has identified a gap in the literature on the use of standardized reading comprehension tests with children with LI. Although a number of studies were found examining the link between oral language skills and scores on reading comprehension tests, no studies specifically examined the interaction of reading comprehension tests with atypical language skills among students with LI. It is important to note that the samples of at least some of the studies reviewed included such children; however, results for these children were not separately explored in these studies to see whether the patterns of results were similar to those of children without LI. Post hoc categories based on discrepancy between reading comprehension and decoding as identified by Nation and Snowling (1997) can give some information regarding the population of children with LI. However, more focused results based upon the performance of children with identified oral language difficulties may result in different conclusions.

Implications for Children with Language Impairment

Based on the results of the included studies, some implications for children with LI and questions for future study are suggested by the authors of this review. In every study, the oral language measures selected accounted for different patterns of variance in reading comprehension scores. Overall, results suggest that tests that are more closely linked to listening comprehension deficits may reveal different weaknesses in students with LI than those that primarily depend upon decoding.

The work of Eason and colleagues (2012) highlights the importance of vocabulary across text genres and question complexity. Given that children with LI frequently have impoverished vocabularies (Paul & Norbury, 2012), their results suggest that it may be clinically useful to augment reading comprehension tests with vocabulary assessment in order to tease out a possible source of reading comprehension weakness. They also found inferencing to predict scores for expository text, but not narrative text. Since children with LI may be limited in inferencing skill (Barnes, Johnston, & Dennis, 2007), a child's reading comprehension might be overestimated by assessment tools that include only narrative text.

Table 3. Definition and Measurement of Oral Language Skills

Authors and year	Theoretical basis	Language construct	Operational measures
Nation & Snowling, 1997	Simple View of Reading	Listening comprehension	Orally presented stories with direct and inferential questions
Spear-Swerling, 2004	Spear-Sternberg Model of RD	Listening comprehension	LC subtest of the <i>Woodcock-Johnson Psychoeducational Battery</i> : an oral cloze task; <i>PPVT III</i>
Keenan et al., 2008	Not directly stated; consistent with Simple View of Reading	Listening comprehension	LC subtest of the <i>Woodcock-Johnson Psychoeducational Battery</i> : an oral cloze task; passage task from <i>Qualitative Reading Inventory</i> with retell and questions; <i>KNOW-IT Test</i> : a passage task with the teaching of background knowledge
Francis et al., 2005	Not directly stated	Receptive language	<i>WISC-R</i> Vocabulary score in one analysis, and a “receptive language” composite in the other (receptive language tests used were not specified)
Cutting & Scarborough, 2004	Simple View of Reading	Oral language (divided into lexical skills and sentence processing)	Lexical Composite: <i>PPVT-III</i> and the <i>Boston Naming Test</i> [expressive vocabulary]; Word Classes subtest <i>CELF-3</i> ; [semantic relations] Sentence processing composite: <i>CELF-3</i> subtests: Concepts and Directions, Formulated Sentences, Recalling Sentences; also a complex sentence comprehension task Verbal memory: Immediate Recall subtest of the <i>Wide Range Assessment of Memory and Learning</i> [story recall]; and a “nonstandardized sentence span measure”
Eason et al., 2012	Simple View and Construction-Integration Model of reading	Oral language (only receptive skills sampled) and inferencing	<i>PPVT-III</i> [“semantic awareness”]; Grammatical Comprehension subtest of the <i>Test of Language Development—Intermediate (3rd Ed.)</i> : sentence-level grammaticality judgement task [morpho-syntactic awareness]; Making Inferences subtest of the <i>Test of Language Competence—Expanded Edition</i> : ability to make inferences based on a passage
Kendeou et al., 2012	Verbal Efficiency Theory	Cognitive processing (rather than language per se)	<i>Dyslexia Early Screening Test</i> [receptive vocabulary]; verbal spatial relations task: sentence-level comprehension task with picture-pointing response [working memory; also preposition comprehension]; sentence repetition: repeat nonsense sentences and answer questions about them [working memory, syntax]

PPVT III: Peabody Picture Vocabulary Test III; a word-level picture pointing task of receptive vocabulary

CELF 3: Clinical Evaluation of Language Fundamentals, Third Edition

LC: listening comprehension

Keenan et al. (2008) found that a given test may measure different skills depending upon the age or decoding skill of the participant. This suggests that further work comparing reading comprehension measures for children with LI may reveal additional variability in test score interpretation, as this group can be expected to include a relatively higher proportion of children with reading problems than the general population and are by definition less skilled in language. Conversely, Cutting and Scarborough (2006) did not find such effects, although they acknowledged that this was not consistent with previous findings.

Kendeou et al. (2012) excluded students with a history of speech or language impairments. It would be interesting to find out whether results regarding cognitive processing demands would be similar with this population. For example, since many children with LI have working memory limitations (Archibald, 2006), these children might be expected to perform more poorly on instruments that tax this resource, such as recall tasks. Note that this would not invalidate the test results, as working memory is generally considered to be a true sub-skill required for reading comprehension. The point is that it would be important to know how children with LI would score across tests for optimal test selection and interpretation. The results obtained by Kendeou et al. (2012) were not consistent with Cutting and Scarborough (2006) who reported that “the prediction of comprehension scores was not enhanced by taking into account any measure of verbal memory, rapid serial naming, IQ, or (with one minor exception) attention” (p. 294). Note that the studies differed in terms of study population, working memory measures, and reading comprehension tests evaluated.

Definition and Measurement of Oral Language Skills

Language measures employed in the studies described here were primarily receptive measures, attributable to the use of the Simple View of Reading framework (Gough & Tunmer, 1986). It may be that the Simple View applies well to children with language systems developing within the expected range, but less well to children with LI who may have inconsistent abilities across language domains. If children with typically developing language skills have relatively more consistent language skills, any one language measure will be more representative of their overall language profiles. For children with LI and potentially with more inconsistent language profiles, a broader selection of language measures may be necessary to appropriately characterize their skills. It has been suggested that the Simple View could be enhanced by more clearly

specifying the language comprehension and cognitive skills that it encompasses (Ricketts, 2011). In addition, the preponderance of receptive measures may limit studies from revealing the impact of expressive language problems on reading comprehension. The study by Spooner et al. (2004) suggested that poor expressive language skills can limit children’s ability to demonstrate their understanding of a text. Simkin and Conti-Ramsden (2006) reported that children’s reading skills varied with their language skills: children with combined receptive and expressive deficits had the most difficulty, but a majority of children with expressive-only deficits also had reading problems. The question remains whether poor expressive language skills are a primary factor in children’s responses in assessment tasks, or whether expressive language represents an underlying skill of reading comprehension itself. Studies comparing results for various reading comprehension tests for children with identified oral language impairments in both receptive and expressive domains will enhance our ability to select appropriate measures for our purposes. This in turn should lead to more effectively targeted intervention programs based upon accurate assessments.

It was interesting to note that the rationale for selection of oral language measures was not consistently reported in these studies. Recall that Kendeou et al. (2012) laid out the research evidence and rationale for their variable selections in detail. This would seem to be sound practice which facilitates the interpretation of results across studies and for practical application.

Aspects of Tests to Consider

As discussed above, reading comprehension tests cannot be considered equivalent. This could be a benefit, given that there are a variety of purposes served by these tests. Selecting the best instrument for one’s purpose then will require careful evaluation of relevant factors. All of the studies in this review included some discussion of test characteristics as potential sources of variation, although they were not necessarily studied as separate variables. In some cases it may be challenging to separate the effects of different test characteristics. For example, it is difficult to disambiguate reading task from response task for cloze presentations as the response item is embedded in the stimulus text. A clear understanding of test characteristics and the resulting ability to select measures for a given purpose has direct relevance to student outcomes. Cutting and Scarborough (2006) discussed the importance of test selection in the identification of reading comprehension problems. They reported the results of a study with a portion of their sample demonstrating that application

of the tests they studied would yield different diagnostic results (Rimrod, Lightman, Roberts, Denckla, & Cutting, 2005). The tables provided in Appendix A summarize variables that may impact test results, as identified in the literature presented above and by task analysis performed by the first author.

Measuring Reading Comprehension in Practice

Based upon the articles reviewed here, it appears that compiling a profile of language and reading skills is an appropriate method to document reading comprehension. Variety in both test formats and tasks sampling component skills must be employed (Spear-Swerling, 2004). Particularly for a child with LI, intervention planning requires a thorough understanding of the learning profile, including receptive and expressive language skills in semantics, syntax, vocabulary and discourse level skills, as well as working memory, reading comprehension with authentic texts, and decoding skills. The most parsimonious method to compile a learning profile for the child might be to coordinate among teacher, speech-language pathologist, and school psychologist. Among them, it is likely that many, if not all, of the measures needed may be readily available. Pooling this knowledge will permit the development of a language and literacy profile that could provide a rich basis for intervention planning.

Conclusions

- 1) Further research needs to be done to document the relationship of oral language skills in children with oral language impairment, particularly expressive skills, to reading comprehension. In particular, comparison of results among standardized reading comprehension tests for children with oral language impairment will facilitate interpretation of the results from these measures.
- 2) Selection of reading comprehension measures in studies of language and reading should be based upon criteria such as those detailed in Snyder and colleagues (2005). This selection rationale should be explicitly stated to facilitate interpretation of results and cross-study comparisons. Oral language measures should be similarly selected and specified for the same reasons.
- 3) Reading comprehension is best represented as a profile of component skills including a variety of oral language skills. No single available instrument appears to document all the applicable skills; measurement selection will vary by the purpose of

the assessment. The necessary comprehensive assessment will be facilitated by inter-professional collaboration among teachers, speech-language pathologists, and psychologists to provide the broadest overview of skills with the least duplication of assessment.

The studies reviewed here comprise a small but remarkably consistent body of literature demonstrating that reading comprehension tests sample different underlying skills. This presents a challenge: test users will need familiarity with the specific skills tapped by the measure in use to best match test to the student and the purpose of assessment. Given that not all reading comprehension tests have been studied, further research in this area should be fruitful in clarifying the skills tapped by these tests.

References

- Archibald, L. M. D., (2006). "Short-term and working memory in children with specific language impairment (Doctoral Dissertation, Western University). Retrieved from <http://ir.lib.uwo.ca/scsdpub/1>
- Arksey, H., & O'Malley, L. (2005) Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19-32.
- Barnes, M. A., Johnston, A. M., & Dennis, M. (2007). Comprehension in a neurodevelopmental disorder, spina bifida myelomeningocele. In Cain, K. & Oakhill, J. (Eds.), *Children's comprehension problems in oral and written language* (pp. 193-217). New York, NY: Guilford Press
- Bishop, D. V. M. (2014). Ten questions about terminology for children with unexplained language problems. *International Journal of Language and Communication Disorders*, 49(4), 381-415.
- Cain, K., & Oakhill, J. (2006). Assessment matters: Issues in the measurement of reading comprehension. *British Journal of Educational Psychology*, 76, 697-708. doi:10.1348/000709905X69807
- Carlisle, J. F. (1991). Planning an assessment of listening and reading comprehension. *Topics in Language Disorders*, 12(1), 17-31.
- Catts, H., Fey, M., Zhang, X., & Tomblin, J. (1999) Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. *Scientific Studies of Reading*, 3, 331-344.
- Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading*, 10, 277-299. doi:10.1207/s1532799xssr1003_5
- Dethorne, L. S., & Watkins, R. V. (2006). Language abilities and nonverbal IQ in children with language impairment: Inconsistency across measures. *Clinical Linguistics & Phonetics*, 20(9), 641-658.
- Dijkers, M. P., Murphy, S. L. & Krellman, J. (2012). Evidence-based practice for rehabilitation professionals: Concepts and controversies. *Archives of Physical Medicine and Rehabilitation*, 93(Supplement 2), S164-S176.
- Eason, S. H., Goldberg, L. F., Young, K. M., Geist M. C., & Cutting, L.E. (2012). Reader-text interactions: How differential text and question types influence cognitive skills needed for reading comprehension. *Journal of Educational Psychology*, 104(3), 515-528.
- Ekins, E., & Schneider, P. (2006). Predicting reading abilities from oral language skills: A critical review of the literature. *Journal of Speech-Language Pathology and Audiology*, 30, 26-36.

- Francis, D., Fletcher, J., Catts, H., & Tomblin, J. (2005). Dimensions affecting the assessment of reading comprehension. In S. Parks & S. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 369-394). Mahwah, NJ: Lawrence Erlbaum Associates.
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*, 6-10.
- Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading, 12*(3), 281-300. doi:10.1080/10888430802132279.
- Kendeou, P., Papadopoulos, T. C., & Spanoudis, G. (2012). Processing demands of reading comprehension tests in young readers. *Learning and Instruction, 22*(5), 354-367. doi:10.1016/j.learninstruc.2012.02.001.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, UK: Cambridge University Press.
- Leonard, L.B. (2000). *Children with specific language impairment*. Cambridge, MA: MIT Press
- Levac, D., Colquhoun, H., & O'Brien, K. (2010) Scoping studies: Advancing the methodology. *Implementation Science, 5*(69). doi:10.1186/1748-5908-5-69.
- Nation, K., & Norbury, C. F. (2005). Why reading comprehension fails: Insights from developmental disorders. *Topics in Language Disorders, 25*(1), 21-32.
- Nation, K., & Snowling, M. (1997). Assessing reading difficulties: The validity and utility of current measures of reading skill. *British Journal of Educational Psychology, 67*(3), 359-370.
- Neale, M. D. (1989). *The Neale Analysis of Reading Ability-Revised*. Windsor: NFER.
- Paul, R & Norbury, C.F. (2012). *Language disorders from infancy through adolescence: Listening, speaking, reading, writing, and communicating*. St. Louis, Missouri: Elsevier Mosby
- Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*(4), 357-383.
- Ricketts, J. (2011). Research review: Reading comprehension in developmental disorders of language and communication. *Journal of Child Psychology and Psychiatry, 52*(11), 1111-1123.
- Rimrod, S., Lightman, A., Roberts, L., Denckla, M. B., & Cutting, L. E. (2005, February). *Are all tests of reading comprehension the same?* Poster presented at the annual meeting of the International Neuropsychological Society, St. Louis, MO.
- Simkin, Z., & Conti-Ramsden, G. (2006). Evidence of reading difficulty in subgroups of children with specific language impairment. *Child Language Teaching and Therapy, 22*(3), 315-331.
- Snyder, L., Caccamise, D., & Wise, B. (2005). The assessment of reading comprehension- Considerations and cautions. *Topics in Language Disorders, 25*(1), 33-50.
- Spear-Swerling, L. (2004). Fourth graders' performance on a state-mandated assessment involving two different measures of reading comprehension. *Reading Psychology, 25*, 121-148. doi:10.1080/02702710490435727
- Spear-Swerling, L., & Sternberg, R.J. (1994). The road not taken: An integrative theoretical model of reading disability. *Journal of Learning Disabilities, 27*(2), 91-103.
- Spooner, A. L., Baddeley, A. D., & Gathercole, S. E. (2004). Can reading accuracy and comprehension be separated in the Neale Analysis of Reading Ability? *British Journal of Educational Psychology, 74*(pt. 2), 187-204.
- van den Broek, P., & Espin, C.A. (2012). Connecting cognitive theory and assessment: Measuring individual differences in reading comprehension. *School Psychology Review, 41*(3), 315-325.

Acknowledgements

The authors would like to thank Camille Dube for her assistance with the reliability procedures for this review.

This work was supported by the Social Sciences and Humanities Research Council of Canada.

Authors' Note

Correspondence concerning this article should be addressed to Melissa Skoczylas, Department of Communication Sciences and Disorders, 2-70 Corbett Hall, University of Alberta, Edmonton, Alberta, Canada, T6G 2G4
Email: mskoczyl@ualberta.ca.

Appendix A Aspects of Tests to Consider

Stimulus Text

- text genre: expository, narrative, poetry, mixed (Eason et al., 2012)
- length of text (Cutting & Scarborough, 2006; Keenan et al., 2008; Kendeou et al., 2012)
- syntactic complexity and language forms (Cutting & Scarborough 2006)
- vocabulary level and familiarity of topic (Francis et al., 2005, Keenan et al., 2008)
- interest level of text (Francis et al., 2005)
- macrostructure (including propositional and intersentential complexity) (Francis et al., 2005)

Reading Task

- oral or silent reading (Keenan et al., 2008)
- level of support: errors corrected in oral reading or not
- time limited or not

Response Task (recognized by numerous authors as sources of inter-test variation)

- cloze (sentence or passage), maze, oral reading fluency, question response (closed or open-ended; direct or inferential; provided orally or written), multiple choice, sentence verification, picture selection, retell, and recall (this characteristic is noted by most authors)
- response mode: oral, written; short or long response; selection/forced choice or open-ended (Francis et al., 2005)
- question complexity relative to text
- level of question (recall, inferencing, etc) (Eason et al., 2012)
- availability of text for answering questions (Cutting & Scarborough, 2006; Kendeou et al., 2012)
- time limited or not (Kendeou et al., 2012)

Additional considerations

- Group or individual administration
- Success criterion: norm-referenced (consider norming sample), criterion-referenced (Snyder et al., 2005)
- Scoring: interpretation required or not (consider reliability) (Francis et al., 2005)
- Norming sample: truncated sample or not; comparability to population of interest (Snyder et al., 2005)
- Administration time: consider effects of attention and fatigue
- Theoretical basis for test construction (Cutting & Scarborough, 2006, van den Broek & Espin, 2012)
- internal factors of the examinee (e.g. motivation) (Francis et al., 2005)