AJSLP

Research Article

Read, Understand, Learn, & Excel (RULE): Development and Feasibility of a Reading Comprehension Measure for Postsecondary Learners

Priya Kucheria,[a] McKay Moore Sohlberg,[a] HyeonJin Yoon,[a] Stephen Fickas,[a] and Jason Prideaux[a]

Purpose: There is a lack of quick, reliable, and valid standardized reading comprehension assessments appropriate for postsecondary readers. We attempted to address this gap by designing Read, Understand, Learn, & Excel (RULE), a reading comprehension measure that employs sentence verification and recall tasks to assess reading comprehension. This article describes the exploratory study undertaken to construct RULE and then examines the preliminary concurrent validity and alternate form reliability of this measure.
Method: The RULE measure was first developed by designing reading stimuli, test items for the sentence verification task, and directions for the recall test for 2 forms based on previous work (Griffiths, Sohlberg, Kirk, Fickas, & Biancarosa, 2016). Thirty undergraduate students who identified themselves as typical readers were administered the RULE measure as well as the Nelson–Denny Reading Test (Brown, Fishco, & Hanna, 1993). Students also completed questionnaires and participated in informal interviews to provide information regarding study and learning habits and academic background.
Results: There was preliminary evidence of alternate form reliability between the sentence verification task sets of 2 chapters of RULE ($r = .38$, $p < .05$). Preliminary evidence for concurrent validity between RULE and the Nelson–Denny Reading Test was provided by correlation coefficients in the low to moderate range (.03–.38).
Conclusion: RULE design and preliminary findings of concurrent validity and alternate form reliability provide "proof of concept" for an ecologically valid testing format that assesses comprehension skills appropriate for the postsecondary level. Suggestions for strengthening validity and reliability of the tool are provided, and clinical contributions of RULE are discussed.
Supplemental Material: https://doi.org/10.23641/asha.6987371

Students with cognitive impairments resulting from acquired brain injury (ABI) are matriculating into postsecondary programs at increased rates to resume or establish career paths (Ackerman, DiRamio, & Mitchell, 2009; Kennedy, Krause, & Turkstra, 2008). This increase is partially due to the passage and refinement of the Americans with Disabilities Act (1990) designed to make education accessible. According to the National Center for Learning Disabilities (Cortiella & Horowitz, 2014), postsecondary education is now the goal of 67% of students with disabilities, including students with ABI. Undermining this goal is the academic struggle these students experience compared with their peers without injury due to difficulty with academic reading activities (Kennedy et al., 2008; Sohlberg, Fickas, & Griffiths, 2011; Todis, Glang, Bullis, Ettel, & Hood, 2011).

Approximately 30% of incoming college freshmen and 80% of adults with developmental or acquired disabilities attending college in the United States have reading comprehension challenges that negatively affect academic performance (National Research Council, 2012; Horn, Berktold, & Bobbitt, 1999). Individuals with reading comprehension deficits at postsecondary levels tend to have higher dropout rates, lower participation in the workplace, and lower earnings (Dodge, 2012; Kennedy et al., 2008; Meulenbroek, Bowers, & Turkstra, 2016; Wilkins & Huckabee, 2014; Wolf, 2001). Students with ABI and developmental reading challenges frequently demonstrate sufficient reading comprehension for functional reading tasks such as comprehending text in magazines or social media; however, they struggle when faced with reading comprehension demands

[a]University of Oregon, Eugene
Correspondence to Priya Kucheria: priyak@uoregon.edu
Editor-in-Chief: Julie Barkmeier-Kraemer
Editor: Kristie Spencer
Received December 18, 2017
Revision received March 15, 2018
Accepted April 16, 2018
https://doi.org/10.1044/2018_AJSLP-17-0221

Disclosure: *The authors have declared that no competing interests existed at the time of publication.*

for the purposes of learning complex, new academic content. Despite the need, there are very few supports for reading comprehension at the postsecondary level. Although our research group has focused on addressing the needs of postsecondary readers with ABI by developing and evaluating assessments and interventions designed to help individuals with deficits in high-level comprehension and retention (Griffiths, Sohlberg, Kirk, Fickas, & Biancarosa, 2016; Sohlberg, Griffiths, & Fickas, 2014), these efforts may have generalized utility for the larger population of struggling postsecondary readers.

Reading-to-learn is a complex process that requires extraction of information while engaging in continual construction, integration, and updating of concepts (Biancarosa & Snow, 2006). Although a review of reading comprehension theory is beyond the scope of this article, established models such as the Landscape Model (Yeari & Van den Broek, 2011) and Structure Building Framework (Gernsbacher, 1997) describe the cognitive processes underlying reading comprehension that are required for constructing mental representations of text via repeated activation of text elements that become linked to form coherent constructions. The effortful steps in the reading comprehension process are highly reliant on working memory. Working memory refers to the ability to temporarily maintain and manipulate information and accurately retrieve goal-relevant information from long-term memory (Baddeley, 2000; Engle & Kane, 2004; Unsworth & McMillan, 2013). Readers must hold on to relevant text as they are reading and integrate it with previously read text or with existing background information to build units of understanding (Yeari & Van den Broek, 2011). Studies with typical adult readers suggest that working memory is sensitive to individual variation in reading comprehension (Hannon, 2012). Although debatable, some researchers assert that this association stems from the ability to exert executive control over mind-wandering behavior (McVay & Kane, 2012). Others posit that this relation is often masked by the degree of correlation between working memory and other constructs such as IQ (Freed, Hamilton, & Long, 2017; Van Dyke, Johns, & Kukona, 2014). The link between working memory impairments and reading comprehension struggles is more strongly implicated in studies involving postsecondary learners with brain injury (Laatsch & Krisky, 2006; Sohlberg, Kucheria, Fickas, & Wade, 2015). Although reading comprehension research recognizes the primacy of working memory, current reading comprehension assessments do not use formats that place demands on working memory. The lack of theoretical grounding in current reading comprehension assessments at the postsecondary level represents one of several shortcomings important to address.

Barriers to Assessing Reading Comprehension at the Postsecondary Level

A major barrier to ensuring that individuals with comprehension impairments are matched with the right support or intervention is the lack of accurate and sensitive instruments for evaluating postsecondary reading competencies (McNamara & Kendeou, 2011). Common standardized assessments such as the Nelson–Denny Reading Test (NDRT; Brown, Fishco, & Hanna, 1993), Gray Oral Reading Tests–Fifth Edition (Wiederholt & Bryant, 2012), and comprehensive batteries such as the Woodcock-Johnson IV Tests of Achievement (McGrew, LaForte, & Schrank, 2014) serve as quick, easy-to-administer, and norm-referenced screeners to classify individuals based on reading levels. However, despite their feasibility, these tools are not designed to assess reading comprehension required in postsecondary contexts (Keenan, Betjemann, & Olson, 2008). Their theoretical frameworks, reading stimuli, and testing formats lack sensitivity to complex comprehension processes (McNamara & Kendeou, 2011; Rapp, Van Den Broek, McMaster, Kendeou, & Espin, 2007). These tools tend to evaluate lower-level reading processes such as phonological processing, word recognition, vocabulary, and sentence comprehension, which are often intact in postsecondary college readers (Hannon, 2012). Working memory, executive functions, and speed of processing are more predictive of postsecondary reading comprehension skills and are constructs that are typically not tapped in currently used tests (Macaruso & Shankweiler, 2010; Rapp et al., 2007; Sullivan, Griffiths, & Sohlberg, 2014).

The use of short, paragraph level text prevents evaluation of reading comprehension for lengthy expository passages required for college level reading. Passages range from 80 to 200 words in the NDRT, Gray Oral Reading Tests, and Woodcock-Johnson IV Tests of Achievement. Although these tests use expository text, the procedures for calculating the difficulty level and rationale for the specific content are not well specified and are not commensurate with postsecondary reading demands. Another challenge to ecological validity is the use of the multiple-choice or maze formats, the common modes used in current assessments. These testing formats primarily rely on recognition or information-locating skills to objectively assess comprehension (Farr, Pritchard, & Smitten, 1990; Ozuru, Best, Bell, Witherspoon, & McNamara, 2007). Many postsecondary reading tasks require the reader to integrate information and recall it without having the text available. Postsecondary reading demands emphasize applying information learned and recalling extensive amounts of information after a time delay. There is a need for a relatively quick, reliable, and valid standardized reading comprehension assessment appropriate for postsecondary readers.

Current Advances in the Measurement of Reading Comprehension

To address the limitations in current assessments of reading comprehension for postsecondary learners, we built on our earlier work evaluating reading comprehension in postsecondary college students with ABI that identified two fruitful evaluation formats, sentence verification task (SVT) and free recall (Griffiths et al., 2016; Sohlberg et al., 2014).

SVT is a widely used recognition paradigm in reading research that provides an alternative to the multiple-choice test format (Carlisle, 1989; Royer, Barbara, & Sinatra, 1987; Stine-Morrow, Soederberg Miller, Gagne, & Hertzog, 2008). The SVT paradigm has respondents read a statement and indicate whether the statement was "true" or "false" in relation to the content of the chapter. The true/false statements are generated based on the construction–integration framework, which operationalizes comprehension into three distinct levels of text comprehension: (a) paraphrased items, (b) local inferences, and (c) global inferences (Royer, 2001). In our previous work, performance on the SVT was sensitive to changes in reading comprehension when readers with cognitive impairment received reading comprehension training compared with the condition without training (Griffiths et al., 2016).

Recognition paradigms can be helpful but are not sufficient for measuring reading comprehension in more advanced learners. For example, they do not capture the active construction of meaning that occurs when text elements are integrated (Klingner, 2004). We thus employed free recall as a second measurement paradigm to complement the SVT (Griffiths et al., 2016). Discourse samples collected via free recall tasks have been used both experimentally and clinically to evaluate cognitive–communicative functions (Coelho, Ylvisaker, Turkstra, 2005; Laing-Gillam, Fargo, & Robertson, 2009). Verbal samples elicited from story retell, story generation, and describing procedures have been shown to be valid measures of listening comprehension, discourse production, and reading comprehension in individuals with cognitive impairments due to brain injury (Coelho, Lê, Mozeiko, Krueger, & Grafman, 2012; Griffiths et al., 2016) as well as typical readers (Keenan et al., 2008; Reed & Vaughn, 2012). Measuring comprehension via discourse analysis of recall transcripts allows direct analysis of the mental representation of textual content constructed by the reader (Pearson & Hamm, 2005). Relevant to our population of interest, text recall activities have been shown to rely on working memory (Towse, Cowan, Hitch, & Horton, 2008).

In an early pilot study, participants were asked to pretend they were telling someone everything they had learned in the textbook passage they had just read (Sohlberg et al., 2011). The free recall samples were analyzed to obtain measures of total communication units and communication units per minute (C-units; Loban, 1976), which provided a quantitative measure of recall and recall efficiency. Our later study (Griffiths et al., 2016) used a more fine-grained analysis of total content recalled (correct information units [CIUs]; Nicholas & Brookshire, 1993) that included immediate and delayed recalls and efficiency measures, which had been shown to distinguish between readers with cognitive impairments and controls (Matsuoka, Kotani, & Yamasato, 2012). These procedures were adopted for use in the proposed instrument that was developed and piloted in this study.

Purpose of the Study

This article is an exploratory study designed to address the current reading comprehension assessment gap. We sought to develop a reliable and valid reading comprehension test that would simulate the demands for postsecondary reading. We also sought to design a test that was feasible and could be given in a standard session by a clinician or an educational staff. We chose to investigate normative performance on a high-level reading comprehension measure as a first step in developing a tool that could be used to identify impaired performance. Our specific project goals were to (a) develop test content that was representative of college level reading in content and length; (b) employ testing formats that placed demands on working memory, speed of processing, and recall processes; (c) construct the test such that it could be administered and scored within 50 min; (d) develop two equivalent assessments that could be used to evaluate reading comprehension pre- and postintervention support; and (e) pilot the administration and analysis on a group of typical undergraduate readers. We called our preliminary assessment RULE (Read, Understand, Learn, & Excel).

Phase 1: Test Development Methods

Our RULE project was carried out in two phases. Phase 1 was dedicated to test development activities, and Phase 2 focused on validation procedures. In this section, we describe the systematic development of reading stimuli, test items, and overall format of the electronic assessment. In Phase 2, we discuss activities undertaken to examine feasibility and content and ecological validity of the RULE measure from the stakeholder's perspective. We also describe the preliminary validation pilot. Specifically, we discuss findings related to concurrent validity and alternate form reliability of this measure when piloted on a small sample of typical undergraduate readers used to establish normative performance.

Passage Development

Our first goal was to create reading stimuli. Texts selected needed to be expository in nature, simulate content typically encountered at the college level, and be equated in length and complexity. We identified two expository passages from college textbooks that represented disciplines encountered in academic learning (Biancarosa & Snow, 2006; Lee & Spratley, 2010; RAND Reading Study Group, 2002). The materials were identified on an open-source website (cnx.org) that allowed us to download and modify college texts without copyright infringement. To avoid confounding factors influencing comprehension, passages needed to share a similar writing structure (Biber, 1985; Kulesz, 2014; Lee & Spratley, 2010). Therefore, we selected opening chapters from two introductory level textbooks: public speaking and social psychology. Introductory social science texts tend to share similar structural patterns

that can be categorized using broad-based expository text typology schemes, such as compare and contrast and cause–effect (Pyle et al., 2017). This selection rationale helped eliminate differences in comprehension arising from organizational differences from different subject matter areas as well as a reader's background knowledge when using the measure to assess pre–post treatment effects. It was necessary to use chapters from different sources despite differences in authors' writing styles, so that content could be understood independently from the rest of the book. Text was eliminated to reach a length of approximately 3,000 words, consistent with the recommendations in the Griffiths et al. (2016) study. Elimination of content began from the end of the chapter to preserve organizational coherence. Texts then underwent latent semantic analysis to ensure that they were equated in length and complexity. Coh-Metrix, an automated text analysis tool, was utilized to compare content within the chapters on discourse and linguistic measures (Graesser, McNamara, & Kulikowich, 2011). Coh-Metrix computes over 50 metrics that describe semantic and syntactic features, text cohesiveness, and text difficulty (Graesser, McNamara, Louwerse, & Cai, 2004). We selected multiple indices, such as argument overlap and sentence-to-sentence cohesion, mentioned in prior studies as sensitive indicators to differences in texts from diverse subject areas to analyze chapters in this study (Lightman, McCarthy, Dufty, & McNamara, 2007; McCarthy, Graesser, & McNamara, 2006; McNamara, Graesser, & Louwerse, 2012).

Several existing indices were used to ensure that chapters were equated in length and complexity. We used index values for complexity and cohesiveness from a previous study (Griffiths et al., 2016) as benchmark criteria. We also made sure that differences in index values across texts fell within the 95% confidence interval range norms created by Coh-Metrix tool developers. Finally, we measured Lexile levels (Stenner, Sanford-Moore, & Williamson, 2012) to confirm that all chapters corresponded to a 12th-grade difficulty level.

To ensure that the chapters were feasible to read within clinical time constraints, we ran an informal pilot study with two speech-language pathologists working in outpatient settings whose caseloads included clients with reading comprehension goals after a mild traumatic brain injury. Feedback from both clinicians was that the chapters were too long to utilize in a clinic setting. Using guidelines for condensing texts for academic purposes and literature on silent reading rates in persons with traumatic brain injury (Green & Hawkey, 2012; Harvey, Hux, Scott, & Snell, 2013), we reduced the chapter length to 2,200–2,400 words. We used the same criteria stated previously to ensure equivalence. The final chapter metrics are presented in Table 1.

Testing Formats

SVT

The SVT is a recognition task that requires the reader to identify whether sentences are "true" or "false" with respect to the corresponding passage. The reader does not have access to the passage when responding. To design the SVT measure, we replicated the approach outlined in previous studies (Griffiths et al., 2016; Royer, Hastings, & Hook, 1979). Three independent raters selected main and supporting details from each chapter and coded them into autonomous and nonautonomous units. Autonomous units could be understood independently of the rest of the text, for example, "Humans have lived on Earth for thousands of years." Nonautonomous units required reference to the text to be understood, for example, "Humans are influenced by the social situation along with their schemas and attitudes regarding a situation." These units were the source of three different types of sentences—local coherence, global coherence, and paraphrase statements—that were the contents of the first SVT set for each chapter. Theories such as Structure Building Framework (Gernsbacher, 1997) driving SVT development postulate that text comprehension occurs at different levels: (a) surface level where the reader has formulated an understanding of the words, syntactic relations, and underlying meaning of a sentence, captured via paraphrased items; (b) microstructural level where the reader formulates an understanding of the meaning of adjacent sentences or ideas captured via local cohesion items; and (c) macrostructural level where ideas within a paragraph or text are integrated with the general theme of the text captured through global cohesion items (Gernsbacher, 1997; Sohlberg et al., 2014). A study evaluating the effects of a strategy training intervention on comprehension noted that these items were sensitive to the intervention (Griffiths et al., 2016).

To avoid passage dependency, we piloted the formulated SVT sets with two groups, individuals who completed SVT with and without reading the chapter. For each group, we organized statements into a progressive hierarchy of difficulty starting with those statements that 100% of the test-takers responded to correctly, to those that only 20% (or less) of test-takers responded to accurately. We eliminated statements that were accurately answered by only 10% of the respondents in either group as well as statements that were accurately answered by 100% of test-takers in the group that did not read the chapter. An equal number of true and false statements from each percentage category from the remaining item bank were then selected, with two statements selected from the "easy" category (80%–100% correct response) and the remaining statements selected from the progressively "hard" categories (60%–79%, 40%–59%, and 20%–39%) to form the SVT set, each consisting of 15 statements.

Free Recall

A goal-directed recall task is considered a high-inference measure because it gives direct access to the mental representation of text content constructed by a reader (Pearson & Hamm, 2005). In RULE, we provided a goal-directed context for eliciting verbal discourse by asking participants to "recall information as though they were helping a friend prepare for a test." To measure comprehension based on free recall performance, we used the approach validated in previous studies that calculated the number of CIUs

Table 1. Summary of text characteristics of Read, Understand, Learn, & Excel chapters.

Variable	Chapter	
	Public speaking	Social psychology
Quantity		
Total words	2,360	2,371
Number of different words	764	753
Total number of sentences	130	128
Complexity		
Mean word length (mean number of syllables per word)	1.70 (1.00)	1.67 (0.95)
Lexical density	31.94%	31.49%
Lexical diversity (type–token ratio)	0.53 [0.80, 0.84]	0.52 [0.67, 0.71]
Mean sentence length	18.73 (11.15)	19.28 (11.59)
Cohesiveness		
Argument overlap	0.59 [0.51, 0.59]	0.62 [0.61, 0.69]
Stem overlap	0.64 [0.36, 0.46]	0.60 [0.61, 0.69]
Sentence-to-sentence cohesion	0.30 (0.20) [0.36, 0.40]	0.32 (0.19) [0.45, 0.49]
Paragraph-to-paragraph cohesion	0.26 (0.20) [0.33, 0.37]	0.26 (0.19) [0.36, 0.42]
Readability metrics		
Gunning Fog Index	16.39	15.25
Lexile level[a]	1320L	1270L

Note. Values in parentheses are standard deviations. Values in brackets are 95% confidence interval ranges based on Coh-Metrix norms.

[a]Lexile levels for postsecondary populations are typically between 1200L and 1380L (Stenner et al., 2012).

generated during the verbal recall of what the reader recollected from the chapter (Griffiths et al., 2016). A CIU is defined as any word that is intelligible in context, accurate, and relevant in relation to the text (Nicholas & Brookshire, 1993). Grammatical correctness is not a factor. Traditionally, CIUs have been used as a measure of discourse during picture description or spontaneous conversation tasks. We used the rule-based scoring system formulated by Nicholas and Brookshire (1993) in our work to analyze verbal recalls of expository text. A related measure that also discriminates between reading comprehension skills of individuals with and without brain injuries is efficiency of recall, calculated by counting the number of CIUs generated per minute (CIUrate; Griffiths et al., 2016; Matsuoka et al., 2012). To calculate the CIUrate, we divided the total number of CIUs by the time taken for completing the recall task. This efficiency metric is critical given that speed of processing has been one of the primary factors/contributors to reading comprehension difficulties after a mild brain injury (Sullivan et al., 2014).

We designed the RULE to contain an immediate free recall and a delayed free recall. Delayed recall allows measurement of retention of material, which is often problematic in postsecondary readers with acquired cognitive impairments (Kennedy et al., 2008). The delayed recall task and CIU analysis were identical to the immediate recall task and were administered 30 min after the immediate condition.

Phase 2: Pilot

Feasibility Study

Several targeted user groups were formed to gather feedback on the RULE assessment from end users including three high school special education teachers, a university academic learning specialist, and six speech-language pathologists whose caseloads included postsecondary students with ABI. The purpose was to examine feasibility and content and ecological validity of RULE. Users were trained to use the RULE assessment and then were given access to it so that they could practice using the tool outside the training session. Users completed surveys querying them on their perceptions of the reading stimuli, the design of the measures, and the overall usability of the tool. In addition, a university student who self-reported high-level reading comprehension difficulties due to developmental attention challenges reviewed the assessment and was asked to give her perception of the stimuli and testing procedures from the student perception.

The clinicians and university learning specialist all endorsed that the RULE (a) filled a measurement gap in their practices, (b) used reading stimuli with a difficulty level and length commensurate with academic demands faced by the populations with whom they worked, and (c) was a tool they would use in their practice. They reported that both SVT and recall provided an objective measurement of an individual's reading approach, which in their current clinical practices was obtained through self-report or was not being addressed. The student reported that the reading level was commensurate with college demands and that the test length was acceptable. The high school teachers shared that students with high-level reading comprehension difficulties did not qualify for special education services, hence they would not be able to use the test in their settings, although they endorsed the existence of students who could benefit from being identified using the tool.

All the clinicians reported a feasibility concern regarding the amount of time it took to transcribe the readers' text recalls necessary for calculating the CIUrate.

Validation Study

After developing RULE and soliciting feedback from end users, we sought to evaluate the reliability and validity of the tool. Specifically, we looked for alternative form reliability and concurrent validity.

Participants

We recruited and consented typical undergraduate readers in accordance with the regulations of our institutional review board protocol. Each participant received $30.00 for participation. Inclusion criteria were as follows: (a) aged 18–30 years, (b) self-identified as fluent speakers of English who had acquired the language before 7 years old, (c) currently enrolled as full-time undergraduate students in a university or college, (d) able to read for daily needs (e.g., street signs, menus, and/or bills), (e) able to comprehend three- to four-paragraph–length material at the 10th grade level, and (f) familiar with using a laptop for reading text, typing, and clicking and scrolling functions of a mouse. Exclusion criteria included (a) diagnosis of a disability or condition that affected basic reading abilities and/or (b) admission to a hospital or outpatient program in the last 12 months for substance abuse or psychiatric issues. Thirty-two individuals (26 women and six men) met criteria and consented to participate. Data for two participants were lost because of attrition and technological difficulties, resulting in a final sample of 30 subjects (24 women and six men).

Table 2 presents participant characteristics. During the waiting portion for the delayed recall test described below, participants answered a questionnaire about their demographics and reading performance, completed the Learning and Study Strategies Inventory (LASSI; Weinstein, Palmer, & Acee, 2016), and were administered the NDRT (Brown et al., 1993). Of note, approximately 40% of the sample reported that it was somewhat typical of them to forget information that they had just read. More than 80% of the subjects reported that it was either fairly or very typical of them to read for extended periods. These ratings implied that, although most of the participants identified themselves as capable of reading for extended periods, many found it challenging to recall and retrieve academic reading content. Mean scores on the information processing, selecting main ideas, and self-testing scale of the LASSI were less than the 50th percentile. According to interpretation guidelines for this measure, half of the sample would benefit from training on learning and study strategies. Mean percentile score (M = 53.20) on the comprehension subtest of the NDRT suggested that the sample consisted of largely "average" readers.

Measures

There were three measures used in the analyses. The SVT and free recall measures were developed for RULE and are described under Test Development. The NDRT (Brown et al., 1993) was used as a reference measure for validity analysis. In the absence of a valid postsecondary reading comprehension measure, we selected the NDRT (Brown et al., 1993) because it is the most commonly used measure to evaluate college-level comprehension performance (Griffiths et al., 2016; Haenggi & Perfetti, 1994; Holmes, 2009; Masterson & Hayes, 2004).

Methods

We collected data between November 2016 and March 2017. Each subject attended two sessions. Sessions were conducted in a private clinic room at a university training clinic associated with a communication disorders and sciences program. The first author or a trained graduate or undergraduate research assistant administered the study protocol. The first session lasted between 1.25 and 1.50 hr. Participants were oriented to the digital features of RULE, which was delivered on a MacBook Pro (Apple). Participants read one

Table 2. Summary of trends on demographic questionnaire and descriptive information on Learning and Study Strategies Inventory subscales.

Demographic questionnaire			
Items	**I can read for extended periods**	**I forget information I have just read**	**I forget information after a delay**
Not at all typical of me	3%	3%	0%
Not very typical of me	3%	37%	43%
Somewhat typical of me	10%	43%	50%
Fairly typical of me	47%	10%	1%
Very much typical of me	37%	7%	1%
Learning and Study Strategies Inventory			
Scale	**Information processing**	**Selecting main ideas**	**Self-testing**
Mean	43.87	64.83	35.77
SD	24.67	21.27	25.58

Note. N = 30. Numerical values on the informal interview section of the table represent percentage of respondents who selected the same response category when presented with an item. Numerical values under the Learning and Study Strategies Inventory section of the table represent mean percentile and standard deviation scores of all participants.

of the two expository texts excerpted from open-source, introductory-level college textbooks using the aforementioned procedures. The order of chapter assignment was determined using the randomization function in Excel. On the basis of results of the feasibility pilot study, we gave participants up to 30 min to read a chapter before completing SVT and recall tasks.

Reading measures were administered on the computer after the allowed reading time or after the participant clicked the "finished reading" button. Participants were told to "retell the text as if you were telling someone all the information you learned and were helping them prepare for a test." Recalls were recorded by the online program and downloaded as audio files on the MacBook. After the recall task, participants completed the SVT. Statements were displayed individually on the computer screen, and participants clicked either "true" or "false" response. Participants had approximately 10 min to complete the SVT and immediate recall tasks. The program was designed to present the delayed recall probe 30 min after a subject had clicked on the "finished reading" button. Participants would receive a "Please wait" screen after completing the SVT. During this wait period, we administered the comprehension subtest of the NDRT (Brown et al., 1993). Participants were given 20 min to complete the comprehension subtest. This was followed by the delayed version of the recall task, which was administered identically to the immediate version.

The second session followed the same format. We administered the remaining expository texts and corresponding SVT and recall tasks. During wait times for the recall task, participants were informally interviewed to gather demographic data or were administered questionnaires about their reading.

Interrater reliability for calculating CIUrate. A second trained rater calculated CIUrate on 25% of the recall transcripts. Intraclass correlation coefficient estimates and their 95% confident intervals were calculated based on a single-rater, absolute-agreement, two-way mixed-effects model with two raters across 34 transcripts. An intraclass correlation coefficient of .89 (95% CI [0.74, 0.95]) suggests good to excellent reliability, particularly for a new measure (Koo & Li, 2016).

Results

Preliminary Analysis

Table 3 displays the descriptive statistics for the two assessment variables and a reference measure for validity analysis (i.e., NDRT scores). Average reading time was approximately 15 (SD = 5.65) and 14 (SD = 4.75) min for the public speaking and social psychology chapters, respectively. The assumption of normality was not supported for the public speaking chapter (Shapiro & Wilk, 1965). A boxplot for the public speaking chapter revealed an outlier case, which we dropped. Therefore, the final analytic sample included 29 subjects. Per Shapiro–Wilk test (Shapiro & Wilk, 1965), normality of data for all distributions was supported for the modified data set, with skewness and kurtosis for the distributions of the SVT scores and CIUrates raising no major concerns across chapters.

We conducted a series of mixed analysis of variance to probe the effects of order of chapter administration, participants' class level (freshman, sophomore, junior, or senior), and chapter type on participants' performance on each of the RULE measures: (a) SVT scores and (b) CIUrate. For CIUrate, the effect of timing of recall task (i.e., immediate or delayed) was additionally included as a within-subject factor. Alpha was adjusted to reduce risk of Type 1 error given multiple tests ($\alpha = .05/2 = .025$). Effect size was reported as partial eta squared (Richardson, 2011).

SVT performance. We ran a $6 \times 4 \times 2$ three-way mixed analysis of variance for SVT performance with order of chapter administration (six possible administration permutations) and class level as between-subject factors and chapter type (social psychology or public speaking) as a within-subject factor. Mauchly's test of sphericity was not defined because the within-subject factor (i.e., chapter type) had only two levels. Levene's test supported homogeneity of error variance of the SVT scores for both chapters ($p > .05$). Results revealed nonsignificant omnibus interaction effect as well as corresponding two-way interaction effects for a combination of factors on SVT scores. There was no main effect of order, $F(1, 21) = 0.001$, $p = .97$, partial $\eta^2 = 0.000$; class level, $F(3, 21) = 0.75$, $p = .54$, partial $\eta^2 = .097$; and chapter type, $F(1, 21) = 0.60$, $p = .59$, partial $\eta^2 = .014$.

CIUrate. The assumption of homogeneity of variance was not supported for CIUrate data because of the unbalanced cell sizes. Therefore, we built a simpler model—a one-way analysis of covariance—to examine the effect of chapter on immediate and delayed CIUrate controlling for administration order. The analysis of covariance showed no significant difference between immediate, $F(1, 55) = 1.72$, $p = .195$, partial $\eta^2 = .030$, and delayed, $F(1, 55) = 0.55$, $p = .462$, partial $\eta^2 = .010$, CIUrates by chapter administration order ($p > .05$). There were also no chapter effects on immediate, $F(1, 55) = 0.38$, $p = .54$, partial $\eta^2 = .007$, and delayed, $F(1, 55) = 0.02$, $p = .883$, partial $\eta^2 = .000$, CIUrates, controlling for chapter administration order as a covariate ($p > .05$).

Concurrent Validity

We computed the Pearson's product–moment correlation coefficients for each of the RULE measures (i.e., SVT scores, immediate and delayed CIUrates) and NDRT comprehension subtest percentiles. As shown in Table 4, the validity coefficients of RULE measures were small to moderate, ranging from .09 to .39.

Because of the gap in the availability of standardized measures of high-level reading comprehension that could be used to evaluate concurrent validity, we ran post hoc analyses using descriptive measures evaluating the reader's self-reported measures and functional measures of reading comprehension. We attempted stratification techniques

Table 3. Means and standard deviations for scores on Read, Understand, Learn, & Excel (RULE) and Nelson–Denny Reading Test (NDRT; Comprehension subtest).

	RULE[a]			RULE[b]			NDRT
Measure	Immediate[c]	Delayed[d]	SVT	Immediate[c]	Delayed[d]	SVT	Subtest
Mean	70.29	59.31	63.45	66.08	57.94	62.07	53.45
SD	28.65	35.80	17.98	23.63	34.11	17.40	27.10

Note. $n = 29$. SVT = sentence verification task.

[a]Public speaking. [b]Social psychology. [c]Immediate number of correct information units generated per minute. [d]Delayed number of correct information units generated per minute.

to informally examine trends in descriptive data between responses on LASSI, the demographic questionnaire, and reported grade point average (GPA) and RULE measures. Our post hoc hypotheses for concurrent validity would be supported if trends were in the same direction, that is, if high versus low means on RULE measures were consistent with trends in means on another measure.

Data were recoded and sorted in three ways to evaluate performance of low scorers against that of high scorers across two or more measures at a time. We computed quartiles to determine scores at the 25th, 50th, and 75th percentiles for measures that posed as continuous variables (e.g., CIUrates) and used those indices to divide the sample into three groups. For measures posing as nominal/ordinal data (demographic questionnaire responses), we dummy coded the data so that subjects were identified either as those who had responded with either "not at all typical of me" or "not very typical of me" to questions regarding their immediate and delayed recall abilities or those who responded with either "somewhat typical of me," "fairly typical of me," or "very much typical of me" when referring to the same construct. Data were also sorted to compare individuals who scored at or below the 25th percentile across all SVT and CIU measures (low RULE performers) against those who scored at or above the 75th percentile across SVT and CIU measures (higher RULE performers).

When using self-report of recall ability as a grouping variable, the group that self-reported that it was either "not very" or "not at all" typical of them to forget information immediately after reading and after a delay had higher mean CIUrates (public speaking: $M_{\text{Immediate}} = 77.04$ and $M_{\text{Delayed}} = 68.10$; social psychology: $M_{\text{Immediate}} = 73.13$ and $M_{\text{Delayed}} = 75.61$) compared with the group that claimed that it was "somewhat typical," "fairly typical," or "very much typical" of them to forget information immediately and after a delay (public speaking: $M_{\text{Immediate}} = 64.45$ and $M_{\text{Delayed}} = 52.20$; social psychology: $M_{\text{Immediate}} = 63.95$ and $M_{\text{Delayed}} = 53.58$; refer to the table provided in Supplemental Material S1). When using GPA as a grouping variable, the group whose GPA was at or below the 25th percentile displayed lower mean CIUrates (public speaking: $M_{\text{Immediate}} = 74.51$ and $M_{\text{Delayed}} = 56.84$; social psychology: $M_{\text{Immediate}} = 75.82$ and $M_{\text{Delayed}} = 64.72$) as compared with the group with individuals whose GPA was in the 75th percentile or a higher range (public speaking: $M_{\text{Immediate}} = 101.26$ and $M_{\text{Delayed}} = 98.67$; social psychology: $M_{\text{Immediate}} = 89.36$ and $M_{\text{Delayed}} = 91.06$; refer to the table provided in Supplemental Material S1). Not all participants ($n = 5$) reported their GPA. Consistent with the questionnaire and GPA findings, the lowest RULE performers also received the lowest scores on the LASSI information processing scale ($n = 1$, $M = 5$th percentile) as compared with the highest RULE performers ($n = 2$, $M = 55$th percentile; refer to the table provided in Supplemental Material S2). It was not possible to test the significance of these trends because the sorting/classification guidelines applied to the data set yielded small numbers within each group.

Table 4. Summary of intercorrelations between Read, Understand, Learn, & Excel (RULE) measures and Nelson–Denny Reading Test (NDRT; Comprehension subtest).

	RULE: Public speaking			RULE: Social psychology		
Measures	Immediate[c]	Delayed[d]	SVT	Immediate[c]	Delayed[d]	SVT
NDRT	0.39*	0.29	0.29	0.09	0.20	0.10
SVT[a]	0.23	0.07	—	0.05	–0.03	0.38*
SVT[b]	–0.03	0.10	0.38*	0.09	0.17	—

Note. $N = 29$. SVT = sentence verification task. Dashes indicate correlation between same variables.

[a]Public speaking. [b]Social psychology. [c]Immediate number of correct information units generated per minute. [d]Delayed number of correct information units generated per minute.

*$p < .05$.

Alternate Form Reliability

We computed the Pearson's product–moment correlation coefficients of each of the RULE measures between chapters. As shown in Table 4, the reliability coefficients were small, ranging from .03 to .38. Although moderate, correlation between SVT forms was significant ($r = .38$, $p < .05$).

Discussion

This article reviews the development and preliminary evaluation of a postsecondary reading comprehension measure, RULE. Findings suggest that RULE has the potential to offer an ecological, valid testing format feasible for use in clinical or educational practice. The study details theory and procedures for developing the reading stimuli and measures. Important contributions include development of stimuli that simulate postsecondary reading demands by using actual college textbook passages that are longer and more complex than those in existing reading comprehension tests. The study also allowed evaluation of a dual testing format that encompasses both a recognition paradigm, SVT, and the analysis of immediate and delayed free recall samples using CIU analysis. The CIUrate is dependent on working memory, speed of processing, and text integration, which represents a novel and critical aspect of reading comprehension measurement. Interestingly, there was not a difference in performance for the immediate versus delayed CIUrates, which in hindsight might have been expected in a typically functioning population. We hypothesize that students with reading comprehension deficits secondary to brain injury would show differences between immediate and delayed recall conditions (Griffiths et al., 2016). Feedback from stakeholders, clinicians, and educators endorsed the need and interest in having a reading comprehension assessment tool such as RULE.

The concurrent validity was small to moderate and in the right direction (positive), which is an important finding for a new clinical tool using a small sample. The small correlations likely reflected the fact that RULE and NDRT are measuring different constructs. For example, like the SVT, NDRT uses a recognition task, of a multiple-choice format, but the test allows the reader to reference the text when filling out the information. This is different than the SVT that requires the reader to recall information when answering the questions. Recall requires working memory and simulates the demands of testing in postsecondary contexts. There is no comparator in the NDRT test for the recall task used to calculate the CIUrate. Given the lack of available instruments measuring reading comprehension using college level materials, we did not have a gold standard comparator for the concurrent validity, hence a small positive correlation would be expected.

Despite the lack of gold standard comparators, trends in descriptive data across measures endorsed our post hoc hypotheses for concurrent validity. A low GPA, difficulty with immediate and delayed recall as measured via self-report, and difficulty with comprehension measured via the LASSI information processing subscale were reflected in RULE scores when compared against individuals who scored relatively higher across the same set of measures. Although not fully conclusive, consistency in performance across measures provides preliminary support for concurrent validity.

The alternate form reliability was lower than desired. However, given the small sample size and novel testing format, a positive correlation is encouraging. Several improvements in the stimuli development process would likely strengthen alternate form reliability. First, it would be helpful to have reading comprehension experts review the two passages to establish content validity. Second, if modifying existing texts, it might be beneficial to examine additional passage features in the Coh-Metrix analyses (e.g., the number of unique information units, syntax structure, and/or types/number of conjunctions used) to increase chapter equivalency. Another option would be the use of existing textbook passages. Since the time of the project, we have become aware of resources (ReadWorks, 2018) with textbook passages already equated; however, they are shorter than our target passages. Perhaps most importantly, it would be valuable to equate the SVT items across chapters on a range of item features (e.g., readability, cohesion, the number of unique information units, syntax structure, the types/number of conjunctions use, item types, and/or item difficulty).

Feasibility was supported as the measure could be administered within the 50-min time frame, the maximum assessment time allowed in most clinical settings. Adoption of clinical tools depends in part on fit with context and provider need, hence it is critical when developing tools to consider implementation from the outset (Damschroder et al., 2009; Sohlberg, Kucheria, et al., 2015). That said, the targeted end users noted that the CIUrate transcription and calculation procedures were too lengthy for RULE to be feasible. This led our research team to begin investigating options for autotranscription and automation of CIU calculation. Currently, we are running a study evaluating the partial automation of the CIU calculation and comparing clinician scoring time using the automated protocol to manual scoring. This work will be important to develop a tool that can be fully adopted.

Limitations

Although the RULE development and initial piloting provide a proof of concept and encourage continued exploration, there were several limitations of the study that warrant caution in interpretation. Our sample size was small. It also is not known whether the sample was representative of typical undergraduates. There was gender disparity in our small sample because of a self-selection bias affecting sampling, which would be important to correct in future studies. We were surprised by the numbers of individuals reporting scores on the LASSI suggesting that they

would benefit from study strategy training and the number who reported feeling challenged when they needed to recall and retrieve academic reading content. There have been some suggestions in the literature that reading comprehension is challenging for today's undergraduates and may reflect changes in educational practices at the elementary and secondary levels (Sohlberg, Griffiths, & Fickas, 2015). Although this may make it difficult to develop norms, decreasing academic reading comprehension would suggest it important to develop a valid, reliable reading comprehension measure.

It is also important to acknowledge the lack of data on the use of discourse analysis for a reading comprehension measure. There have been fewer than 10 studies that have employed retell or recall tasks to assess reading comprehension at the postsecondary level (Reed & Vaughn, 2012; Watter, Copley, & Finch, 2017). Although it is well supported theoretically and in research practice, it has not been used as a clinical tool. Furthermore, verbal recall is dependent on verbal ability, and like "think aloud" protocols, verbal ability may confound results as opposed to reading comprehension.

Conclusions and Next Steps

The results of this development and preliminary evaluation are informing future research. Our immediate next steps are to focus on the passage and item development activities and ensure that the passage and SVT items are equated across chapters. Improved alternate form reliability is critical to develop a measure that can be used to measure treatment effects. Another step that would contribute to the robustness of concurrent validity findings is to use a mixed methods design to assess comprehension. This would entail using functional measures of reading (such as cumulative GPA for courses heavily reliant on reading) and a variety of self-report measures that provide data on comprehension ability and effort or difficulty level experienced when completing tests.

It is hoped that lessons learned in this study will stimulate and guide other reading comprehension research focused on developing and evaluating tools appropriate for postsecondary readers. Study findings provide preliminary support for using ecological materials that match the demands of college readers and validate assessments that use both recognition paradigms and recall measures that tap working memory.

Acknowledgments

This study was supported by the National Science Foundation PFI-AIR TT Grant 1640492 awarded to McKay Moore Sohlberg at the University of Oregon.

The authors would like to thank all the participants who contributed to the study, as well as Kayla Davis, Haley Hash, Elise Heater, Rachel Lacewell, Kelsey Lewey, and Garrett Porter, who acted as research assistants.

References

Ackerman, R., DiRamio, D., & Mitchell, R. L. G. (2009). Transitions: Combat veterans as college students. *New Directions for Student Services, 2009*(126), 5–14. https://doi.org/10.1002/ss.311

Americans With Disabilities Act of 1990, Pub. L. No. 101-336, 104 Stat. 328 (1990).

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4,* 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Biancarosa, C., & Snow, C. E. (2006). *Reading next—A vision for action and research in middle and high school literacy: A report to Carnegie Corporation of New York* (2nd ed.). Washington, DC: Alliance for Excellent Education.

Biber, D. (1985). Investigating macroscopic textual variation through multifeature/multidimensional analyses. *Linguistics, 23*(2), 337–360. https://doi.org/10.1515/ling.1985.23.2.337

Brown, J. I., Fishco, V. V., & Hanna, G. S. (1993). *Nelson–Denny Reading Test*. Rolling Meadows, IL: Riverside Publishing.

Carlisle, J. F. (1989). The use of the sentence verification technique in diagnostic assessment of listening and reading comprehension. *Learning Disability Research, 5,* 33–44.

Coelho, C., Lê, K., Mozeiko, J., Krueger, F., & Grafman, J. (2012). Discourse production following injury to the dorsolateral prefrontal cortex. *Neuropsychologia, 50*(14), 3564–3572.

Coelho, C., Ylvisaker, M., & Turkstra, L. S. (2005). Non-standardized assessment approaches for individuals with traumatic brain injury. *Seminars in Speech and Language, 26,* 223–241.

Cortiella, C., & Horowitz, S. H. (2014). *The state of learning disabilities: Facts, trends and emerging issues*. New York, NY: National Center for Learning Disabilities.

Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implementation Science, 4,* 50. https://doi.org/10.1186/1748-5908-4-50

Dodge, K. M. (2012). Examining the lived experience of students with reading comprehension learning disabilities and the perceived value of the accommodations received (p. 300). *Theses and Dissertations*. The University of Toledo, Toledo, OH.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 44, pp. 145–199). New York, NY: Elsevier Science.

Farr, R., Pritchard, R., & Smitten, B. (1990). A description of what happens when an examinee takes a multiple-choice reading comprehension test. *Journal of Educational Measurement, 27*(3), 209–226.

Freed, E. M., Hamilton, S. T., & Long, D. L. (2017). Comprehension in proficient readers: The nature of individual variation. *Journal of Memory and Language, 97,* 135–153. https://doi.org/10.1016/j.jml.2017.07.008

Gernsbacher, M. A. (1997). Two decades of structure building. *Discourse Processes, 23,* 265–304.

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher, 40*(5), 223–234.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, and Computers, 36*(2), 193–202.

Green, A., & Hawkey, R. (2012). An empirical investigation of the process of writing: Academic Reading test items for the International English Language Testing System. *Studies in Language Testing,* 270–378.

Griffiths, G. G., Sohlberg, M. M., Kirk, C., Fickas, S., & Biancarosa, G. (2016). Evaluation of use of reading comprehension strategies to improve reading comprehension of adult college students with acquired brain injury. *Neuropsychological Rehabilitation, 26*(2), 161–190. https://doi.org/10.1080/09602011.2015.1007878

Haenggi, D., & Perfetti, C. A. (1994). Processing components of college-level reading comprehension. *Discourse Processes, 17*(1), 83–104. https://doi.org/10.1080/01638539409544860

Hannon, B. (2012). Understanding the relative contributions of lower-level word processes, higher-level processes, and working memory to reading comprehension performance in proficient adult readers. *Reading Research Quarterly, 47*(2), 125–152. https://doi.org/10.1002/RRQ.013

Harvey, J., Hux, K., Scott, N., & Snell, J. (2013). Text-to-speech technology effects on reading rate and comprehension by adults with traumatic brain injury. *Brain Injury, 27*(12), 1388–1394.

Holmes, V. M. (2009). Bottom-up processing and reading comprehension in experienced adult readers. *Journal of Research in Reading, 32*(3), 309–326. https://doi.org/10.1111/j.1467-9817.2009.01396.x

Horn, L., Berktold, J., & Bobbitt, L. (1999, June). *Students with disabilities in postsecondary education: A profile of preparation, participation, and outcomes* [statistical analysis report]. Washington, DC: National Center for Education Statistics. U.S. Department of Education.

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. https://doi.org/10.1080/10888430802132279

Kennedy, M. R. T., Krause, M. O., & Turkstra, L. S. (2008). An electronic survey about college experiences after traumatic brain injury. *Neurorehabilitation, 23,* 511–520.

Klingner, J. K. (2004). Assessing reading comprehension. *Assessment for Effective Intervention, 29*(4), 59–70.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliabilty research. *Journal of Chiropractic Medicine, 15,* 155–163.

Kulesz, P. A. (2014). *The effects of reader characteristics, text features, and comprehension processes on reading comprehension* (Doctoral dissertation). Retrieved from University of Houston Institutional Repository (http://hdl.handle.net/10657/752)

Laatsch, L., & Krisky, C. (2006). Changes in fMRI activation following rehabilitation of reading and visual processing deficits in subjects with traumatic brain injury. *Brain Injury, 20*(13-14), 1367–1375.

Laing-Gillam, S. L., Fargo, J. D., & Robertson, K. S. C. (2009). Comprehension of expository text: Insights gained from think-aloud data. *American Journal of Speech-Language Pathology, 18*(1), 82–94.

Lee, C.D., & Spratley, A. (2010). *Reading in the disciplines: The challenges of adolescent literacy*. New York, NY: Carnegie Corporation of New York.

Lightman, E., McCarthy, P., Dufty, D., & McNamara, D. (2007). Cohesion and structural organization in high school texts. In *Proceedings of the Florida Artificial Intelligence Research Society International Conference (FLAIRS)* (pp. 235–240). Menlo Park, CA: AAAI Press.

Loban, W. (1976). *Language development: Kindergarten through grade twelve* (NCTE Committee on Research Report No. 18). Urbana, IL: National Council of Teachers of English.

Macaruso, P., & Shankweiler, D. (2010). Expanding the simple view of reading in accounting for reading skills in community college students. *Reading Psychology, 31,* 454–471. https://doi.org/10.1080/02702710903241363

Masterson, J., & Hayes, M. (2004). UK data from 197 undergraduates for the Nelson–Denny Reading Test. *Journal of Research in Reading, 27*(1), 30–35. https://doi.org/10.1111/j.1467-9817.2004.00212.x

Matsuoka, K., Kotani, I., & Yamasato, M. (2012). Correct information unit analysis for determining the characteristics of narrative discourse in individuals with chronic traumatic brain injury. *Brain Injury, 26,* 1723–1730. https://doi.org/10.3109/02699052.2012.698789

McCarthy, P., Graesser, A. C., & McNamara, D. S. (2006, July). *Distinguishing genre using Coh-Metrix indices of cohesion*. Poster presented at the annual meeting of the Society for Text and Discourse, Minneapolis, MN.

McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). *Technical Manual. Woodcock-Johnson IV*. Rolling Meadows, IL: Riverside.

McNamara, D. S., Graesser, A. C., & Louwerse, M. M. (2012). Sources of text difficulty: Across the ages and genres. In J. P. Sabatini & E. Albro (Eds.), *Assessing reading in the 21st century: Aligning and applying advances in the reading and measurement sciences*. Lanham, MD: R&L Education.

McNamara, D. S., & Kendeou, P. (2011). Translating advances in reading comprehension research to educational practice. *International Electronic Journal of Elementary Education, 4,* 33–46.

McVay, J. C., & Kane, M. J. (2012). Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention. *Journal of Experimental Psychology: General, 141*(2), 302–320. https://doi.org/10.1037/a0025250

Meulenbroek, P., Bowers, B., & Turkstra, L. S. (2016). Characterizing common workplace communication skills for disorders associated with traumatic brain injury: A qualitative study. *Journal of Vocational Rehabilitation, 44*(1), 15–31. https://doi.org/10.3233/JVR-150777

National Research Council. (2012). *Improving adult literacy instruction: Options for practice and research*. Lesgold, A. M., & Welch-Ross, M. (Eds.). Washington, DC: The National Academies Press.

Nicholas, L. E., & Brookshire, R. H. (1993). A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech and Hearing Research, 36*(2), 338–350. https://doi.org/10.1044/jshr.3602.338

Ozuru, Y., Best, R., Bell, C., Witherspoon, A., & McNamara, D. S. (2007). Influence of question format and text availability on the assessment of expository text comprehension. *Cognition and Instruction, 25*(4), 399–438. https://doi.org/10.1080/07370000701632371

Pearson, D. P., & Hamm, D. N. (2005). The assessment of reading comprehension: A review of practices—Past, present, and future. In S. G. Paris & S. A. Stahl (Eds.), *Center for improvement of early reading achievement (CIERA). Children's reading comprehension and assessment* (pp. 13–69). Mahwah, NJ: Erlbaum. https://doi.org/10.4324/9781410612762

Pyle, N., Vasquez, A. C., Lignugaris/Kraft, B., Gillam, S. L., Reutzel, D. R., Olszewski, A., . . . Pyle, D. (2017). Effects of expository text structure interventions on comprehension: A meta-analysis. *Reading Research Quarterly, 52*(4), 469–501. https://doi.org/10.1002/rrq.179

RAND Reading Study Group. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND Corporation.

Rapp, D. N., Van Den Broek, P., McMaster, K. L., Kendeou, P., & Espin, C. A. (2007). Higher-order comprehension processes in struggling readers: A perspective for research and intervention. *Scientific Studies of Reading, 11*(4), 289–312. https://doi.org/10.1080/10888430701530417

ReadWorks. (2018). Retrieved from https://www.readworks.org/

Reed, D., & Vaughn, S. (2012). Retell as an indicator of reading comprehension. *Scientific Studies of Reading, 16*(3), 187–217. https://doi.org/10.1080/10888438.2010.538780

Richardson, J. T. E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review, 6*(2), 135–147.

Royer, J. M. (2001). Developing reading and listening comprehension tests based on the sentence verification technique (SVT). *Journal of Adolescent & Adult Literacy, 45*(1), 3–41.

Royer, J. M., Hastings, N. C., & Hook, C. (1979). A sentence verification technique for measuring reading comprehension. *Journal of Reading Behavior, 11,* 355–363.

Royer, J. M., Barbara, A. G., & Sinatra, G. M. (1987). The sentence verification technique: A practical procedure for testing comprehension. *Journal of Reading, 30,* 414–422.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52*(3/4), 591–611.

Sohlberg, M. M., Fickas, S., & Griffiths, G. G. (2011, November). *Reading comprehension strategies delivered via tablet for individuals with acquired brain injury*. Presented at the annual convention of the American Speech-Language-Hearing Association, San Diego, CA.

Sohlberg, M. M., Griffiths, G. G., & Fickas, S. (2014). An evaluation of reading comprehension of expository text in adults with traumatic brain injury. *American Journal of Speech-Language Pathology, 23,* 160–175. https://doi.org/10.1044/2013_AJSLP-12-0005

Sohlberg, M. M., Griffiths, G. G., & Fickas, S. (2015). An exploratory study of reading comprehension in college students after acquired brain injury. *American Journal of Speech-Language Pathology, 24*(3), 358–373. https://doi.org/10.1044/2015_AJSLP-14-0033

Sohlberg, M. M., Kucheria, P., Fickas, S., & Wade, S. L. (2015). Developing brain injury interventions on both ends of the treatment continuum depends upon early research partnerships and feasibility studies. *Journal of Speech, Language, and Hearing Research, 58*(6), S1864–S1870. https://doi.org/10.1044/2015_JSLHR-L-15-0150

Stenner, A. J., Sanford-Moore, E., & Williamson, G. L. (2012). The Lexile Framework for Reading quantifies the reading ability needed for "college and career readiness"(MetaMetrics Research Brief). Durham, NC: MetaMetrics.

Stine-Morrow, E. A., Soederberg Miller, L. M., Gagne, D. D., & Hertzog, C. (2008). Self-regulated reading in adulthood. *Psychology and Aging, 23*(1), 131–153.

Sullivan, M. P., Griffiths, G. G., & Sohlberg, M. M. (2014). Effect of posttraumatic stress on study time in a task measuring four component processes underlying text-level reading. *Journal of Speech, Language, and Hearing Research, 57*(5), 1731–1739.

Todis, B., Glang, A., Bullis, M., Ettel, D., & Hood, D. (2011). Longitudinal investigation of the post-high school transition experiences of adolescents with traumatic brain injury. *The Journal of Head Trauma Rehabilitation, 26,* 138–149.

Towse, J. N., Cowan, N., Hitch, G. J., & Horton, N. J. (2008). The recall of information from working memory: Insights from behavioural and chronometric perspectives. *Experimental Psychology, 55*(6), 371–383.

Unsworth, N., & McMillan, B. D. (2013). Mind wandering and reading comprehension: Examining the roles of working memory capacity, interest, motivation, and topic experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(3), 832–842.

Van Dyke, J. A., Johns, C. L., & Kukona, A. (2014). Low working memory capacity is only spuriously related to poor reading comprehension. *Cognition, 131*(3), 373–403. https://doi.org/10.1016/j.cognition.2014.01.007

Watter, K., Copley, A., & Finch, E. (2017). Discourse level reading comprehension interventions following acquired brain injury: A systematic review. *Disability and Rehabilitation, 39*(4), 315–337.

Weinstein, C. E., Palmer, D. R., & Acee, T. W. (2016). *LASSI user's manual: Learning and study strategies* (3rd ed.). Clearwater, FL: H&H.

Wiederholt, J. L., & Bryant, B. R. (2012). *Gray Oral Reading Test–Fifth Edition: Examiner's record booklet*. Austin, TX: Pro-Ed.

Wilkins, J., & Huckabee, S. (2014). *A literature map of dropout prevention interventions for students with disabilities*. Clemson, SC: National Dropout Prevention Center for Students with Disabilities, Clemson University.

Wolf, L. E. (2001). College students with ADHD and other hidden disabilities. *Outcomes and interventions. Annals of the New York Academy of Sciences, 931,* 385–395. https://doi.org/10.1111/j.1749-6632.2001.tb05792.x

Yeari, M., & Van den Broek, P. (2011). A cognitive account of discourse understanding and discourse interpretation: The landscape model of reading. *Discourse Studies, 13*(5), 635–643.