CrossMark

Understanding the influence of text complexity and question type on reading outcomes

Mercedes Spencer[1] · Allison F. Gilmour[2] · Amanda C. Miller[3] ·
Angela M. Emerson[4] · Neena M. Saha[1] · Laurie E. Cutting[1]

© Springer Nature B.V. 2018

Abstract

In the current study, we examined how student characteristics and cognitive skills, differing levels of text complexity (cohesion, decoding, vocabulary, and syntax), and reading comprehension question types (literal, inferential, critical analysis, and reading strategy) affected different types of reading outcomes (multiple-choice reading comprehension questions, free recall, and oral reading fluency) in a sample of 181 native English-speaking adolescents (9 to 14.83 years). Results from item response theory one-parameter models and multilevel models suggested that different cognitive skills predicted performance across the three reading outcomes. After controlling for student characteristics and cognitive skills, text complexity negatively impacted reading outcomes, particularly oral reading fluency and free recall. Critical analysis and inferential questions emerged as the most difficult types of comprehension questions. The implications of these findings are discussed.

Keywords Item response theory · Reading comprehension · Oral reading fluency · Assessment · Multilevel models

✉ Laurie E. Cutting
laurie.cutting@vanderbilt.edu

1 Department of Special Education, Vanderbilt Kennedy Center, Vanderbilt University, PMB 40, 230 Appleton Place, Nashville, TN 37203, USA

2 Department of Teaching and Learning, Temple University, Philadelphia, USA

3 Department of Psychology and Neuroscience, Regis College, Regis University, Denver, USA

4 Center for Neuromodulation, Stanley D. and Joan H. Ross Center for Brain Health and Performance, Ohio State University, Columbus, USA

Introduction

According to recent statistics, nearly one-fourth of eighth graders in the United States attain reading comprehension (RC) scores below the basic level (National Center for Education Statistics [NCES], 2015). This suggests approximately 25% of eighth graders likely struggle with locating pertinent information within a text, engaging in inference making, and using context clues to infer the meaning of words in context (NCES, 2009). This is problematic given that RC predicts academic achievement and academic achievement is predictive of later student outcomes (Reynolds, Chan, & Temple, 1998; Reynolds, Temple, Robertson, & Mann, 2002). Although RC is fairly consistently operationalized in the literature, this skill can be assessed using a variety of different methods. For instance, the National Assessment of Educational Progress reading assessment contains both multiple-choice questions (MCQs) and open-ended questions (NCES, 2013). Further, RC measures can also include items that have a free recall or a cloze format (see Fuchs, Fuchs, Hosp, & Jenkins, 2001; Garcia & Cain, 2014). This fact is important because the way in which RC is assessed may affect an individual's RC performance (e.g., Keenan & Betjemann, 2006).

Moreover, an individual's ability to comprehend written text is also related to how fluently he/she is able to read the words (i.e., oral reading fluency [ORF]; Fuchs, Fuchs, & Maxwell, 1988; National Institute of Child Health and Human Development, 2000). These strong associations between ORF and RC (Adams, 1990; Buck & Torgesen, 2003; Fuchs et al., 2001) suggest that ORF is not only highly related to an individuals' RC performance but may also serve as a reliable indicator of reading proficiency (Deno, 1985; Fuchs et al., 2001). Thus, a comprehensive assessment and understanding of a reader's proficiency would need to incorporate multiple measures of reading, including measures of RC and fluency, especially for struggling readers (e.g., dyslexia; Shaywitz & Shaywitz, 2005). An additional aspect of written text that may be impacting individuals' reading proficiency is text complexity. Thus, in the current study, we investigated the association between multiple reading outcomes (as indexed by MCQs, free recall, and ORF) and text complexity.

Within the last 6 years, discussions of text complexity have emerged on a national scale with the adoption of Common Core State Standards (CCSS; www.corestandards.org; CCSS Initiative, 2010). A requirement of these standards is that students are expected to have a mastery of complex text by the end of 12th grade (CCSS Initiative, 2010).[1] Although the focus on reading skills in state standards is not new, the consideration of text complexity is an important and interesting addition to these standards, especially given the relative paucity of research examining how aspects of text can impact reading outcomes.

Text complexity and text difficulty are frequently used interchangeably (Hiebert & Mesmer, 2013). However, text complexity refers to specific features within the text that can be manipulated (e.g., syntactic complexity) whereas text

---

[1] Less than half 12th graders would have met or exceeded this benchmark based on the NAEP reading assessment; only 37% of 12th graders scored at or above proficiency in 2015 (NCES, 2015).
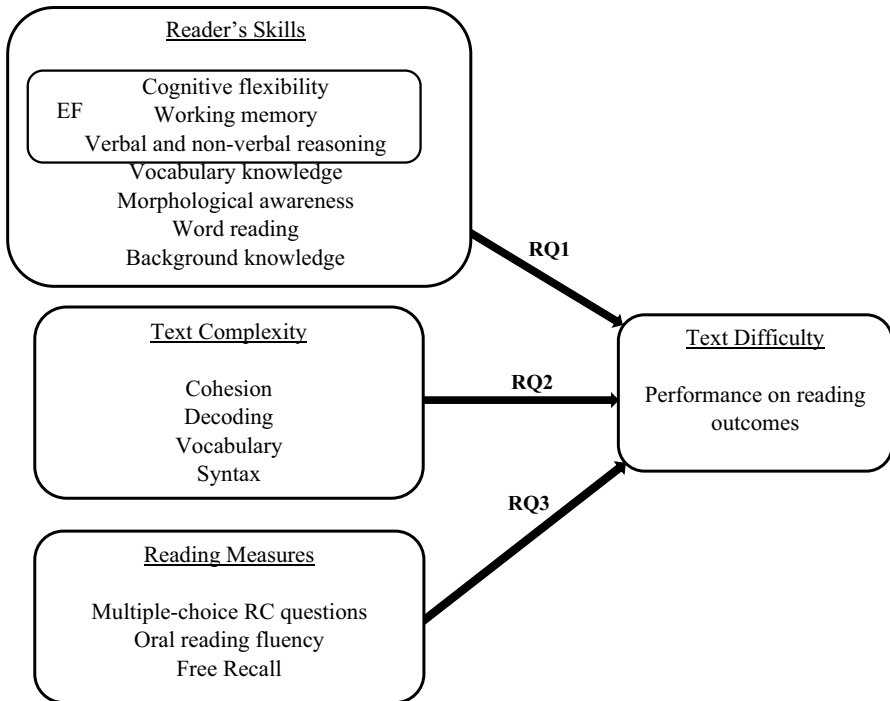
Fig. 1 Relations investigated in the present study. *RC* Reading comprehension, *RQ* research question, *EF* executive function

difficulty refers to the interaction between complexity and reader characteristics and, subsequently, how challenging a text may be to comprehend (e.g., Lexiles; Cunningham & Mesmer, 2014; Mesmer, Cunningham, & Hiebert, 2012; see Fig. 1). Multiple factors can determine text complexity, including syntactic complexity (Frantz, Starr, & Bailey, 2015), novel vocabulary words and/or unfamiliar topics (Freebody & Anderson, 1983), and literal or grammatical cohesiveness (Graesser, McNamara, Louwerse, & Cai, 2004; Halliday & Hasan, 2014; Van Dijk & Kintsch, 1983). The impact of complex texts on both RC and ORF may be due to increased cognitive demands associated with more effortful cognitive processing and a subsequent reduction in automaticity (e.g., decoding unfamiliar vs. familiar words; Benjamin & Schwanenflugel, 2010; Goswami, Gombert, & de Barrera, 1998; Miller et al., 2014; O'Connor et al., 2002). We focus on four aspects of text complexity—decodability, syntax, vocabulary, and cohesion—given that these characteristics are frequently associated with text complexity (see Fitzgerald et al., 2016; Hiebert & Mesmer, 2013; Mesmer et al., 2012).

Text complexity can additionally interact with readers' cognitive skills in ways that are less predictable (e.g., McNamara, Kintsch, Songer, & Kintsch, 1996), likely due to the multitude of skills that are used during reading. For example, McNamara et al. (1996) found that readers with high background knowledge

performed better on passages with low (local and global) cohesion than high cohesion while the opposite was true for readers with low background knowledge. Although one might expect all readers to perform better when reading highly cohesive passages, this counterintuitive finding emphasizes the necessity of examining the interplay between reader- and text-based factors. One potential explanation for this finding is that readers with high background knowledge were better able to actively process passages with low cohesion whereas readers with low background knowledge were unable to bridge these gaps in information. Further, readers with low background knowledge tend to have impaired inference making skills (Elbro & Buch-Iversen, 2013) and would have greater difficulty comprehending texts that did not explicitly include necessary information.

Several decades of research suggest that numerous skills are associated with reading outcomes, including vocabulary, morphological awareness, IQ, phonological awareness, decoding ability, and ORF (e.g., Carlisle 2003; Cutting & Scarborough, 2006; Engen, & Hoien, 2002; Fuchs et al., 2001; Quinn, Wagner, Petscher, & Lopez, 2015; Ready, Chaudry, Schatz, Strazzullo, 2013; Tannenbaum, Torgesen, & Wagner, 2006). Additionally, components of executive function (EF; e.g., working memory [WM], cognitive flexibility, and verbal and nonverbal reasoning; Denckla, 1989), may be associated with RC (e.g., Arrington, Kulesz, Francis, Fletcher, & Barnes, 2014; Sesma, Mahone, Levine, Eason, & Cutting, 2009). We therefore argue that the impact of text complexity on reading outcomes cannot be adequately investigated without first accounting for these cognitive predictors.

Assessing reading skills

In both research and educational practice, reading skills are assessed using a variety of different measures, and it is likely that these measures vary in their complexity and skill demands (see Keenan, Betjemann, & Olson, 2008). Further, it is unlikely that any one assessment comprehensively captures RC (Cutting & Scarborough, 2006; Keenan et al., 2008; Miller et al., 2014). Although various types of RC assessments all purport to measure a single underlying skill, RC assessments vary in several notable ways, including response format (Keenan et al., 2008) and question type (Eason, Sabatini, Goldberg, Bruce, & Cutting, 2013). For instance, Keenan and Meenan (2014) administered three widely-used standardized assessments of RC to 995 8- to 18-year-olds. Correlations across all three measures ranged from .45 to .68, indicating that different assessments were tapping different components of RC. Moreover, the availability of the text and the format of RC questions can also affect performance (Ozuru, Best, Bell, Witherspoon, & McNamara, 2007; Schoeder, 2011), with some highly structured RC questions even exhibiting a tendency for passage independence (Keenan & Betjemann, 2006). Similarly, RC can be assessed using a variety of different formats, including MCQs, free recall, and cloze among others (Garcia & Cain, 2014), all of which can impact the types of cognitive skills taxed to successfully complete the task (e.g., Best, Floyd, & McNamara, 2008). Given the variability associated with these assessments, it remains important to

investigate the extent to which different reading outcomes may be addressing similar versus dissimilar cognitive skills.

Such variability among assessments is potentially problematic if one considers that RC and ORF assessments are used for progress monitoring (e.g., Deno, 1989; Hosp & Fuchs, 2005; Shinn, 1989) and as a means of identifying at-risk and/or struggling readers (e.g., Catts, Fey, Zhang, & Tomblin, 2001; Nation & Snowling, 1998; National Reading Panel, 2000; Wexler, Vaughn, Edmonds, & Reutebuch, 2008). Subsequently, the interplay of text complexity and outcome type can potentially lead to an over- or under-identification of reading difficulties. As a consequence, over-identification can result in unnecessary expenditures of both time and effort for targeted remediation services that may not be needed whereas under-identification can result in an absence of much needed early identification/remediation services and reading problems that are more difficult to remediate later on (Lyon, 1998; National Reading Panel, 2000). Additionally, the success of reading intervention programs are evaluated using both experimenter-created and standardized RC and ORF assessments, resulting in even greater variability (Fuchs, Fuchs, & Kazdan, 1999; Solis et al., 2012).

Text complexity and RC

Multiple theories of RC aim to account for factors that likely impact the RC process. For example, Kintsch's (1988) construction–integration (C–I) model and Perfetti's (1999) Reading Systems Framework (RSF) both assert that RC is affected by bottom-up text-based information *and* top-down reader-based characteristics. According to these frameworks, RC can either be degraded or supported by such processes because these influences are not contained within a single domain (e.g., linguistic and semantic components [C–I model] and the phono-orthographic and lexical areas [RSF]). Similarly, student-level characteristics map on to various components. For instance, IQ, non-verbal reasoning, and background knowledge fall within global (C–I model) or general knowledge (RSF). Verbal reasoning, vocabulary, and morphological awareness are contained within the semantic (C–I model) and linguistic system knowledge levels (RSF). Inference making, cognitive flexibility, and WM map onto integration (C–I model) and comprehension processes (RSF).

Although such frameworks acknowledge the contributions of both reader- and text-based factors to the RC process, few empirical investigations have focused explicitly on this interplay within the scope of how reading is being assessed. Three previous studies lay the groundwork for the present investigation. Eason, Goldberg, Young, Geist, and Cutting (2012) investigated the relations between reader-based characteristics and features of a RC assessment in 126 children. Results indicated that functional texts were the least difficult to comprehend while critical analysis and process strategies questions were the most difficult. Additionally, children with the greatest inference making and planning/organizing skills attained higher scores on more complex texts and more difficult question types.

In another study, Miller and colleagues (2014) examined the effect of various text-based characteristics on RC performance for 94 children. Results indicated that

children's reading speed was substantially slower for more difficult passages and that inferential and reading strategy questions were more difficult than literal questions. However, the authors did not explicitly investigate the relation between text complexity and RC performance. In a more recent investigation, Kulesz, Francis, Barnes, and Fletcher (2016) examined the impact of various features of RC passages from the Gates-MacGinitie Reading Tests for 1,190 adolescents. The findings indicated that genre (narrative vs. expository) was the strongest predictor of text difficulty and that background knowledge and vocabulary were the strongest predictors of RC.

Text complexity and ORF

Similar to RC, there exist several theories of word recognition that acknowledge the influence of multiple sources of information. Three prominent models of word recognition – the triangle model (Seidenberg & McClelland, 1989), the dual-route cascaded model (DRC; Coltheart, Rastle, Perry, & Ziegler, 2001), and the connectionist dual process model (CDP; Zorzi, Houghton, & Butterworth, 1998)—each argue that word reading is affected by textual features and/or reader characteristics. Verbal reasoning, vocabulary, and morphological awareness fall within orthographic, phonological, and semantic knowledge (triangle model) and the lexical route (DRC and CDP). Inference making and cognitive flexibility maps onto the integration of multiple sources of information (triangle, DRC, and CDP) as well as the activation (DRC) and bootstrapping of word representations (triangle).

Empirical evidence suggests that text complexity also impacts ORF (e.g., Benjamin & Schwanenflugel, 2010). In a recent review, Amendum, Conradi, and Hiebert (2017) examined the influence of text complexity to both RC and reading fluency across 26 empirical studies. Across studies, text difficulty was based on a variety of factors, including cohesion, decodability, word frequency, and vocabulary, among others. Seventy-three percent of the studies reviewed (11 out of 15) indicated that as text complexity increased, reading rate decreased (no relationship was found for the other four studies). Similarly, 54% of studies (seven out of 13) suggested that increasing text complexity resulted in decreased RC (no relationship was found for five studies; optimum difficulty level was found for 1 study). Overall, the pattern of results indicated that increasing the complexity of the text resulted in a decrease in reading fluency in addition to RC. However, the authors acknowledged that these relations may be sensitive to specific types of text complexity manipulations, indicating that this is an area that needs future investigation.

The current study

There are two gaps in the literature that we aim to address in the present study: (a) the impact of text complexity on reading ability in the context of multiple cognitive skills, and (b) the impact of RC question type across identical passages. Thus, we use item response theory (IRT) and multilevel models to further elucidate the

relations among various text- and student-based characteristics. The current investigation was guided by three research questions (RQs):

RQ1	What are the relations between students' cognitive skills and performance on MCQs, ORF, and free recall over and above sex, age, and IQ? Further, how consistent are students' reading scores across measures (free recall, MCQs, and ORF)?

RQ2	Does text complexity predict reading outcomes over and above student characteristics and cognitive skills?

RQ3	Do question types differ in difficulty over and above text complexity (cohesion, decoding, vocabulary, and syntax) and student characteristics and cognitive skills?

We build upon previous investigations in three important ways. First, we include a rich battery of language, literacy, and EF assessments within a single model in order to account for as many cognitive skills as possible that could influence the association between text complexity and RC. Second, we examine the associations between student characteristics and cognitive skills across three measures of reading using identical passages. This approach minimizes the variability across passages, and thus, can further elucidate whether different measures of reading rely on different cognitive skills. Third, in addition to accounting for age (e.g., Eason et al., 2012, Kulesz et al., 2016, and Miller et al., 2014), we also include sex, IQ, socioeconomic status (SES), and background knowledge within our models. The rationale for this approach is that (a) sex differences in reading do exist (see Klecker, 2006), (b) IQ and SES are related to numerous skills examined in the present investigation (e.g., Bowey, 1995; Engle, Tuholski, Laughlin, & Conway, 1999; Hackman & Farah, 2009; Oakhill, Cain, & Bryant, 2003), and (c) background knowledge is often cited as a strong predictor of reading outcomes (e.g., Priebe, Keenan, & Miller, 2012). Further, IQ and SES were not considered in previous investigations (e.g., Eason et al., 2012; Kulesz et al., 2016; Miller et al., 2014).

Method

Participants

The current study took place in an urban area within the southeastern United States. We recruited participants through advertisements in schools, clinics, and doctors' offices. The original sample included 226 students. In the current study, we report data from students with complete data only (missing data ranged from 0.5 to 7.5% across covariates, and we excluded one student with an out-of-range score on morphological awareness). The final analysis sample included 181 native English-speaking adolescents ranging from 9 to 14.83 years ($M = 11.65$, $SD = 1.39$) with a broad range of reading abilities. Graduate-level research assistants and/or staff members assessed students across 2 days. The sample was 71.82% Caucasian, 11.05% African

American, 2.21% Asian, and 4.97% multi-racial; 9.94% did not specify. Half were male (50.28%).

Participants were excluded if they had: (a) previous diagnosis of intellectual disability; (b) known uncorrectable visual impairment; (c) treatment of any psychiatric disorder (other than ADHD) with psychotropic medications; (d) history of known neurologic disorder; (e) documented hearing impairment greater than or equal to a 25 dB loss in either ear; (f) known full-scale IQ below 80 prior to testing, or a score below 70 on either performance or verbal scales using the Wechsler Abbreviated Scale of Intelligence 4th edition (WASI; Wechsler, 1999) determined after enrollment; and/or (g) the history of or presence of a pervasive developmental disorder. Children with ADHD who were treated with medications other than stimulants were excluded. These criteria allowed us to reduce the likelihood of comorbidities that may affect reading ability.

Measures

Student variables

We included 11 student-level variables related to RC (see Table 1). We describe each of these measures in greater detail below.

Word reading We measured word reading with the *Test of Word Reading Efficiency* (TOWRE; Torgensen, Wagner, & Rashotte, 1999). Participants had 45 s to read aloud a list of real words and 45 s to read a list of non-words. Test–retest coefficients range from 0.83 to 0.96 for all ages (Torgensen et al. 1999).

Vocabulary We measured vocabulary using three subtests from the *Test of Word Knowledge* (TOWK; Wiig & Secord, 1992): Expressive and Receptive Vocabulary and Synonyms. For Expressive Vocabulary, participants saw a picture and described the picture using one word. For Receptive Vocabulary, participants saw a set of pictures and selected the picture that best represented the orally-presented word. The Synonyms subtest required children to identify a word from a list of four printed words that was most like the stimulus word. Test–retest coefficients for 9–12-year-olds were reported as .96 (expressive), .87 (receptive), and .86 (synonyms).

Morphological awareness We used the test of Morphological Relatedness adapted from Mahony, Singson, and Mann (2000). Participants determined if a derived word was related to a root word. For each item, participants heard the word and saw it in print. Participants scored one point for each correct answer across 12 items. In our sample, $\alpha = .60$ and Spearman-Brown split-half reliability was .83.

Inferential reasoning We measured cohesion and elaborative inferencing using the *Know-It* task (Barnes, Dennis, & Haefele-Kalvaitis, 1996). Participants listened to a seven-part narrative story and answered two questions following each story that required them to make elaborative (gap-filling) and cohesive inferences (text-connecting; see Hall, 2016). Participants scored one point for each correct response across 14 items. Scores were calculated for elaborative and cohesion questions separately. In our sample, $\alpha = .61$ (cohesive) and .64 (elaborative). Both

Table 1 Descriptive statistics across measures

Level	Type	Variable	Mean	SD	Min.	Max.
Response	Outcome					
(R = 17,259)		MCQs[a]	0.80	0.40	0	1
(R = 2,142)		ORF[b]	132.27	40.25	32.89	264.47
(R = 2,150)		Free recall[c]	29.85	16.80	0	90
Student (J = 215)						
	Covariate					
		Sex (*Male N = 107*)[a]	0.50	0.50	0	1
		Age	11.87	1.32	9.92	14.83
		IQ[d]	107.53	15.02	67	140
		SES	49.66	7.77	26	62
		Non-verbal reasoning[d]	108.93	13.83	66	151
		Word reading[d]	100.55	11.96	72	134
		Vocabulary[e]	9.64	3.20	4	17
		Cohesive inferencing[f]	4.62	1.66	0	7
		Elaborative inferencing[f]	3.48	1.83	0	7
		Cognitive flexibility[e]	9.63	2.74	1	16
		Working memory[f]	6.29	2.30	0	12
		Morphological awareness[f]	9.62	1.21	3	12
		Background knowledge[c]	4.53	5.22	0	53.85
		Verbal reasoning[d]	10.59	3.14	1	16

SD Standard deviation, *Min* Minimum, *Max* Maximum, *MCQs* Multiple-choice questions, *ORF* Oral reading fluency

[a]Dichotomous; [b]words read per minute; [c]percentage; [d]standard score; [e]scaled score; [f]raw score

subtests were double-coded for 20% of our sample. Inter-rater agreement was 96.11% (cohesive) and 94.8% (elaborative).

Cognitive flexibility We used the Card Sorting subtest of the *Delis Kaplan Executive Function System* (DKEFS; Delis, Kaplan, & Kramer, 2001) to measure verbal and nonverbal cognitive flexibility. Participants were required to sort cards into two groups in as many different ways as they could (up to 36 sorts across two conditions and sets). Sorts were counted as correct if they fell into one of the sorting descriptions for that set of cards. Repeat sorts were not counted. Reported internal consistency for this subtest ranges from .55 to .82 depending on age. Test–retest reliability is .49.

Verbal reasoning We measured verbal abstract reasoning and concept formation with the 20 Questions subtest from the DKEFS (Delis et al., 2001). Participants looked at a page of pictures and guessed which picture the examiner picked by asking the fewest number of yes/no questions as possible across four trials. Internal consistency for this subtest ranges from .72 to .87 depending on age. Reported test–retest reliability coefficients are low (Delis et al., 2001), which may be due to learning effects (Homack, Lee, & Riccio, 2005).

Nonverbal reasoning We used the Analysis-Synthesis subtest from the *Woodcock-Johnson-III* (Woodcock, McGrew, & Mather, 2001) to assess nonverbal abstract reasoning. The participant was required to solve a number of problems using a series of "keys", made up of combinations of colored square boxes to solve the puzzles. Correct answers were given one point across 35 items. Reported test–retest coefficients were over 0.88 for all ages (Woodcock et al., 2001).

WM We measured verbal WM based on the *Sentence Span Task* (Daneman & Carpenter, 1980). Participants were asked to listen to sentences, decide whether the sentence was true or false, then to recall the last word of the sentence. The number of sentences increased with each trial and consisted of three spans across four sets. Each correct answer was given one point across 42 items. Administration was discontinued if/when the participant missed three spans. Reported internal consistency estimates range from .61 to .76 depending on the length of the sentence span, and the test–retest correlation is .44 (Waters & Caplan, 1996).

Background knowledge We assessed background knowledge by asking participants to tell the tester everything they knew about each passage topic prior to reading. Responses were recorded and later scored using an idea checklist (inter-rater agreement = 76.74% for 20% of participants averaged across passages). Participants received one point for each answer they gave that mapped onto an idea unit from the passage. Background knowledge was scored as a percentage of the total idea units possible for each passage. In our sample, α = .88.

IQ We assessed IQ with the WASI (Wechsler, 1999). Reliability is reported as .93.

SES SES was assessed using the Hollingshead Four-Factor Index (Hollingshead, 1975), which is a commonly used continuous measure of SES (Edwards-Hewitt & Gray, 1995) and is based on parents' educational and occupational information.

Outcomes

All participants read a series of experimental passages and then completed three assessments: MCQs, ORF, and free recall. These measures are described in greater detail below.

Passages Ten baseline passages included manipulations for cohesion, decoding, syntax, and vocabulary as a means of increasing text complexity and resulted in 18 total passages (see Appendices A, B and C). We manipulated passages using Coh-Metrix (Graesser et al., 2004) and bootstrapping with 90% confidence intervals (CI) for cohesion, vocabulary, and syntax manipulations. The decodability of passages (pre- and post-manipulation) was examined using the Decoding System Measure (DSyM; Cutting, Saha, & Hasselbring, 2017) and bootstrapping with 90% CI. Briefly, the DSyM quantifies several sub-lexical features of words, such as frequency of grapheme to phoneme mappings, and assigns a score based on decoding difficulty. Preliminary evidence (unpublished data) has shown that the DSyM score explains unique variance in children's word reading fluency both on individual words as well as longer text, such as passages. All baseline passages were within the 90% CI during the bootstrapping procedure, with the exception of one (Sap and Syrup), which fell below the 90% CI (i.e., was easier to decode). Similarly, all

decoding manipulated passages fell at or above the 90% CI (regardless of whether the Sap and Syrup passage was included in the analysis).

Passages were piloted on adults to ensure increased difficulty. Order of administration and passage manipulated was counter-balanced across 10 lists. Baseline passages were matched on word and sentence length, word frequency, and text complexity using Coh-Metrix. We designed our passages to be on obscure topics to reduce participants' reliance on background knowledge. All passages were expository because (a) text genre is known to influence comprehension (e.g., Eason et al., 2012), and we wanted to control for this, and (b) the use of expository texts with participants in this age range increases our study's external validity given that expository texts are common in academic learning.

MCQs Participants answered 7 to 11 MCQs about each passage (81 MCQs total). Participants were allowed to reference the passage but received no explicit instructions to do so. Literal questions (29 items) measured participants' understanding of concepts directly stated in the passage. Reading strategy questions (12 items) measured participants' ability to identify text-based characteristics and strategy use. Inferential questions (31 items) measured participants' ability to make inferences and/or predictions, and/or draw conclusions beyond what was explicitly directly stated. Critical analysis questions (9 items) measured participants' ability to determine the author's purpose, recognize types of text, and determine patterns of text organization. In our sample, overall $\alpha = .94$. By question type, literal $\alpha = .81$, inferential $\alpha = .84$, reading strategy $\alpha = .70$, and critical analysis $\alpha = .60$.[2]

ORF ORF was operationalized as the number of words read per minute. Participants were told to read passages aloud and that they should read at a pace that allows them to also comprehend the text. We recorded the number of seconds participants took to read each text and divided by the number of words in that text. In our sample, $\alpha = .99$.

Free recall After the participant read each passage, the examiner asked the participant to verbally recall that passage. Participants could not reference the passages and there was no time delay or masking between the reading of the passage and recall. We divided the text into meaningful idea units and created a checklist of these idea units and used this to calculate the percentage of total unweighted idea units recalled for each passage (i.e., memory of the text's microstructure; Kintsch, 1988). Inter-rater agreement was 87.94%. In our sample, $\alpha = .98$.

Analytic approach

We first examined the pairwise correlation coefficients for scores on the three reading outcomes. We created one score each for MCQs, ORF, and free recall by averaging participants' performance for each measure across the 10 reading passages. These scores were unadjusted for text complexity, student characteristics, and question type. We then specified multilevel models for free recall and ORF (continuous)

---

[2] Alpha is directly related to the number of items; thus, lower reliability for some question types was likely due to the small number of items.

and explanatory cross-classified random-effects one-parameter IRT models (Rasch models) for MCQs (dichotomous). These data are cross-classified because the outcomes are both nested in students (i.e., multiple outcomes for each student) and passage topics. Some students read a baseline passage without any manipulations while others read the same passage with manipulations. The outcomes from these different students are therefore nested within the passage topic. Models included dummy variables for passage manipulations, mean-centered student characteristics, and item type (MCQs model only). Each type of question was dummy coded with literal questions as the reference group. Question type was used as an item-level covariate.

MCQs

We used IRT models (van den Noortgate, de Boeck, & Meulders, 2003) to evaluate how item, passage, and student covariates related to RC difficulty. These models result in an outcome, the latent trait of RC, on the logit scale. A positive or negative coefficient for a passage manipulation would suggest that the manipulation resulted in the passage being easier than the baseline passage or more difficult than the baseline passage, respectively. Prior to IRT modeling, we ensured the unidimensionality of our data by using principal component analysis (PCA). We established the need for random effects by running a series of unconditional models with random intercepts for students, items, and passage topic. We compared the fit of these models using the AIC, BIC, and change in log-likelihood (LL) criteria. We fit all models with the "glmer" function in R (Bates, Maechler, Bolker, & Walker, 2015).

ORF and free recall

We used the approach described above to establish the need for random effects. We then fit a series of multilevel models for each outcome using the "lmer" function in R (Bates et al., 2015). Cross-classified models have unknown distributional characteristics resulting in biased standard errors (Meyers & Beretvas, 2006). We reported simulation based 95% confidence intervals using the "boot" function (Canty & Ripley, 2015) with 1000 replications.

Results

Descriptive statistics are presented in Table 1. Correlations across assessments were small to moderate (see Table 2). The results of the PCA suggested that the MCQ results were unidimensional and supported the use of an IRT model. We included random effects for items and students in the MCQ models because the addition of a random effect for passage topic did not improve the model fit $(\Delta\chi^2[1] = 0, p = .998)$. Including the passage topic and student random effects improved the model fit for ORF over models including only the passage topic random effect $(\Delta\chi^2[1] = 4940.1, p < .001)$ or only the student random effect $(\Delta\chi^2[1] = 209.92, p < .001)$. Similarly, including passage topic and student random effects improved the fit of models addressing free recall beyond only including

Table 2 Correlations between student-level predictors

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Sex	–													
2. Age	.111	–												
3. IQ	–.045	**–.134**	–											
4. SES	.021	–.067	**.431**	–										
5. Non-verbal reasoning	.011	–.057	**.648**	**.315**	–									
6. Word reading	–.067	.053	**.448**	**.164**	**.327**	–								
7. Vocabulary	.075	**–.175**	**.647**	**.267**	**.374**	**.425**	–							
8. Cohesive inferencing	–.075	–.044	**.440**	**.190**	**.304**	**.329**	**.366**	–						
9. Elaborative inferencing	–.038	.006	**.397**	.144	**.311**	**.278**	**.405**	**.532**	–					
10. Cognitive flexibility	**–.185**	–.068	**.573**	**.241**	**.427**	**.286**	**.393**	**.287**	**.279**	–				
11. Working memory	.124	**.226**	**.497**	**.182**	**.365**	**.418**	**.370**	**.325**	**.320**	**.334**	–			
12. Morphological awareness	.028	.049	**.454**	**.153**	**.347**	**.222**	**.260**	**.261**	**.220**	**.270**	**.249**	–		
13. Background knowledge	.032	.028	**.138**	.092	.043	**.201**	**.207**	**.134**	.087	-.019	**.143**	.072	–	
14. Verbal reasoning	.030	–.087	**.427**	**.332**	**.420**	**.216**	**.349**	**.293**	**.226**	**.357**	**.226**	**.049**	.126	–

Correlations significant at the $p < .05$ level are bolded

passage random effects ($\Delta\chi^2[1] = 1855.9, p < .001$) or student random effects ($\Delta\chi^2[1] = 119.25, p < .001$).

RQ1: student-level predictors across three reading outcomes

Correlations between MCQs and free recall ($r = .66$), MCQs and ORF ($r = .68$) and free recall and ORF ($r = .56$) were moderate and statistically significant at $p < .05$. Next, we examined which student-level cognitive skills significantly predicted performance on MCQs, ORF, and free recall and whether this varied across the three types of RC measures (see Tables 3 and 4). Age and IQ were significant predictors of performance on MCQs, ORF, and free recall.

We also included several additional predictors of reading outcomes within these models. As shown in Table 3, students who scored higher on word reading, vocabulary, or cohesive inferencing had higher average scores on MCQs after controlling for other text and student characteristics. Word reading and cohesive inferencing skills were also associated with average increases in ORF (Table 4). Only WM was associated with an average increase in free recall after accounting for other student and passage characteristics (Table 5).

RQ2: manipulated text complexity

MCQs

After controlling for student and item characteristics, cohesion manipulated passages were more difficult than baseline passages (see Table 3). On average, students scored 0.28 logits lower on RC when the passage had a cohesion manipulation than on baseline. None of the other passage manipulations were more difficult than baseline.

ORF

All passage manipulations were associated with a decrease in words read per minute after controlling for student characteristics. Cohesion manipulated passages resulted in an average decrease of 7.06 words read and decoding manipulated passages resulted in a decrease of 7.68 words read (see Table 4). Vocabulary manipulation resulted in an average decrease of 11.11 words read. Syntax manipulation resulted in a decrease of 3.43 words read.

Free recall

Free recall scores also decreased due to text complexity (see Table 5). Cohesion, syntax, and vocabulary manipulations were negatively associated with free recall. Cohesion manipulations resulted in an average decrease of 3.30 percentage points in the percentage of idea units recalled, syntax manipulated was associated with a 2.21

Table 3 Explanatory item response theory results for multiple-choice questions

	Null model			Full model			
	Est.	SE	Z	Est.	SE	Z	β
Fixed Effects							
Grand mean intercept	**1.862**	**0.135**	**13.810**	**2.410**	**0.192**	**12.542**	**2.410**
Passage covariates							
Cohesion				**− 0.281**	**0.107**	**− 2.624**	**− 0.281**
Decoding				− 0.108	0.099	− 1.099	− 0.108
Syntax				− 0.020	0.109	− 0.179	− 0.020
Vocabulary				− 0.141	0.113	− 1.240	− 0.141
Question covariates							
Critical analysis				**− 0.911**	**0.361**	**− 2.514**	**− 0.911**
Reading strategy				− 0.461	0.327	− 1.440	− 0.472
Inferential				**− 0.700**	**0.247**	**− 2.837**	**− 0.699**
Student covariates							
Sex				− 0.104	0.103	− 1.014	− 0.104
Age				**0.348**	**0.040**	**8.787**	**0.461**
IQ				**0.021**	**0.006**	**3.381**	**0.310**
SES				0.004	0.007	0.619	0.034
Non-verbal reasoning				0.003	0.005	0.774	0.049
Word reading				**0.010**	**0.005**	**2.045**	**0.121**
Vocabulary				**0.096**	**0.022**	**4.432**	**0.306**
Cohesive inferencing				**0.074**	**0.035**	**2.098**	**0.122**
Elaborative inferencing				0.002	0.032	0.052	0.003
Cognitive flexibility				0.011	0.023	0.486	0.030
Working memory				0.030	0.027	1.101	0.069
Morphological awareness				0.075	0.045	1.665	0.090
Background knowledge				− 0.002	0.006	− 0.336	− 0.008
Verbal reasoning				0.011	0.020	0.560	0.034
Random Effects: variances							
Intercept (Item)	0.968			0.851			
Intercept (Student)	1.171			0.295			
Fit statistics							
AIC	14,143.1			11,397.1			
BIC	14,166.4			11,579.1			
LL (*df*)	− 7068.5 (3)			− 5674.5 (24)			

Est. Estimate, *SE* standard error, *AIC* akaike information criterion, *BIC* Bayesian information criterion, *LL* Loglikelihood, *df* degrees of freedom

Coefficients are in logits on an easiness scale. Estimates significant at the $p < .05$ level are bolded. A random effect is not included for passage topic because the inclusion of the random effect did not improve model fit

Table 4 Multilevel model results for oral reading fluency (ORF)

	Null model		Full model		
	Est.	95% CI	Est.	95% CI	β
Fixed effects					
Intercept	**132.100**	**126.907, 137.871**	**138.918**	**133.667, 143.974**	**0.165**
Passage covariates					
Cohesion			**−7.062**	**−8.874, −5.234**	**−0.175**
Decoding			**−7.680**	**−9.608, −5.836**	**−0.191**
Syntax			**−3.427**	**−5.388, −1.503**	**−0.085**
Vocabulary			**−11.108**	**−12.979, −9.314**	**−0.276**
Student covariates					
Sex			−5.214	−11.878, 1.512	−0.130
Age			**12.694**	**10.295, 15.279**	**0.418**
IQ			**0.397**	**0.037, 0.772**	**0.148**
SES			−0.022	−0.464, 0.432	−0.004
Verbal reasoning			0.201	−1.071, 1.370	0.016
Non-verbal reasoning			−0.168	−0.455, 0.124	−0.058
Word reading			**1.940**	**1.660, 2.235**	**0.578**
Vocabulary			0.590	−0.648, 1.985	0.047
Morphological awareness			−0.686	−3.611, 2.055	−0.021
Cohesive inferencing			**2.376**	**0.119, 4.683**	**−0.176**
Elaborative inferencing			−0.221	−2.323, 1.703	−0.010
Cognitive flexibility			−0.358	−1.670, 1.188	−0.024
Working memory			0.672	−0.881, 2.350	0.038
Background knowledge			−0.043	−0.144, 0.065	−0.004
Random Effects: Variances					
Intercept (Student)	1526.08	–	410.47	–	0.253
Intercept (Topic)	12.67	–	16.33	–	0.010
Residual	95.52	–	86.17	–	0.053
Fit statistics					
AIC	16,969.85		13,978.28		
BIC	16,992.53		14,093.86		
LL (*df*)	−8480.926 (4)		−6968.138 (21*)*		

CI Confidence interval, *β* standardized coefficient, *AIC* akaike information criterion, *BIC* Bayesian information criterion, *LL* loglikelihood, *df* degrees of freedom

Estimates excluding 0 in the 95% CI are bolded

Table 5 Multilevel model results for free recall

	Null model		Full model		
	Est.	95% CI	Est.	95% CI	β
Fixed effects					
Intercept	**29.855**	**27.456, 32.237**	30.635	27.580, 33.786	0.046
Passage covariates					
Cohesion			**−3.301**	**−5.144, −1.540**	**−0.197**
Decoding			−0.536	−2.547, 1.463	−0.032
Syntax			**−2.210**	**−4.151, −0.249**	**−0.132**
Vocabulary			**−2.845**	**−4.696, −0.881**	**−0.169**
Student covariates					
Sex			0.463	−2.524, 3.702	0.028
Age			**2.729**	**1.563, 3.835**	**0.216**
IQ			**0.221**	**0.041, 0.403**	**0.198**
SES			0.110	−0.098, 0.310	0.051
Verbal reasoning			−0.179	−0.819, 0.412	−0.034
Non-verbal reasoning			0.035	−0.114, 0.164	0.029
Word reading			0.131	−0.020, 0.284	0.094
Vocabulary			0.535	−0.105, 1.154	0.102
Morphological awareness			0.607	−0.799, 1.992	0.044
Cohesive inferencing			0.355	−0.816, 1.491	0.035
Elaborative inferencing			0.573	−0.409, 1.599	0.062
Cognitive flexibility			0.382	−0.301, 1.068	0.062
Working memory			**0.888**	**0.041, 1.694**	**0.122**
Background knowledge			0.051	−0.062, 0.159	0.012
Random Effects: Variances					
Intercept (Student)	192.32		89.74		0.318
Intercept (Topic)	6.63		10.16		0.036
Residual	84.81		88.48		0.313
Fit statistics					
AIC	16,359.80		13,817.81		
BIC	16,382.49		13,933.45		
LL (*df*)	−8175.898 (4)		−6887.907 (21)		

CI Confidence interval, *β* standardized coefficient, *AIC* akaike information criterion, *BIC* Bayesian information criterion, *LL* loglikelihood, *df* degrees of freedom

Estimates excluding 0 in the 95% CI are bolded

percentage point decrease, and vocabulary was associated with a 2.85 percentage point decrease.

RQ3: difficulty of RC question types

Critical analysis and inferential questions were significantly more difficult than literal questions after controlling for text complexity and student characteristics (see Table 3). Critical analysis questions were associated with a −0.91 logit change in RC. Inferential questions were associated with a −0.70 logit change in RC.

Discussion

In the present study, we investigated the relationship between text complexity and reading outcomes in the context of students' cognitive skills. The importance of examining text complexity is highlighted by the recent implementation of many reading-focused benchmark assessments across states in the United States that focus on student ability to read increasingly complex text (e.g., Common Core). Nonetheless, there continues to be relatively little discussion as to how text complexity may impact reading. This is a practical issue because benchmark assessments and measures used to identify children who meet criteria for reading difficulties can vary greatly in terms of the complexity of their passages.

Student-level predictors

Results indicate that textual features are but a single component of the RC process; reader-based characteristics significantly impact RC (Eason et al., 2012; Kintsch, 1988; Miller & Keenan, 2009; Perfetti, 1999). The interactions between reader- and text-based characteristics are frequently complex and are not often examined comprehensively. In our study, however, we investigated the associations between a myriad of cognitive skills and reading outcomes while controlling for the effects of age, sex, IQ, text complexity, and type of RC questions.

Results indicated that after accounting for passage-level covariates, performance on MCQs was predicted by students' performance on measures of word reading, vocabulary, and cohesive inferencing. These findings are important because they suggest that decoding, vocabulary, and syntax manipulations do not substantially impact students' performance on average while individual differences in cognitive skills do predict performance (i.e., passage characteristics are less important for MCQs relative to students' skills). This specific pattern makes sense for a variety of reasons. First, word reading is essential for successful RC (LaBerge & Samuels, 1974); effortful decoding is detrimental because it interrupts a reader's mental model of the text (e.g., Perfetti, 1985). Second, language skills have been identified as robust predictors of RC, both theoretically (Kintsch, 1988; Perfetti, 1999; Hoover & Gough, 1990) and empirically (e.g., Ouellette, 2006). For instance, the RSF (Perfetti, 1999) asserts that a reader's vocabulary

(lexicon) is a bridge between bottom-up (word reading) and top-down processes (higher-level comprehension) and that a reader's lexical knowledge plays a central role in RC (see also Kintsch, 1988). Third, cohesive inferencing likely emerged as a significant predictor based on the nature of our RC assessment. Most questions required students to make inferences (inferential), indicate how the text was organized or structured (reading strategy), and/or determine patterns of text organization (critical analysis), all of which tap some aspect of cohesive inferencing.

For free recall, cohesion, syntax, and vocabulary manipulations negatively affected students' ability to recall the passages, on average, and WM was the only cognitive skill that predicted free recall after accounting for text complexity and other student characteristics. Therefore, passage characteristics were more important for students' recall compared to individual differences in their cognitive skills. The outcome that WM predicted free recall was somewhat expected given the high reliance on WM for verbal recall (Baddeley & Hitch, 1974). The finding that ORF was predicted by word reading and cohesive inferencing is explained by the fact that efficient word recognition is critical for ORF (National Reading Panel, 2000) and also the idea that ORF may represent automatic inference making during reading (Thurlow & Van den Broek, 1997 as cited in Fuchs et al., 2001). Similar to free recall, passage-level characteristics also substantially impacted students' ORF. Given the amount of residual variance that was present, however, it is important to note that the inclusion of other student-level variables (e.g., listening comprehension) may account for additional variance in reading outcomes. It is important to note that students' ORF performance may have been affected by their awareness of follow-up RC questions (English, Barnes, Fletcher, Dennis, & Raghubar, 2010; van den Broek, Lorch, Linderholm, & Gustafson, 2001).

Although it may seem surprising that background knowledge was not a significant predictor of any reading outcome (e.g., Barnes et al., 1996; Cain, Oakhill, & Bryant, 2004; McNamara et al., 1996), this finding was anticipated because we created passages on unfamiliar topics to decrease students' reliance on background knowledge. We must acknowledge, however, that this pattern may also be due to the fact that our measure of background knowledge exhibited some floor effects. Nonetheless, the finding that background knowledge did not predict reading outcomes is interesting because it allowed us to better identify which cognitive skills predicted reading outcomes aside from background knowledge, and may explain why some commonly-associated reading-related skills did not significantly predict reading outcomes (e.g., cognitive flexibility).

Another interesting outcome was that SES and sex were not significant predictors across any of the three models. However, this pattern of results may be due to the fact that (a) a substantial portion of the sample (86%) came from upper-middle class backgrounds and (b) several student-level predictors included within the models are correlated with SES (e.g., IQ; von Stumm & Plomin, 2015). Furthermore, in the current study, sex was significantly correlated with only cognitive flexibility, which was a nonsignificant predictor across all models. In other words, this observed lack of association between sex and other predictors of reading may explain why sex did not emerge as a significant predictor in this sample.

Our findings also build on the current literature of RC assessments in a novel way. Previous studies have shown that different measures of RC tax different cognitive skills (see Keenan et al., 2008 and Cutting & Scarborough, 2006); however, these studies have examined this phenomenon across different measures of RC, and subsequently, different passages. As a result, previous approaches included additional variability across reading measures that may have affected the observed differences. A unique aspect of the current investigation is that three types of reading outcomes were measured across the *same* passages. Thus, the current results strengthen the argument that different types of reading tasks are truly relying on different cognitive skills.

Question- and passage-level predictors

Results indicated that the three measures were only moderately correlated ($r = .56–.68$), suggesting that the different measures were, in fact, assessing relatively separable components of reading. It is important to acknowledge, however, that the magnitude of these correlations were likely also affected by the psychometric properties of the assessments (e.g., imperfect reliabilities). We found that each of the manipulated passages was more difficult than the baseline passage and that critical analysis and inferential comprehension questions were more difficult than literal questions. We also found some evidence that different cognitive skills were differentially predictive of performance across the three measures. For instance, word reading and cohesive inferencing predicted performance for MCQs and ORF only. On the other hand, vocabulary predicted MCQ performance only, and WM predicted free recall only.

The moderate correlations found across measures of RC and free recall and also between recall and ORF are in line with those found in other investigations (e.g., Barth, Catts, & Anthony, 2009; Fuchs et al., 1988; Hansen, 1978; Keenan et al., 2008; Riedel, 2007). Although we identified slightly higher correlations between MCQs and ORF than others (Cutting & Scarborough, 2006; Eason et al., 2013; Keenan, et al., 2008), this may be because previous studies examined these relations using a standardized measure whereas we used a researcher-created measure. However, our assessment allowed us to make comparisons across the same passages, which reduced measure-specific variability. Thus, our findings suggest that different approaches to measurement may indeed be capturing different components.

The outcome that literal questions were the easiest was perhaps not unexpected (Eason et al., 2012; Geiger & Millis, 2004). Questions that require critical analysis or inference-making are inherently more difficult given that such questions require students to expand on the text and infer unstated meaning (e.g., Cain & Oakhill, 1999). Our results are informative because these findings demonstrate that this pattern exists even after controlling for student- and passage-level effects and emphasize the importance of including different questions types; measuring RC using only literal MCQs, for instance, could mask student deficits.

Assessment difficulty was also affected by text complexity, although this impact varied across outcomes. Decreasing the cohesiveness consistently

increased the difficulty; however, increasing the syntactic and vocabulary complexity was negatively associated with only ORF and free recall and decreasing decodability negatively impacted ORF only. It may be the case that for RC (MCQs), inference making is substantively impacted by cohesion manipulations such that with less cohesive passages inferences are more difficult to make and thus results in less overall integration of the text into a reader's situation model. Syntax, vocabulary, and decoding manipulations, on the other hand, may have had less of an impact on students' inference making ability and their subsequent performance on MCQs. Another possible explanation for this pattern of findings may be due to the assessment features; MCQs may have provided enough structure and cues that such prompts may have required less active comprehension of the text (e.g., Coleman, Lindstrom, Nelson, Lindstrom, & Gregg, 2010; Keenan & Betjemann, 2006) and therefore relied less on text-based features. In other words, given the format of MCQs, individuals may continue to answer at least some of these items correctly (see Tuinman, 1974). The finding that decoding and language skills (vocabulary) continued to explain additional unique variance in students' RC over and above textual features and other cognitive skills provides further support for the Simple View (Hoover & Gough, 1990), which argues that these components are critical for RC success.

All forms of text manipulations were found to substantially affect students' ORF. Given that ORF is based on students' direct interaction with the text, the observed impact of text-based manipulations on ORF was expected; more complex texts tend to negatively affect an individual's reading rate (e.g., Zvonik & Cummins, 2003). For instance, children tend to exhibit slower reading fluency when reading expository texts than narrative texts (Saenz & Fuchs, 2002), which may be due to the fact that expository texts are more difficult relative to narrative texts with regards to both text structure and vocabulary (see Saenz & Fuchs, 2002). The finding that only word recognition predicted ORF makes sense if one considers that (a) decoding skills are foundational for reading fluency (Pikulski & Chard, 2005) and (b) certain cognitive abilities, such as verbal knowledge, explain differential amounts of unique variance in ORF across differing levels of fluency (Torgesen & Hudson, 2006). Similar to ORF, textual manipulations had a greater impact on students' free recall of the text relative to cognitive skills (aside from WM which was anticipated). This outcome makes sense because more complex text is generally more difficult for individuals to recall relative to simpler text (e.g., Lorch, Lorch, & Inman, 1993). Further, because we ensured that students had little to no background knowledge about the topics presented in the passages, text complexity likely became a more important factor for students' processing and recall of the text (Moravcsik & Kintsch, 1993 cited in Kintsch, 1994). Overall, the observed inconsistency of the text complexity manipulations as predictors of the different reading outcomes provides support for the distinction between text complexity and text difficulty outlined by Mesmer and colleagues (e.g., Cunningham & Mesmer, 2014; Mesmer et al., 2012) and will likely be a fruitful area for future investigation.

Theoretical and practical implications

The findings from the present investigation provide additional empirical support to interactive theories of RC, including the C-I model and the RSF (Kintsch, 1988; Perfetti, 1999), which posit that RC is affected by a myriad of interactions between the reader and the text. For instance, RC performance (as indexed by MCQs) was impacted by the cohesiveness of the text in addition to readers' cohesive inferencing ability, providing evidence for the interaction between textual features and reader characteristics (i.e., text difficulty). Thus, our results highlight the need to take into account the potentially complex nature of textual features and their impact on the RC process in addition to commonly cited predictors of individual differences in RC (e.g., word reading). These findings also provide rationale for the inclusion of aspects of EF within future RC frameworks (e.g., WM).

The results of our investigation have several practical implications as well. First, we provide evidence that text complexity and question type can result in differences in reading performance. This outcome is less than ideal within the scope of identification and remediation of reading difficulties given that identification status may vary as a function of the type of assessment that is being used. Thus, we provide evidence for and argue that a student's reading performance should be assessed using multiple measures (e.g., reading fluency and RC), which is also emphasized by the National Reading Panel (2000) and in progress monitoring (e.g., AIMSweb; Shinn & Shinn, 2002). More importantly, the current findings additionally suggest that scores be interpreted in the context of text complexity, format, and type of reading skills assessed (see Cutting & Scarborough, 2006; Keenan et al., 2008). Second, we have identified several cognitive skills related to reading performance, including vocabulary, word reading, inference making, and WM, that would likely be a positive supplement to intervention programs targeting children with reading difficulties (Hall, 2016; Loosli, Buschkuehl, Perrig, & Jaeggi, 2012; Reed, 2008; Snow, Burns, & Griffin, 1998; Snowling & Hulme, 2011). It is important to note, however, that there remains some debate about the malleability of WM (see Melby-Lervag & Hulme, 2013).

Limitations and future directions

We acknowledge that the present investigation has several limitations. First, because the results regarding reader characteristics were correlational, we are unable to make any causal statements about these findings. Second, these data were not longitudinal. Therefore, it is difficult to say whether the effects seen would be maintained throughout development. Third, because the present sample included native English-speaking adolescents only, we are unable to extrapolate to adults or second-language learners. Also related to the sample, most students were from middle to upper-middle class backgrounds. Fourth, although the selection of expository texts only was intentional, different genres can require different cognitive skills (e.g., Eason et al., 2012). This being said, future investigations should examine the impact of text genre

on the interaction between text- and reader-based characteristics as well as identify whether these interactions vary based on reading ability. Fifth, we examined only the impact of increasing text complexity; thus, we cannot be certain that decreasing text complexity would result in an identical pattern of results. Finally, we acknowledge that the inclusion of IQ within our models may have led to a more conservative estimates than if we did not include IQ; however, the inclusion of this covariate allowed for the identification of the most robust associations among predictors. In addition to the inclusion of multiple measures of EF (e.g., inhibitory control), future studies should utilize factor analytic methods for determining which measures best represent EF. A comprehensive approach to studying RC is necessary to further elucidate the mechanisms behind the elaborate interplay of text complexity, type of RC questions, reader characteristics, and reading outcomes and will lead to a better understanding of why a substantial proportion of children continue to struggle to read.

Appendix A

See Table 6.

Table 6 Passage topics and manipulations

Topics	Manipulations	# of passages per topic
1. Toads	Baseline only	1
2. Sap and syrup	Baseline and vocabulary	2
3. Octopuses	Baseline and vocabulary	2
4. Mustangs (animal)	Baseline and cohesion	2
5. Hot air balloons	Baseline and cohesion	2
6. Moths	Baseline and decoding	2
7. The west branch flood	Baseline and decoding	2
8. Igloos	Baseline and syntax	2
9. Bugs of the Amazon	Baseline and syntax	2
10. Deserts	Baseline only	1

Appendix B: examples of cohesion-, decoding-, syntax-, and vocabulary-manipulated passages

Baseline versus cohesion-manipulated passages

Baseline passage

Humans are born without the well-designed wings that birds have. But humans all over have looked up at the sky and dreamed of flying. The goal of flight was met first by two brothers. They created the world's first hot air balloon. They lived in France 200 years ago. These brothers knew that air gets lighter when it gets hot, and so the men ran a vital test. The brothers put an upside-down paper bag over heat. The hot air was trapped inside the bag, like a balloon. Then they used a small bit of string for tying the bag shut. The bag floated up toward the sky when they let it go, just as they expected it would. So, the brothers then set out to change the world.

The brothers then made a much bigger hot air balloon from silk fabric. At first, the brothers were not sure if the hot air balloon could carry adults safely. As a result, they chose to use a lamb and a duck from a nearby farm. An amazed group met to see the first flight. Even the king of France decided to attend! The hot air balloon soared in the sky for a full 6 min. It landed gently a few miles away. The hot air balloon's passengers were very scared, but they were safe. The brothers then redid the test with humans for passengers.

Since the day that the first humans flew, the dream of flight has taken off. Today, crowds meet from all over the world to fly hot air balloons together. Most of them fly in hot air balloons just for fun. We fly in airplanes to get from one place to another. However, scientists still use balloons to understand the climate. They are also used to collect vital information from inside hurricanes, tornados, and storms.

Cohesion-manipulated passage

We are born without the well-designed wings that birds have. Humans all over have looked up at the sky and dreamed of what it would be like to fly. Two brothers created the very first hot air balloon in France over 200 years ago. They were the first to meet the goal of sending a person up into the air. Knowing that air gets lighter as it heats up, the men ran a vital test. They hung an upside-down paper bag over an open flame. The hot air got trapped inside. Using a small bit of string, it was tied shut. The bag floated up toward the sky when it was let go, just as expected. The brothers set out to change the world.

They made a much bigger hot air balloon from some light fabrics, such as silk. At first, nobody was sure that the new one could carry adults safely. They chose to use a lamb and a duck from a nearby farm instead. An amazed group met to watch the first trip. Even the king of France decided to attend! It soared for a full

6 min before coming down, landing gently a few miles away. The passengers were very scared, but also safe. The innovators redid the same test with humans. Those were the first brave guys to ever lift off of the ground.

That first day of flight brings the dream of flying within reach. Today, crowds gather all over the world to fly hot air balloons together. There are festivals every few months. Most enjoy ballooning as a hobby that is practiced during spare time. In general, airplanes get us from one place to another. With balloons, scientists can better understand severe climates. They are used to collect vital information from inside hurricanes, tornados, and storms.

Baseline versus decoding-manipulated passages

Baseline passage

Most folks do not grasp the difference between moths and butterflies. Both insects have six legs and are the same size. Both come out of cocoons. You may think that butterflies are more appealing, but that is not always correct. In fact, several butterflies are a tawny or pale color. By contrast, the wings of some moths have nice patterns or bright colors. A more dependable method to identify them is to note when they appear. Butterflies are awake during the daytime. Most moths only wake up after sunset. There is also a visible difference in their feelers, which are used for sensing their surroundings. Moths have thick feelers that look like feathers. A butterfly's feelers are skinny with a minor bump at the tip.

Most moths use their noses like straws for drinking nectar from flowers. In fact, one has a huge nose that is three times as long as the rest of its body! Another moth has a nose that is as sharp as a blade. It uses its nose to stab an animal's skin and suck out the blood. It is called a "vampire" moth because it likes to drink blood. Some odd names have been given to others, too. There are "silk," "wax," and "hawk" moths.

Everyone is aware of the fact that flying insects are enticed by lamps. Yet for hundreds of years, nobody knew why. However, scientists now think they can explain this fact. They have seen that the light from the moon and stars enable moths to find their way in the dark. When moths see a light bulb's glow, they get confused because they think the light is up in the sky. The moths you see swirling around lamps at night are lost. By turning off the lights, you will assist them to find their way again.

Decoding-manipulated passage

Most folks do not realize the difference between moths and butterflies. Both insects have six legs and are the same size. Both come out of cocoons. You may think that butterflies are more colorful, but that is not always accurate. In fact, various butterflies are a beige or neutral color. By contrast, the wings of some moths have pretty designs or beautiful colors. A more reliable way to distinguish them is to observe when they appear. Butterflies are awake during the daytime. Most moths only wake up after the sun sets. There is also an obvious difference in their feelers, which are used for sensing

their surroundings. Moths have thick feelers that look like feathers. A butterfly's feelers are wiry with a miniature knob at the tip.

Most moths use their noses like straws for drinking nectar from flowers. In fact, one has a giant nose that is three times as long as the rest of its body! Another moth has a nose that is as sharp as a knife. It uses its nose to pierce an animal's skin and suck out the blood. It is called a "vampire" moth because it likes to drink blood. Some unusual names have been given to other ones, too. There are "emperor," "rough," and "gypsy" moths.

Everyone is familiar with the knowledge that flying insects are beguiled by lamps. Yet for hundreds of years, nobody knew why. However, scientists now know they can explain this curiosity. They have learned that the light from the moon and stars enable moths to find their way in the dark. When moths see a light bulb's glow, they get confused because they believe the light is up in the sky. The moths you see circling around lamps at night are lost. By turning off the lights, you will encourage them to find their way again.

Baseline versus syntax-manipulated passages

Baseline passage

In the winter, the Arctic landscape consists of snow, rocks, and icy waters. Shelter keeps a person from freezing to death in the icy climate. Eskimos have lived in the Arctic for hundreds of years. They survived harsh winters by building homes out of snow. These homes are shaped like domes and are actually very warm. They are called "igloos." The word comes from the Eskimo word for "ice home." An arctic wind can push down the walls or topple the roof of a regular home. But even in the worst storm, the harsh wind sweeps gently over the round igloo.

Eskimos use snow from a single snowfall when building an igloo. Since each block comes from the same snow, they are all just as strong. Eskimos have built homes out of snow the same way for hundreds of years. First, big blocks of snow are placed in a ring. Next, layers are added in a spiral. The top of each block is shaped to lean inward. The final block goes in the middle of the dome. It seals the igloo shut. Finally, snow is stuffed between cracks, and a doorway is dug out. An Eskimo family can connect some of these "ice homes" with tunnels. One of these domes may serve as a family room. Other domes could be used as a kitchen or living room.

If Eskimos get caught in a sudden storm while traveling, they make a small igloo for shelter. After they get inside, they will block the entryway with a final block of snow. Finally, they only have to wait for the storm to stop. Today, very few Arctic families live in homes made of snow all year round. But lots of Eskimos still know how to build one. They practice this very old skill by building igloos when they go ice fishing.

Syntax-manipulated passage

In the winter, snow, rocks, and icy waters make up the Arctic landscape. Without shelter in this icy climate, humans would freeze to death in minutes. For hundreds

of years, the Arctic has been the home of the Eskimos. Through the long, harsh winters, Eskimos stayed warm by building homes out of snow. These remarkably warm homes are shaped like domes. The Eskimo word for "ice home" is actually "igloo". The walls or roof of a regular home would quickly be crushed by strong arctic winds. But even in the worst storm, the harsh wind sweeps gently over the round igloo.

The blocks of snow in an igloo consist of snow from a single snowfall. Since each block comes from the same snow, they are all just as strong. For hundreds of years, Eskimos have built homes the same way. First, big blocks of snow are placed in a ring. Next, layer after layer is added in a spiral. The top of each block is shaped to lean inward. Then, the final block in the middle of the dome seals the igloo shut. At the end of the process, snow is stuffed between cracks and a doorway is dug out. Sometimes, an Eskimo family with more than one dome will connect them with tunnels. One of these domes may serve as a family room. The other domes can be used as kitchens or living rooms.

If Eskimos get caught in a sudden storm while traveling, they make a small igloo for shelter. A final block of snow will block the entryway after everyone is safe inside the igloo. Finally, they only have to wait for the storm to stop. Today, very few Arctic families live in homes made of snow all year round. However, this very old skill is still practiced when Eskimos go ice fishing.

Baseline versus vocabulary-manipulated passages

Baseline passage

Maple trees need a lot of water. They simply cannot live without it. On rainy days, the roots take up as much water as is needed. Then the sap carries the water up to the treetop. The leaves use the water and turn sunlight into food so that the tree can grow. Next, the sap brings extra food from the leaves back down to the roots. The roots store food for another day.

The sap of "sugar maple" trees is distinctive. It has a sweet taste, unlike the sap of other trees. It is the kind of sap that farmers use to cook the maple syrup that we put on pancakes. Cooking syrup is a custom that is 200 years old. First, a farmer drills holes in the trunks of his maple trees. The sap runs right under the bark, so the holes are not too deep. Then the farmer pounds a spout into the hole and hangs a bucket below. On a warm day the sap drips slowly into the bucket. When it is full, the farmer moves the sap into a cooking pot. Finally, the pot is set on a hot stove where it boils until the water is gone. Making just one gallon of maple syrup takes forty gallons of sap.

Sometimes the farmer decides to make some candy by cooking the syrup a little longer on the stove. He waits and waits for the boiling liquid to get very thick. Then the farmer takes it off the stove and puts it in the freezer. The syrup cools until it is finally hard. The end result is a tasty candy. It is a nice treat for the farmer's children. Anybody can make maple candy this way. If you do not have a sugar maple tree, you can get some syrup at a store.

Vocabulary-manipulated passage

Maple trees need a heap of water. They simply cannot subsist without it. On sodden days, the roots sop up as much water as is vital. Then the sap relays the water up to the treetop. The leaves use the water to turn sunlight into food so that the tree can grow. Next the sap transports leftover food from the leaves back down to the roots. The roots amass food for another day.

The sap of "sugar maple" trees is matchless. It boasts a sweet taste, unlike the sap of other trees. It is the kind of sap that farmers employ to cook the maple syrup that we pour on pancakes. Cooking syrup is a custom that is 200 years old. First, a farmer drills holes in the trunks of his maple trees. The sap runs right under the bark, so the holes are not too deep. Then the farmer hammers a spout into the hole and suspends a bucket below. On a balmy day the sap drips sluggishly into the bucket. When it is finally full, the farmer moves the sap into a cooking pot. Finally, the pot is set on a hot stove where it simmers until the water is gone. Cooking just one gallon of maple syrup entails forty gallons of sap.

Sometimes the farmer opts to make some candy by cooking the syrup a tad longer on the stove. He waits and waits for the boiling liquid to get very dense. Then the farmer takes it off the stove and puts it in the freezer. The syrup cools until it is finally stiff. The finishing result is a yummy candy. It is a nice tidbit for the farmer's children. Anybody can make maple candy this way. If you do not have a sugar maple tree, you can get some syrup at a supermarket.

Appendix C: example reading comprehension passage and multiple-choice question types

Toads can be found in almost every part of the world. Freezing temperatures cause unsafe conditions for these cold-blooded animals. In northern climates, toads hibernate all through winter. They protect themselves by making nests underground. Toads sleep in those nests all day. They are mostly awake after dark. That is when the hunt for food takes place. Toads eat mostly bugs and worms. In some places, they have learned to try new foods. Toads capture prey with their sticky tongue. They hunt on land and in water.

Toads have odd eyes that help them to swim and hunt in low light. Their eyeballs have a see-through lid, which enables them to see underwater. Their vision is really good at spotting movement, even in a murky pond or on a cold, moonless night. But a toad's eyesight is not very sharp. An insect can be standing right in front of the toad and as long as that bug stays very still, the toad will not see anything. But as soon as the insect moves, the toad will suddenly react. In an instant, the long sticky tongue will dart out and eat the victim. The hunter's aim is spot-on, so it catches many snails, slugs, and flies.

In many parts of the world, toads are considered to be pests. A long time ago, humans used to bring toads on ships to new lands. In those days, humans used to believe that toads would eat the bugs that like to eat crops. But without natural predators, toads quickly grew in numbers. Now, toads harm local habitats by eating up

all the food. They also harm many animals who try to eat them. The animals do not know that toads have poison in their skin. So now, humans are trying to get rid of toads in gardens and parks.

Factual Humans used to believe that toads:

(a) were a symbol of wealth.
(b) would eat the bugs that like to eat crops.
(c) could cure disease.

Inferential According to the information in the passage, if a fox eats a toad, it would probably:

(a) be very sick.
(b) be unable to digest the toad's eyes.
(c) choke on it.

Process Strategy You will most likely find this passage in:

(a) an email.
(b) a diary.
(c) an encyclopedia.

Critical Analysis In the second paragraph, the author shows that toads:

(a) may not see a moving firefly.
(b) can see a bug swimming in a murky pond.
(c) cannot see through their eyelids.

References

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge: MIT Press.

Amendum, S. J., Conradi, K., & Hiebert, E. (2017). Does text complexity matter in the elementary grades? A research synthesis of text difficulty and elementary students' reading fluency and comprehension. *Educational Psychology Review*. https://doi.org/10.1007/s10648-017-9398-2.

Arrington, C. N., Kulesz, P. A., Francis, D. J., Fletcher, J. M., & Barnes, M. A. (2014). The contribution of attentional control and working memory to reading comprehension and decoding. *Scientific Studies of Reading, 18,* 325–346. https://doi.org/10.1080/10888438.2014.902461.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation, 8,* 47–89. https://doi.org/10.1016/s0079-7421(08)60452-1.

Barnes, M. A., Dennis, M., & Haefele-Kalvaitis, J. (1996). The effects of knowledge availability and knowledge accessibility on coherence and elaborative inferencing in children from fifteen years of age. *Journal of Experimental Child Psychology, 61,* 216–241.

Barth, A. E., Catts, H. W., & Anthony, J. L. (2009). The component skills underlying reading fluency in adolescent readers: A latent variable analysis. *Reading and Writing: An Interdisciplinary Journal, 22,* 567–590. https://doi.org/10.1007/s11145-008-9125-y.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear Mixed-Effects Models using 'Eigen' and S4. R package version 1.1-8.

Benjamin, R. G., & Schwanenflugel, P. J. (2010). Text complexity and oral reading prosody in young readers. *Reading Research Quarterly, 45,* 388–404.

Best, R. M., Floyd, R. G., & McNamara, D. S. (2008). Differential competencies contributing to children's comprehension of narrative and expository texts. *Reading Psychology, 29,* 137–164. https://doi.org/10.1080/02702710801963951.

Bowey, J. A. (1995). Socioeconomic status differences in preschool phonological sensitivity and first-grade reading achievement. *Journal of Educational Psychology, 87,* 476–487. https://doi.org/10.1037/0022-0663.87.3.476.

Buck, J., & Torgesen, J. (2003). *The relationship between performance on a measure of oral reading fluency and performance on the Florida Comprehensive Assessment Test.* Tallahassee, FL: Florida Center for Reading Research.

Cain, K., & Oakhill, J. V. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing: An Interdisciplinary Journal, 11,* 489–503. https://doi.org/10.1023/a:1008084120205.

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology, 96,* 31–42. https://doi.org/10.1037/0022-0663.96.1.31.

Canty, A. & Ripley, B. (2015). boot: Bootstrap functions. R package version 1.3-17.

Carlisle, J. F. (2003). Morphology matters in learning to read: A commentary. *Reading Psychology, 24,* 291–322. https://doi.org/10.1080/02702710390227369.

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (2001). Estimating the risk of future reading difficulties in kindergarten children: A research-based model and its clinical implementation. *Language, Speech, and Hearing Services in Schools, 32,* 38–50. https://doi.org/10.1044/0161-1461(2001/004).

Coleman, C., Lindstrom, J., Nelson, W., Lindstrom, W., & Gregg, K. N. (2010). Passageless comprehension on the Nelson–Denny reading test: Well above chance for university students. *Journal of Learning Disabilities, 43,* 244–249. https://doi.org/10.1177/0022219409345017.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108,* 204–256. https://doi.org/10.1037/0033-295x.108.1.204.

Common Core State Standards Initiative. (2010). *Common Core State Standards for English language arts & literacy in history/social studies, science, and technical subjects.* Washington, DC: CCSSO & National Governors Association.

Cunningham, J. W., & Mesmer, H. A. (2014). Quantitative measurement of text difficulty: What's the use? *The Elementary School Journal, 115*(2), 255–269. https://doi.org/10.1086/678292.

Cutting, L., Saha, N., & Hasselbring, T. (2017). U.S. Patent Application No. 62509856. Washington, DC: U.S. Patent and Trademark Office.

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. https://doi.org/10.1207/s1532799xssr1003_5.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450–466. https://doi.org/10.1016/s0022-5371(80)90312-6.

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis–Kaplan executive function system (D-KEFS).* San Antonio, TX: Pearson.

Denckla, M. B. (1989). Executive function, the overlap zone between attention deficit hyperactivity disorder and learning disabilities. *International Pediatrics, 4,* 155–160.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52,* 219–232. https://doi.org/10.1177/001440298505200303.

Deno, S. L. (1989). Curriculum-based measurement and alternative special education services: A fundamental and direct relationship. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 1–17). New York: Guilford Press.

Deno, S. L. (2003). Developments in curriculum-based measurement. *The Journal of Special Education, 37,* 184–192.

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,* 515–528. https://doi.org/10.1037/a0027182.

Eason, S. H., Sabatini, J., Goldberg, L., Bruce, K., & Cutting, L. E. (2013). Examining the relationship between word reading efficiency and oral reading rate in predicting comprehension among different types of readers. *Scientific Studies of Reading, 17,* 199–223. https://doi.org/10.1080/10888 438.2011.652722.

Edwards-Hewitt, T., & Gray, J. J. (1995). Comparison of measures of socioeconomic status between ethnic groups. *Psychological Reports, 77,* 699–702. https://doi.org/10.2466/pr0.1995.77.2.699.

Elbro, C., & Buch-Iversen, I. (2013). Activation of background knowledge for inference making: Effects on reading comprehension. *Scientific Studies of Reading, 17,* 435–452. https://doi.org/10.1080/10888438.2013.774005.

Engen, L., & Høien, T. (2002). Phonological skills and reading comprehension. *Reading and Writing, 15,* 613–631. https://doi.org/10.1023/a:1020958105218.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of Experimental Psychology: General, 128,* 309–331. https://doi.org/10.1037/0096-3445.128.3.309.

English, L., Barnes, M. A., Fletcher, J. M., Dennis, M., & Raghubar, K. P. (2010). Effects of reading goals on reading comprehension, reading rate, and allocation of working memory in children and adolescents with spina bifida meningomyelocele. *Journal of the International Neuropsychological Society, 16,* 517–525. https://doi.org/10.1017/s1355617710000123.

Fitzgerald, J., Elmore, J., Hiebert, E. H., Koons, H. H., Bowen, K., Sanford-Moore, E. E., et al. (2016). Examining text complexity in the early grades. *Phi Delta Kappan, 97*(8), 60–65. https://doi.org/10.1177/0031721716647023.

Frantz, R. S., Starr, L. E., & Bailey, A. L. (2015). Syntactic complexity as an aspect of text complexity. *Educational Researcher, 44*(7), 387–393. https://doi.org/10.3102/0013189x15603980.

Freebody, P., & Anderson, R. C. (1983). Effects of vocabulary difficulty, text cohesion, and schema availability on reading comprehension. *Reading Research Quarterly, 18,* 277–294. https://doi.org/10.2307/747389.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5,* 239–256. https://doi.org/10.1207/s1532799xssr0503_3.

Fuchs, L. S., Fuchs, D., & Kazdan, S. (1999). Effects of peer-assisted learning strategies on high school students with serious reading problems. *Remedial and Special Education, 20,* 309–318. https://doi.org/10.1177/074193259902000507.

Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education, 9,* 20–28. https://doi.org/10.1177/074193258800900206.

García, J. R., & Cain, K. (2014). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of Educational Research, 84,* 74–111. https://doi.org/10.3102/0034654313499616.

Geiger, J. F., & Millis, K. K. (2004). Assessing the impact of reading goals and text structures on comprehension. *Reading Psychology, 25,* 93–110. https://doi.org/10.1080/02702710490435637.

Goswami, U., Gombert, J. E., & de Barrera, L. F. (1998). Children's orthographic representations and linguistic transparency: Nonsense word reading in English, French, and Spanish. *Applied Psycholinguistics, 19,* 19–52.

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, & Computers, 36,* 193–202. https://doi.org/10.3758/bf03195564.

Hackman, D. A., & Farah, M. J. (2009). Socioeconomic status and the developing brain. *Trends in Cognitive Sciences, 13,* 65–73. https://doi.org/10.1016/j.tics.2008.11.003.

Hall, C. S. (2016). Inference instruction for struggling readers: A synthesis of intervention research. *Educational Psychology Review, 28,* 1–22. https://doi.org/10.1007/s10648-014-9295-x.

Halliday, M. A. K., & Hasan, R. (2014). *Cohesion in English*. New York, NY: Routledge.

Hansen, C. L. (1978). Story retelling used with average and learning disabled readers as a measure of reading comprehension. *Learning Disability Quarterly, 1,* 62–69. https://doi.org/10.2307/1510938.

Hiebert, E. H., & Mesmer, H. A. E. (2013). Upping the ante of text complexity in the Common Core State Standards examining its potential impact on young readers. *Educational Researcher, 42,* 44–51. https://doi.org/10.3102/0013189x12459802.

Hollingshead, A. B. (1975). Four factor index of social status. Unpublished manuscript, Yale University, New Haven, CT.

Homack, S., Lee, D., & Riccio, C. A. (2005). Test review: Delis–Kaplan executive function system. *Journal of Clinical and Experimental Neuropsychology, 22,* 599–609. https://doi.org/10.1080/13803390490918444.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal, 2,* 127–160. https://doi.org/10.1007/bf00401799.

Hosp, M. K., & Fuchs, L. S. (2005). Using CBM as an indicator of decoding, word reading, and comprehension: Do the relations change with grade? *School Psychology Review, 34,* 9–26.

Keenan, J. M., & Betjemann, R. S. (2006). Comprehending the gray oral reading test without reading it: Why comprehension tests should not include passage-independent items. *Scientific Studies of Reading, 10,* 363–380. https://doi.org/10.1207/s1532799xssr1004_2.

Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading, 12,* 281–300. https://doi.org/10.1080/10888430802132279.

Keenan, J. M., & Meenan, C. E. (2014). Test differences in diagnosing reading comprehension deficits. *Journal of Learning Disabilities, 47,* 125–135. https://doi.org/10.1177/0022219412439326.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological Review, 95,* 163–182. https://doi.org/10.1037/0033-295x.95.2.163.

Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist, 49,* 294–303. https://doi.org/10.1037/0003-066x.49.4.294.

Klecker, B. M. (2006). The gender gap in NAEP fourth-, eighth-, and twelfth-grade reading scores across years. *Reading Improvement, 43,* 50–56.

Kulesz, P. A., Francis, D. J., Barnes, M., & Fletcher, J. M. (2016). The influence of properties of the test and their interactions with reader characteristics on reading comprehension: An explanatory item response study. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000126.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6,* 293–323. https://doi.org/10.1016/0010-0285(74)90015-2.

Loosli, S. V., Buschkuehl, M., Perrig, W. J., & Jaeggi, S. M. (2012). Working memory training improves reading processes in typically developing children. *Child Neuropsychology, 18,* 62–78. https://doi.org/10.1080/09297049.2011.575772.

Lorch, R. F., Lorch, E. P., & Inman, W. E. (1993). Effects of signaling topic structure on text recall. *Journal of Educational Psychology, 85,* 281–290. https://doi.org/10.1037/0022-0663.85.2.281.

Lyon, G. R. (1998). Why reading is not a natural process. *Educational Leadership, 55*(6), 14–18.

Mahony, D., Singson, M., & Mann, V. (2000). Reading ability and sensitivity to morphological relations. *Reading and Writing: An Interdisciplinary Journal, 12,* 191–218. https://doi.org/10.1023/a:1008136012492.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14,* 1–43.

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology, 49,* 270–291. https://doi.org/10.1037/a0028228.

Mesmer, H. A., Cunningham, J. W., & Hiebert, E. H. (2012). Toward a theoretical model of text complexity for the early grades: Learning from the past, anticipating the future. *Reading Research Quarterly, 47,* 235–258. https://doi.org/10.1002/rrq.019.

Meyers, J. L., & Beretvas, N. (2006). The impact of inappropriate modeling of cross-classified data structures. *Multivariate Behavioral Research, 41,* 473–497. https://doi.org/10.1207/s15327906mbr4104_3.

Miller, A. C., Davis, N., Gilbert, J. K., Cho, S. J., Toste, J. R., Street, J., et al. (2014). Novel approaches to examine passage, student, and questions effects on reading comprehension. *Learning Disabilities Research and Practice, 29,* 25–35. https://doi.org/10.1111/ldrp.12027.

Miller, A. C., & Keenan, J. M. (2009). How word reading skill impacts text memory: The centrality deficit and how domain knowledge can compensate. *Annals of Dyslexia, 59,* 99–113. https://doi.org/10.1007/s11881-009-0025-x.

Moravcsik, J. E., & Kintsch, W. (1993). Writing quality, reading skills, and domain knowledge as factors in text comprehension. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 47,* 360–374.

Nation, K., & Snowling, M. J. (1998). Semantic processing and the development of word-recognition skills: Evidence from children with reading comprehension difficulties. *Journal of Memory and Language, 39,* 85–101. https://doi.org/10.1006/jmla.1998.2564.

National Center for Education Statistics. (2009). *The nation's report card: Reading 2009 (NCES 2010–458)*. Washington, D.C.: U.S. Department of Education, Institute of Education Sciences.

National Center for Education Statistics. (2013). *National assessment of educational progress (NAEP), 2013 mathematics and reading assessments*. Washington, D.C.: U.S. Department of Education, Institute of Education Sciences.

National Center for Education Statistics. (2015). *NAEP 2015 reading: A report card for the nation and the states*. Washington, D.C.: U.S. Department of Education, Institute of Education Sciences.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). Washington, DC: Government Printing Office.

National Reading Panel (US), National Institute of Child Health, & Human Development (US). (2000). *Report of the national reading panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups*. National Institutes of Health: National Institute of Child Health and Human Development.

O'Connor, R. E., Bell, K. M., Harty, K. R., Larkin, L. K., Sackor, S. M., & Zigmond, N. (2002). Teaching reading to poor readers in the intermediate grades: A comparison of text difficulty. *Journal of Educational Psychology, 94,* 474–485. https://doi.org/10.1037/0022-0663.94.3.474.

Oakhill, J. V., Cain, K., & Bryant, P. E. (2003). The dissociation of word reading and text comprehension: Evidence from component skills. *Language and Cognitive Processes, 18,* 443–468. https://doi.org/10.1080/01690960344000008.

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology, 98,* 554–566. https://doi.org/10.1037/0022-0663.98.3.554.

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

Perfetti, C. A. (1985). *Reading ability*. New York, NY: Oxford University Press.

Perfetti, C. A. (1999). Comprehending written language: A blueprint of the reader. In P. Hagoort & C. M. Brown (Eds.), *Neurocognition of language processing* (pp. 167–208). Oxford: Oxford University Press.

Pikulski, J. J., & Chard, D. J. (2005). Fluency: Bridge between decoding and reading comprehension. *The Reading Teacher, 58,* 510–519. https://doi.org/10.1598/rt.58.6.2.

Priebe, S. J., Keenan, J. M., & Miller, A. C. (2012). How prior knowledge affects word identification and comprehension. *Reading and Writing: An Interdisciplinary Journal, 25,* 131–149. https://doi.org/10.1007/s11145-010-9260-0.

Quinn, J. M., Wagner, R. K., Petscher, Y., & Lopez, D. (2015). Developmental relations between vocabulary knowledge and reading comprehension: A latent change score modeling study. *Child Development, 86,* 159–175. https://doi.org/10.1111/cdev.12292.

Ready, R. E., Chaudhry, M. F., Schatz, K. C., & Strazzullo, S. (2013). "Passageless" administration of the Nelson–Denny reading comprehension test: Associations with IQ and reading skills. *Journal of Learning Disabilities, 46,* 377–384. https://doi.org/10.1177/0022219412468160.

Reed, D. K. (2008). A synthesis of morphology interventions and effects on reading outcomes for students in grades K–12. *Learning Disabilities Research & Practice, 23,* 36–49. https://doi.org/10.1111/j.1540-5826.2007.00261.x.

Reynolds, A. J., Chan, H., & Temple, J. A. (1998). Early childhood intervention and juvenile delinquency: An exploratory analysis of the Chicago child-parent centers. *Evaluation Review, 22,* 341–372. https://doi.org/10.1177/0193841x9802200302.

Reynolds, A. J., Temple, J. A., Robertson, D. L., & Mann, E. A. (2002). Age 21 cost-benefit analysis of the Title I Chicago child-parent centers. *Educational Evaluation and Policy Analysis, 24,* 267–303. https://doi.org/10.3102/01623737024004267.

Riedel, B. W. (2007). The relation between DIBELS, reading comprehension, and vocabulary in urban first-grade students. *Reading Research Quarterly, 42,* 546–567. https://doi.org/10.1598/rrq.42.4.5.

Saenz, L. M., & Fuchs, L. S. (2002). Examining the reading difficulty of secondary students with learning disabilities: Expository versus narrative text. *Remedial and Special Education, 23*(1), 31–41. https://doi.org/10.1177/074193250202300105.

Schroeder, S. (2011). What readers have and do: Effects of students' verbal ability and reading time components on comprehension with and without text availability. *Journal of Educational Psychology, 103,* 877–896. https://doi.org/10.1037/a0023731.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96,* 523–568. https://doi.org/10.1037/0033-295x.96.4.523.

Sesma, H. W., Mahone, E. M., Levine, T., Eason, S. H., & Cutting, L. E. (2009). The contribution of executive skills to reading comprehension. *Child Neuropsychology, 15,* 232–246. https://doi.org/10.1080/09297040802220029.

Shaywitz, S. E., & Shaywitz, B. A. (2005). Dyslexia (specific reading disability). *Biological Psychiatry, 57,* 1301–1309. https://doi.org/10.1016/j.biopsych.2005.01.043.

Shinn, M. R. (Ed.). (1989). *Curriculum-based measurement: Assessing special children.* New York, NY: Guilford.

Shinn, M. M., & Shinn, M. R. (2002). *AIMSweb training workbook: Administration and scoring of reading curriculum-based measurement (R-CBM) for use in general outcome measurement.* Eden Prairie, MN: Edformation.

Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children committee on the prevention of reading difficulties in young children.* Washington, DC: National Research Council.

Snowling, M. J., & Hulme, C. (2011). Evidence-based interventions for reading and language difficulties: Creating a virtuous circle. *British Journal of Educational Psychology, 81,* 1–23. https://doi.org/10.1111/j.2044-8279.2010.02014.x.

Solis, M., Ciullo, S., Vaughn, S., Pyle, N., Hassaram, B., & Leroux, A. (2012). Reading comprehension interventions for middle school students with learning disabilities: A synthesis of 30 years of research. *Journal of Learning Disabilities, 45,* 327–340. https://doi.org/10.1177/0022219411402691.

Tannenbaum, K. R., Torgesen, J. K., & Wagner, R. K. (2006). Relationships between word knowledge and reading comprehension in third-grade children. *Scientific Studies of Reading, 10,* 381–398. https://doi.org/10.1207/s1532799xssr1004_3.

Thurlow, R., & van den Broek, P. (1997). Automaticity and inference generation. *Reading and Writing Quarterly, 13,* 165–184. https://doi.org/10.1080/1057356970130205.

Torgensen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of word reading efficiency (TOWRE).* Austin, TX: Pro-Ed.

Torgesen, J. K., & Hudson, R. (2006). Reading fluency: critical issues for struggling readers. In S. J. Samuels & A. Farstrup (Eds.), *Reading fluency: The forgotten dimension of reading success.* Newark, DE: International Reading Association.

Tuinman, J. J. (1974). Determining the passage dependency of comprehension questions in 5 major tests. *Reading Research Quarterly, 9,* 206–223.

van den Broek, P., Lorch, R. F., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory and Cognition, 29,* 1081–1087. https://doi.org/10.3758/bf03206376.

van den Noortgate, W., de Boeck, P., & Meulders, M. (2003). Cross-classification multilevel logistic models in psychometrics. *Journal of Educational and Behavioral Statistics, 28,* 369–386. https://doi.org/10.3102/10769986028004369.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension.* New York, NY: Academic Press.

Von Stumm, S., & Plomin, R. (2015). Socioeconomic status and the growth of intelligence from infancy through adolescence. *Intelligence, 48,* 30–36. https://doi.org/10.1016/j.intell.2014.10.002.

Waters, G. S., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation for reading comprehension. *The Quarterly Journal of Experimental Psychology, 49,* 51–79.

Wechsler, D. (1999). *Wechsler abbreviated scale of intelligence.* New York, NY: Harcourt Brace & Company.

Wexler, J., Vaughn, S., Edmonds, M., & Reutebuch, C. K. (2008). A synthesis of fluency interventions for secondary struggling readers. *Reading and Writing: An Interdisciplinary Journal, 21,* 317–347. https://doi.org/10.1007/s11145-007-9085-7.

Wiig, E. H., & Secord, W. A. (1992). *Test of word knowledge (TOWK).* San Antonio: Psychological Corporation.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock–Johnson III.* Boston, MA: Houghton Mifflin Harcourt.

Zorzi, M., Houghton, G., & Butterworth, B. (1998). Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1131–1161. https://doi.org/10.1037/0096-1523.24.4.1131.

Zvonik, E. & Cummins, F. (2003). The effect of surrounding phrase lengths on pause duration. In *Eighth European conference on speech communication and technology* (Vol. 82, pp. 176–193). https://doi.org/10.2307/329207