A Tale of Two Tests: The Role of Topic and General Academic Knowledge in Traditional Versus Contemporary Scenario-Based Reading

Zuowei Wang[a,b], Tenaha O'Reilly[a], John Sabatini[b], Kathryn S. McCarthy[c],

Danielle S. McNamara[d]

[a]Educational Testing Service; [b]University of Memphis; [c]Georgia State University;

[d]Arizona State University

Correspondence should be addressed to Zuowei Wang (email: wzw678@gmail.com)

A Tale of Two Tests: The Role of Topic and General Academic Knowledge in Traditional Versus Contemporary Scenario-Based Reading

Abstract

We compared high school students' performance in a traditional comprehension assessment requiring them to identify key information and draw inferences from single texts, and a scenario-based assessment (SBA) requiring them to integrate, evaluate and apply information across multiple sources. Both assessments focused on a non-academic topic. Performance on the two assessments were moderately correlated (r=.57), but the SBA was more difficult (Study 1; n=342). The two assessments similarly depended on basic reading skills but diverged in the relation to academic knowledge and (non-academic) topic knowledge (Study 2; n= 1,107). Academic knowledge was highly predictive of traditional comprehension, but less so for SBA. Topic knowledge was more predictive of SBA than traditional comprehension. Thus, the two assessments tap into similar constructs related to comprehension; however, the level of topic knowledge is more important for performance on scenario-based, multiple-source reading tasks, whereas academic knowledge is more important for traditional reading comprehension tasks.

Keywords: Scenario-based reading; multiple text comprehension; basic reading skills; prior knowledge; assessment

A Tale of Two Tests: The Role of Topic and General Academic Knowledge in Traditional

Versus Contemporary Scenario-Based Reading

1   Introduction

In the past few decades, researchers have reconsidered models of reading

comprehension in light of the affordances and challenges faced by the typical

"modern reader" (Alexander, 2012; Goldman et al., 2016; Magliano, McCrudden,

Rouet, & Sabatini, 2018). Modern readers are faced with unprecedented access to a

vast amount of information afforded by the internet (Alexander, 2012; Goldman et

al., 2016; Leu, Kinzer, Coiro, Castek, & Henry, 2013; Rouet, Durik, & Britt, 2017). More

than ever, readers must find, evaluate, and integrate information across multiple

sources and do so in the service of a reading purpose, task or scenario -- all while

considering the relevance, credibility, and importance of these sources and resolving

discrepancies among them (Britt, Rouet, & Durik, 2017; Magliano et al., 2018). These

reading tasks might be relatively simple and low stakes, such as following a recipe to

prepare a meal; or complex, multistep, and effortful, such as an in-depth internet

search to evaluate the risk of travelling during the Covid-19 pandemic. These

considerations highlight important facets of modern reading; most importantly that

it is inherently context-driven and scenario-based (McCrudden & Schraw, 2007;

Sabatini, O'Reilly, Halderman, & Bruce, 2014).

At present, many reading comprehension assessments have not yet adapted to

these affordances and challenges faced by the modern reader. Traditionally, many

standardized reading comprehension assessments have focused primarily on single-

text comprehension and the topics covered in the passages are expected to be similarly familiar to a wide range of students (Johnston, 1984; Shapiro, 2004). In other words, assessment designers try to minimize the influence of background knowledge on reading comprehension. However, this logic may not reflect how a reader faces real-world comprehension tasks, threatening their validity. As such, they do not reflect contemporary discourse comprehension theories or reading practices of the typical modern reader.

The purpose of this study is to compare two types of reading comprehension assessments: a traditional style test, and a scenario-based reading comprehension test, with the latter design intended to encompass more of the facets of modern reading practices and discourse comprehension theories. We examine differences in performance on these two types of assessments and, critically, explore the roles that different types of knowledge can play in reading comprehension tasks.

1.1 Theoretical Foundations: Relation Between Knowledge and Comprehension

Theories of discourse comprehension assume that readers construct a multi-layered mental model as they read (McNamara & Magliano, 2009). This mental model includes both information explicit in the text, and *inferences* that fill in conceptual gaps encountered in text (Graesser, Singer, & Trabasso, 1994; Kintsch, 1988; McNamara & Kintsch, 1996; McNamara & O'Reilly, 2009). These inferences are generated by connecting information from various parts of the text, as well as through connecting information in the text to information in the reader's background knowledge. Thus, background knowledge can be critical to comprehension in at least

two ways. First, existing knowledge serves as a structure or schema (Anderson, 1978;

Paul & Christopher, 2017) for integrating new information in the formation of a

mental model (Kintsch, 1988, 1998). Second, knowledge provides readers with the

necessary resources (i.e., knowledge-based inferences) to fill in conceptual gaps

encountered in text (McNamara & Kintsch, 1996). Indeed, the relations between

background knowledge and comprehension are well-documented in prior research

(Alexander & Kulikowich, 1991; Alexander, Kulikowich, & Schulze, 1994; Kintsch,

1994; O'Reilly, Wang, & Sabatini, 2019; Shapiro, 2004). Meta-analyses suggest that

background knowledge is one of the strongest predictors of learning (Dochy, Segers,

& Buehl, 1999).

Although it is well known that background knowledge predicts comprehension,

much remains to be investigated regarding how different types of background

knowledge affect the reading comprehension process (Dochy et al., 1999). For

example, general world knowledge is a strong predictor of comprehension for

developing readers (Cain, Oakhill, Barnes, & Bryant, 2001; Talwar, Tighe, &

Greenberg, 2018); whereas domain or topic knowledge tends to be a stronger

predictor of comprehension success for more advanced students, particularly when

faced with informational texts (Alexander et al., 1994; Murphy & Alexander, 2002).

1.2   Two Kinds of Relations Between Knowledge and Comprehension

In this paper, we distinguish two types of knowledge when examining the

relation between knowledge and comprehension: topic knowledge, namely, in-depth

knowledge on the specific topic of the reading task, and general knowledge that

covers a broad range of topics across domains that is not directly related to the reading task. The relation between reading comprehension and topic knowledge is well specified in contemporary reading comprehension models (McNamara & Magliano, 2009). When a student reads a text, relevant knowledge is activated in long-term memory (Cook & O'Brien, 2014; Kendeou & O'Brien, 2014). If a reader has a lot of relevant topic knowledge, more information can be activated and used more efficiently to construct and elaborate upon the mental model being formed as one reads (McNamara, 2009). Therefore, reading comprehension theories suggest that topic knowledge should strongly predict comprehension of content in the topic.

In contrast, the relation between reading comprehension and knowledge that is not directly related to the reading topic (e.g. general academic knowledge) is less straightforward. On one hand, such knowledge may be related to comprehension because it reflects general knowledge acquired across a student's lifetime of learning (i.e., crystalized intelligence; Cattell, 1963), and thus is de facto evidence of good text processing, memory, and perhaps conscientiousness. It may also be that as students become better readers, they simply acquire more knowledge when reading (Cunningham & Stanovich, 1998; Stanovich, 1993), that is, they have more frequent engagement and exposure to text-based knowledge sources. In either case (or in combination), as these students progress through formal schooling, they simultaneously improve their reading comprehension ability and gain ready access to knowledge, especially academic knowledge (Pearson & Hamm, 2005). This analysis would suggest that students' academic knowledge should predict their

comprehension, even when the target readings are on a topic not directly related to a particular academic domain.

Because topic knowledge and academic knowledge are often highly correlated, especially for an academic topic, in this study we focused on the comprehension of contents around a non-academic topic, American football. Knowledge on American football can be gained without requiring one to read (e.g., watching games). One can have high academic knowledge but know little about American football, and vice versa. Indeed, previous work in expertise and domain knowledge demonstrates that students' knowledge of academic and non-academic domains are not always highly related (Griffin, Jee, & Wiley, 2009; Hall & Edmondson, 1992; Schneider, Körkel, & Weinert, 1989; Spilich, Vesonder, Chiesi, & Voss, 1979; Voss, Vesonder, & Spilich, 1980). This allows us to better separate specific topic knowledge (in this study, of American football) and general academic knowledge when examining their roles comprehension. Specifically, we designed a traditional reading comprehension assessment and an SBA, both using texts on the topic of American football. We expect to observe a relatively low correlation between students' knowledge on American football and their academic knowledge given how these types of knowledge can be acquired.

It is not our intent to argue that academic and non-academic knowledge are qualitatively different, but rather that the distinction between academic and non-academic knowledge serves as a means to disentangle reading skill and knowledge. Academic knowledge is more likely to be learned in a formal setting and more likely

to be learned from text. By contrast, non-academic knowledge is often learned in informal contexts. Notions of *funds* of knowledge (González, Moll, & Amanti, 2006; Moll, Amanti, Neff, & González, 1992) suggest that non-academic knowledge and knowledge acquired in informal environments is often robust and can and should be leveraged in the classroom. However, academic knowledge is often the privileged knowledge derived primarily within educational contexts.

1.3   Two Types of Reading Comprehension Assessments

Comprehension can be evaluated in different ways. In a typical standardized reading comprehension assessment (Dunn & Markwardt, 1970; Leslie & Caldwell, 2016; MacGinitie, MacGinitie, Maria, Dreyer, & Hughes, 2000; Wiederholt & Bryant, 1992; Woodcock, McGrew, & Mather, 2001), students are given a multiple-choice test in which they read a passage about a topic and then answer questions about the passage (e.g., locate key details, draw inferences). Once students have completed the questions associated with one passage, they move on to the next passage and associated questions. In such cases, there is no explicitly provided goal for reading the passages. We take the position that all reading is bound by situation and context and is, therefore, always in service of some task (Britt et al., 2017; Rapp & McCrudden, 2018). In these types of test-taking situations, the implied task, and by extension the readers' goal in these assessments is often to race the clock by answering the greatest number of questions correctly before time runs out (Rupp, Ferne, & Choi, 2006).

These traditional comprehension assessments are often designed with the

assumption that students have little direct knowledge of the content; thus, correct answers require reference to the text, rather than drawing on prior knowledge to answer questions independent of textual processing (Shapiro, 2004). Many traditional comprehension tests include series of isolated (e.g., unrelated) texts on idiosyncratic topics in hopes that individual differences in prior knowledge can be mitigated across passages and topics (Johnston, 1984). As each topic only has one passage, and because of assessment developers' efforts to avoid writing items that can be answered solely on the basis of prior knowledge (Shapiro, 2004), the comprehension tasks tend to stay close to the text, requiring readers to find information in the passage or to make inferences across sentences. Higher level tasks that target deeper comprehension and integration of multiple skill operations (e.g. application, analysis, synthesis and evaluation in Bloom, 1956; also see McNamara, Jacovina, & Allen, 2015) are usually not the focus of these comprehension tests.

Of course, the type of reading required in a traditional standardized comprehension test is different from many of our everyday reading experiences. Recognizing this discrepancy, several large-scale assessments have been designed to explore students' comprehension skill in the context of context-driven reading. For example, the PISA 2018 Reading Framework (OECD, 2019) defined reading literacy as "understanding, using, evaluating, reflecting on and engaging with texts in order to achieve one's goals, to develop one's knowledge and potential and to participate in society" (p. 28). Following this definition of reading literacy, the 2018 PISA Reading assessment (OECD, 2019) requires students to read multiple texts not only from

traditional sources, such as literature and textbooks, but also from electronic texts such as emails, blogs and websites. These different sources of texts are connected by what PISA terms as a reading *scenario*, which provides students with "an overarching purpose for reading a collection of thematically related texts in order to complete a higher-level task (e.g. responding to some larger integrative question or writing a recommendation based on a set of texts)" (OECD, 2019, p.41). In a reading scenario, students interact with (virtual) peers and use different text materials to solve a realistic problem, which requires both "basic and higher-level reading and reasoning skills" (OECD, 2019, p.42). These scenario-based assessments (SBAs) parallel the national and international standards (National Assessment Governing Board, 2017; OECD, 2019) for digital literacy that highlight the need for students to think critically, evaluate, and integrate information from across a variety of sources and media. As such, SBAs arguably better resemble reading activities in society than many traditional reading comprehension assessments (OECD, 2019; Rouet, Britt, & Durik, 2017; Sabatini, O'Reilly, Wang, & Dreier, 2018). Thus, while the overarching goal of both traditional tests and SBAs is to "do well on the test", these assessments are designed in ways that afford different reading goals such that the traditional assessments do not demand cross-text integration or deeper application of ideas in the text.

There are several differences between traditional reading comprehension and contemporary scenario-based reading comprehension assessments as exemplified by 2018 PISA Reading (OECD, 2019). First, traditional comprehension assessment does

not explicitly provide students with an overarching goal (other than answering individual questions correctly), whereas in scenario-based comprehension assessment, the reading scenario defines a reading goal that frequently requires complex, multi-step planning and monitoring of sub-goals towards achieving the overarching goal. Second, traditional comprehension assessment often uses single, isolated passages in which no integration across passages is necessary, whereas scenario-based comprehension assessment requires students to process multiple, related texts and integrate information across texts. Third, traditional comprehension assessment focuses on isolated comprehension skills (e.g., identify main idea, make cross-sentence inferences), whereas scenario-based comprehension assessment not only requires students to understand materials piecemeal, but also to apply their understanding in achieving the goal of the reading scenario. Collectively, these differences indicate that SBAs require more complex processing, even when other factors such as text complexity are fixed.

Relevant to this paper, one important difference between traditional comprehension assessment and SBA is the role of prior topic knowledge. While efforts are made to mitigate the role of prior topic knowledge in traditional comprehension assessments (Johnston, 1984; Shapiro, 2004), this constraint becomes somewhat relaxed in SBAs. It is reasonable to believe that topic knowledge can facilitate the integration of information across multiple sources (Kendeou & O'Brien, 2016; McNamara, 2001; Rouet, 2003), which is often required in SBAs. Topic knowledge affords faster activation and processing (Carson, 2007; Ericsson & Kintsch,

11

1995; McNamara & McDaniel, 2004), which can free up cognitive resources (Hahnel et al., 2019) students need to focus on the "higher-level reading and reasoning" tasks posed by SBAs (OECD, 2019, p.42). Thus, we hypothesize the relation to topic knowledge should be stronger for SBAs than traditional comprehension assessments.

The attempts to mitigate topic knowledge in traditional comprehension assessment implies that the assessment is aimed at measuring the ability to process and understand reading materials about novel content. Since understanding novel content is necessary to learning of new content (Kendeou & O'Brien, 2016; Mayer, 2002), traditional comprehension assessment should also predict the ability to learn new content (Pearson & Hamm, 2005). In other words, those who do well on traditional comprehension assessments tend to be efficient academic learners of new content, and consequently, over time they are more likely to acquire more general academic knowledge. Based on this reasoning, we hypothesize that performance on a traditional comprehension test will be strongly correlated with academic knowledge, even when the reading sources and academic knowledge are on different topics and domains.

The strength of the relation between general academic knowledge and SBA performance is less theoretically predictable. If SBA is also highly correlated to academic knowledge as is traditional comprehension assessment, then it would suggest that students who do well academically also tend to be successful in scenario-based, multiple-source reading. If, however, the correlation between SBA and academic knowledge is lower, then it would indicate that a high achieving

student at school may not necessarily be well prepared for the complexities of scenario-based, multiple-source reading. We do not have a specific a-priori hypothesis regarding the direction of this comparison. Nevertheless, given the differences between SBA and traditional comprehension assessment, we hypothesize that the role of academic knowledge might be different in the two assessments.

1.4 The Current Study

The current study examines how background knowledge is related to reading comprehension. Critically, we investigate how both topic-specific knowledge and general academic knowledge (not related to the reading topic) predict comprehension in two types of comprehension assessments.

We address the following research questions:

RQ1. How does students' performance in a traditional single-passage reading comprehension test compare to that in a scenario-based, multiple-text reading comprehension test?

RQ2. Does the role of academic knowledge differ across the two reading comprehension tests?

RQ3. Does the role of topic-specific knowledge differ across the two reading comprehension tests?

2 Study 1

2.1 Method

2.1.1 Participants

Participants were 342 students from three high schools located in three states of the U.S. (California, Idaho and Oklahoma): 178 (52%) were female; 114 (33%) were

from grade 9, 115 (34%) were from grade 10, 69 (20%) were from grade 11 and 44

(13%) were from grade 12. The largest group of students 117 (34%) identified

themselves as White; 107 (31%) mixed race, 53 (16%) Hispanic/Latino, 22 (6%) as

Black, 16 (5%) as American Indian or Alaskan Natives, 12 (4%) as Asian, and the

remaining reporting as other minority groups.

2.1.2   Comprehension Assessments

Participants completed a traditional reading comprehension assessment and an

SBA, both on the topic of American football. The primary concern in the selection of

the targeted non-academic topic was to identify a domain for which variability in

students' knowledge was not due to age, gender, socio-economic status, or region.

Nielsen ratings (Nielsen, 2014) suggested that football captures the most regionally

and gender diverse viewership among major sports in the U.S. Because football

viewership tends to lean slightly male (Salkowitz, 2018), we also examined

performance as a function of gender.

The traditional assessment comprised 10 short passages. The passages ranged

between 242 and 512 words (M=337 words, SD=77 words). The specific topics of

these passages ranged from the history of American football, its difference compared

to rugby, to football rules, the role of the football referee, football equipment such as

helmet and gloves. For each passage, reading comprehension questions were

developed targeting students' understanding of the passage (e.g., identifying key

ideas and details, making inferences).

We created parallel forms (Form A and Form B) of the reading comprehension

test, each including five of the 10 passages. Form A included 41 questions; Form B

included 40 questions. The average Flesch-Kincaid grade level of passages in Form A

was 8.9 (*SD*=.9), and for Form B was 9.4 (*SD*=1.1). Using the TextEvaluator[TM] tool

(Sheehan, Kostin, Napolitano, & Flor, 2014), the average grade level of passages in

Form A was 8.6 (*SD*=.9), and Form B was 8.4 (*SD*=1.5). Consistent with the design of

traditional comprehension tests, items were designed to probe for students'

understanding of a single passage. Items did not require students to integrate

information across passages. The reliability of both forms was .95 (Cronbach's alpha

was used for all reliability measures in this study). The decision to use two parallel

forms instead of merging them into a single form was to reduce students' testing

time so that they could complete the test within a single testing session (i.e., class

period).

The SBA in this study was largely consistent with the PISA 2018 Reading

assessment framework (OECD, 2019). In the SBA, participants were provided with a

realistic purpose – to chat with peers on the topic of American football and decide

whether they would support a proposed rule change. Participants read four

thematically related sources of texts: introduction to football - 832 words, recent rule

changes in professional football – 248 words, a comparison between professional

and college football – 273 words, and understanding football data – 410 words. The

Flesch-Kincaid grade level for the four main SBA passages ranged between 5.5 and

8.3, *M*=6.9, *SD*=1.2. Using the TextEvaluator[TM] tool, the average grade level of the

four passages was *M*=7.5, *SD*=1.3. The four passages provided different perspectives

and differed in format (e.g., formal introductory texts, football play-by-play descriptions). While the passages and tasks were designed to build upon one another, SBA items themselves were independent. Students were required to integrate information across the sources to understand the rules of football and apply the information to interpret football results. For example, one SBA question provides students with a play-by-play description of a football game, and students need to read texts from three sources to predict what happens next. Another question asks students to identify the best player based on performance data. The football SBA test included 24 questions. The reliability for the SBA questions was .87.

Both assessments consisted of multiple-choice questions. Except that one SBA item was a yes/no question, all other items in both assessments had four options. For two SBA items, students were asked to write a short explanation about their selection, but these constructed responses were not scored.

There were several differences between the two assessments: 1) the SBA test had an explicit overarching reading goal (i.e. discussing rule change with peers) whereas the traditional reading comprehension test only required answering typical reading comprehension questions (e.g., identification of key information, basic inferences); 2) the SBA test required the integration of information from multiple texts in order to answer questions, whereas the traditional reading comprehension test used discrete passages; 3) because of the task structure of the SBA test, towards the end of the test students were asked to apply their understanding to a new situation (e.g., interpreting football data to decide who was the best player). The

application questions were not a part of the design or sequence of the traditional

reading comprehension task. In the traditional reading test, there was no task

structure that linked the texts and there were no cross-passage questions.

2.1.3   Procedure

Participants completed the study in two sessions. They completed one form of

the traditional reading comprehension test in one session and the SBA in another.

The order of the assessments was counterbalanced across participants.

2.1.4   Analysis

The main goal of Study 1 was to answer RQ1, that is, examining performance

differences of the same group of students on the two comprehension assessments. A

t-test was conducted to compare the difficulty level of SBA and traditional

comprehension, and the correlation between the two tests was also examined.

Additionally, the effects of assessment type (within-subjects) and gender

(between-subjects) on performance were explored using linear mixed effect models.

As a baseline model, we allowed the intercept of items to randomly vary but fixed

the slope of items. In other words, item difficulty was assumed to vary across items,

but the effects of gender and assessment type were treated as fixed across items. In

the main model, we allowed both the intercept and the slope of items to randomly

vary; namely, the effects of gender and assessment type were assumed to randomly

vary across items. We compared the two models using a likelihood ratio test to

examine whether allowing the effects of gender and assessment type to randomly

vary across items significantly improved model fit. This allowed us to examine

whether the effects, if significant, were stronger in certain items. We used the

statistical software *R* and the *nlme* package for this analysis (Pinheiro, Bates, DebRoy,

& Sarkar, 2017).

2.2   Results

In a preliminary analysis, we examined potential differences across the two

forms of the traditional assessment. Participants who completed Form A, *M*=27.42

(69% correct), *SD*= 10.99 (27%) scored equally well compared to those who

completed Form B, *M*=27.70 (69% correct), *SD*=10.23 (26%). Thus, we collapsed

across these forms in subsequent analyses.

The correlation between students' performance on SBA and that on traditional

comprehension was *r*(340) = .57, *p*<.01. A paired-sample *t*-test showed that SBA

scores were significantly lower than the traditional comprehension test, *t*(341)=8.99,

p<.01, Cohen's *d* =.49, indicating that the SBA was more difficult than the traditional

comprehension assessment.

Linear-mixed effect modelling revealed a significant main effect of assessment

type, with students' accuracy on the SBA about 15% (*B*= -.15) lower than that on the

traditional comprehension assessment, *t*(63)=-3.59, *p*<.01. The effect of gender was

not significant, *B*=.02, *t*(15111)=1.95, *p*=.052. The interaction between gender and

assessment type was significant, *B*=.11, *t*(15111)=6.42, *p*<.01, with male students

less affected by the difficulty of SBA. For random effects, the *SD* of item level

intercept was .12, the *SD* of the slope of gender was .02, the *SD* of the slope of

assessment type was .10, and the *SD* of the slope of the interaction between gender

and assessment type was .06, and the *SD* of residual was .46.

Because in the estimation we allowed these effects to randomly vary across items, we compared this model with the baseline model in which these effects were fixed across items. Results showed that allowing these effects to randomly vary across items did not improve model fit, $\chi^2(9)=12.0$, *p*=.21. Thus, these effects did not significantly differ across items.

Simple effect analysis showed that male students and female students both correctly answered 68% of the traditional comprehension questions, with *SD*'s 26% and 27% respectively, *t*(340)=.08, *p*=.94; however, female students did worse on the SBA (*M*=50%, *SD*=20%) than male students (*M*=64%, *SD*=20%), *t*(340)= -5.43, *p*<.01, Cohen's *d* =.59.

2.3   Discussion

The goal of Study 1 was to address RQ1 – investigating the similarity and differences in students' performance on the traditional and SBA comprehension assessments. Results showed that the SBA was significantly more challenging than the traditional reading comprehension test, and the two comprehension tests were moderately correlated.

What makes the SBA more difficult? Is it because of the *text*, or is it because of the *task*? Due to the inherent differences between SBA and traditional reading comprehension assessment, it is difficult to fix either the text or the task across the two assessments to find a clear answer to this question. For example, passages in an SBA are all connected by the reading scenario, whereas passages in a traditional

reading assessment are isolated and independent, rendering the use of the same set of passages in both assessments difficult. Similarly, while it is common to find a task that requires the student to integrate information from multiple texts in an SBA, such a task is rarely seen in traditional reading comprehension assessments. Nevertheless, we attempt to address this question by looking at text complexity data. As we reported in the Method section, multiple text complexity measures indicated that texts used in the traditional reading comprehension assessment was about one to two grade levels higher than those used in the SBA. Despite the relatively easier texts used in the SBA, the SBA was more difficult than the traditional reading comprehension assessment. Therefore, the higher difficulty level of the SBA is likely more related to the task in this study, rather than the text.

The moderate correlation between the two assessments indicates that the constructs behind the two assessments are related, but not identical. This is further supported by the gender differences within and across the two reading tasks: female and male students performed comparably on the traditional assessment, but female students had lower scores on the SBA than their male counterparts. This finding is counter to most findings in which females tend to outperform their male peers on reading comprehension tests (Stoet & Geary, 2013). One potential explanation for this is that there is a slight gender bias in football viewership (Nielsen, 2014). Thus, this difference in SBA performance may reflect differences in prior topic knowledge. This is one of the issues explored in Study 2.

3 Study 2

In Study 2, we collected measures of both general academic (e.g., science, history) and topic knowledge (football) to explore how each relates to performance on the two types of comprehension tests. Additionally, gender differences in topic knowledge (on American football) are examined in light of the relation between topic knowledge and comprehension. Critically, we also include measures of basic reading skill to more clearly evaluate the unique contributions of knowledge above and beyond lower-level reading skills.

Broadly, we hypothesized that basic reading skill, general academic knowledge, and topic knowledge would all predict reading comprehension performance in both the traditional assessment and SBA. However, we anticipated that the strengths of the relations between these individual differences would vary across the two assessment types.

3.1 Method
3.1.1 Participants

Participants were 1,107 high school students from seven schools in six U.S. states (Alabama, California, Georgia, Idaho, Oklahoma and Rhode Island): 532 (48%) were female; 361 (33%) were from grade 9, 339 (31%) were from grade 10, 195 (18%) were from grade 11 and 212 (19%) were from grade 12. With regards to race, the largest group of students 478 (43%) identified themselves as White, 230 (21%) mixed race, 180 (16%) Hispanic/Latino, 74 (7%) as Black, 30 (3%) as American Indian or Alaskan Natives, 26 (2%) as Asian, and the remaining did not report or reported as other minority groups. We decided not to use multilevel modeling to account for the

multilevel nature of the data (i.e. students nested under schools) because our preliminary analysis showed that school level variance was negligible and not affecting correlation estimates.

3.1.2   Materials

The same traditional reading comprehension assessment and SBA from Study 1 were used in Study 2. In Study 2, participants completed either the SBA or the traditional comprehension assessment.

In order to account for the effects of lower-level, basic reading skills, participants completed adapted versions of three subtests from the RISE reading component tests (Sabatini et al., 2019): Vocabulary, Morphology, and Sentence Processing. All three subtests have good reliability (Cronbach's alphas > .82; Sabatini et al., 2019). In the RISE Vocabulary task, students view a word and select one from three words that share meaning with the target word. In the RISE Morphology task, students fill in a missing word from a sentence by selecting from one of three options. Each option shares the same word root but includes different prefixes or suffixes. The RISE Sentence Processing task also requires students to fill in a missing word by selecting one of three words so that the sentence makes sense. We elected to use shortened versions of the three subtests to make sure students had enough time to complete all assigned tasks in two sessions. As a result, the Vocabulary subtest had 24 items; Morphology had 9 items, and Sentence Processing had 5 items.

Participants completed two background knowledge tests. One evaluated academic knowledge, while the other assessed topic knowledge relevant to the

football (non-academic) texts in the comprehension assessments. Knowledge in two academic domains was evaluated: science and history. Using these two distinct domains allowed us to evaluate the generalizability of our findings. For each domain, 25 multiple-choice items were developed. One example science knowledge question was "Which of the following is the fundamental element found in all living organisms?" One example history question was "Which invention had the greatest impact on the settlement of the Great Plains after the Civil War?" These questions were developed from high school students' released state exams. After the data collection, item properties were examined, including average item difficulty as well as item-total correlation. Three history knowledge questions were excluded because they showed negative item-total score correlation. All science questions showed positive item-total correlation.

Students' knowledge on the topic of American football was evaluated by two types of items: topic vocabulary (30 items) and multiple-choice (40 items). To answer the topic vocabulary items, students needed to indicate if each word they saw in a word list was related to the topic of American football. For example, the word "cornerback" should be selected as related to the topic, whereas the word "yellow card" should be selected as unrelated to the topic. To reduce guessing behavior, students were also given the option to select "I don't know". In this study, correct answers were given 1 point, whereas incorrect answers and "I don't know" were awarded 0 points. The 40 multiple-choice questions were designed to test knowledge of American football. Items included questions about game play (e.g. "How many

yards do you need to get a first down?") as well as questions about National League

Football (e.g., "Which NFL team features a helmet decal only on one side of the

helmet?"). Topic knowledge was scored by totaling correct answers for both the

vocabulary and multiple-choice items. Additionally, students were asked seven

questions to examine their experience and interest with American football, and they

provided their responses on a six-point scale. Some example questions are: "Football

is my favorite sport", "I frequently watch football games on television", and "I enjoy

reading articles about football".

Reliability for all measures calculated from the current sample is provided in

Table 1.

3.1.3   Procedure

This study was completed in two sessions (Figure 1). In the first session,

participants completed the topic knowledge test on American football, a brief survey

about their experiences and interest in American football (seven questions), the

three RISE subtests (Vocabulary, Morphology, and Sentence Processing). In addition,

they were randomly assigned to complete either the science or history academic

knowledge test. In the second session, participants were randomly assigned to

complete either the traditional assessment (randomly assigned to Form A or Form B)

or the SBA.

Due to practical limitations, students only completed one of the two reading

comprehension assessments and only one of the academic knowledge tests. Data

were collected during class time and our collaborating schools preferred two sessions

rather than three. To be considerate of this request, we randomly assigned students

to one general knowledge test and one reading comprehension test. Random

assignments happened within each testing session (classroom), so that students in

the same classroom took either the science or history knowledge test, and either the

SBA or the traditional comprehension assessment.

3.1.4   Analysis

We first replicated some of the Study 1 findings. We compared the difficulty level

of the SBA and the traditional reading comprehension assessment with a *t*-test.

Gender differences on both comprehension measures, topic knowledge, academic

knowledge (science and history), and basic reading skills were examined using

planned *t*-test with Bonferroni correction.

To understand the differences in performance on SBA and traditional

comprehension, we examined the correlation between each comprehension test and

1) basic reading skills; 2) academic knowledge; and 3) topic knowledge. To further

understand the unique contribution of different types of knowledge to

comprehension in SBA and traditional comprehension assessment, we conducted a

set of hierarchical regression analyses.

3.2   Results
3.2.1   Preliminary Analyses

Table 1 shows means, *SD*s and sample sizes on all measures. We first replicated

the analyses from Study 1. These analyses revealed, as in Study 1, that the SBA

(*M*=48% correct, *SD*=23%) was significantly more difficult than the traditional reading

comprehension assessment (*M*=56% correct, *SD*=28%), $t(841)= -4.76$, $p<.01$, Cohen's

d=.33.

Also similar to Study 1, the effect of gender was not significant on the traditional

reading comprehension assessment, *t*(442)= -.32, *p*=.95, but female students did

slightly worse on the SBA, *t*(397)= -2.10, *p*=.036, Cohen's *d*=.21.

As might be expected from the Nielsen (2014) data, female students had less

football knowledge than male students, *t*(1083)= -12.55, *p*<.01, Cohen's *d*=.77. The

effect of gender was not statistically significant on any of the following measures: 1)

history knowledge, *t*(568)= -1.77, *p*=.08; 2); science knowledge, *t*(513)=-.23, *p*=.82; 3)

vocabulary, *t*(1083)=1.67, *p*=.10; or 4) morphology, *t*(1083)=1.80, *p*=.07. While

female students performed slightly better on the sentence processing tasks,

t(1083)=2.17, *p*=.03, Cohen's *d*=.13, this difference does not reach statistical

significance when applying a Bonferroni correction to account for potential Type 1

error due to the multiple *t*-tests calculated.

3.2.2   Correlation Analyses

Table 2 shows the correlation among measures (lower left below diagonal).

Given that students only completed a subset of the tasks (i.e., science or history,

traditional or SBA), Table 2 also includes the sample size of each correlation (upper

right above diagonal). Several patterns emerged. Basic reading skills were similarly

related to traditional reading comprehension performance (*r's* .56 - .57) and SBA

performance (*r's* between .50 - .54). These correlations suggest, perhaps

unsurprisingly, that reading comprehension performance is impacted by basic

reading skills, but that there are additional factors that influence comprehension.

Basic reading skills were also correlated with both types of background knowledge. However, basic reading skills were more strongly correlated with the academic knowledge tests (science and history; r's = 0.43 - 0.57) than with non-academic topic knowledge (r's = 0.26 - 0.29). Topic knowledge was only weakly related to academic knowledge (r= 0.20 for history and r= 0.25 for science).

Critical to our research questions, background knowledge was significantly correlated with reading comprehension, but this varied across background knowledge type and assessment type. Academic knowledge was strongly correlated with traditional reading comprehension test score (r= 0.70), but less so with SBA reading comprehension (r's = 0.53 or 0.57). In contrast, topic knowledge only weakly correlated with traditional reading comprehension (r= 0.25), but was moderately correlated with SBA reading comprehension (r=.50). These findings were in line with our predictions that topic knowledge would be more relevant to scenario-based assessments than for traditional assessments.

Football interest positively predicted students' football knowledge, $r(764)$= 0.44, p<.01, and their performance in both comprehension tests, $r(400)$= 0.14, p<.01 and $r(362)$= 0.26, p<.01, for traditional and SBA respectively. The two correlation coefficients did not significantly differ, z=-1.72, p=.08. Partial correlation analyses showed that once the effect of football knowledge was controlled, football interest no longer predicted students' comprehension performance.

3.2.3   Regression Analyses

To further examine the unique contributions of basic reading skills, academic

knowledge, and football topic knowledge to performance in the two types of reading comprehension tasks, two sets of hierarchical regressions were performed separately for students who completed the traditional reading comprehension task and for those who completed the SBA reading comprehension task. In these regressions, reading comprehension performance was the dependent variable; the three basic reading skills were entered in the first regression model; academic knowledge (science or history), was added in the second model; and football topic knowledge was added in the third regression model. In addition, due to the gender effect wherein female students scored lower than male students in football knowledge, we added gender to the final model. Of particular interest was sr^2, a measure that reflects the unique contribution of each predictor to the dependent variable after all other predictors are controlled.

For students who completed the traditional reading comprehension test, about half completed the science knowledge test (Table 3) and the other half completed the history knowledge test (Table 4). For both groups, basic reading skills explained about 40% of variance in their reading comprehension performance in the first regression model. After adding the academic knowledge predictor into the model, science knowledge explained another 14% of variance, while history knowledge explained another 20% of variance in reading comprehension performance. Finally, in the third regression model, students' football knowledge explained no more than 1% of variance in traditional reading comprehension performance. An examination of Model 3 for both regression analyses showed that the largest contribution to

students' traditional reading comprehension performance was their academic knowledge, with science and history knowledge uniquely contributing to 24% and 19% of variance in reading comprehension, respectively, as reflected by sr^2. Additionally, when gender was added to the multiple regression in Model 4, it failed to explain any variance in comprehension, and it did not alter the effects of other variables (Model 4 not shown in these tables due to space limit).

For students who completed the SBA reading comprehension test, again, about half of them took the science knowledge test (Table 5) and the other half took the history knowledge test (Table 6). For those who completed the science knowledge test, basic reading skills alone explained 30% of the variance in SBA reading comprehension performance. Adding science knowledge to the model explained another 8% of variance in SBA reading comprehension performance. In the third model, after the effects of basic reading skills and science knowledge were controlled, football knowledge explained an additional 10% of variance in SBA reading comprehension performance. For the group who took the history knowledge test, basic reading skills explained 33% of the variance in SBA performance. In Model 2, history knowledge explained another 6% of the variance. In Model 3, topic knowledge explained an additional 11% of variance in SBA reading comprehension. For both groups, Model 3 of the regression analyses showed that the biggest unique contributor to students' SBA reading comprehension performance was football knowledge, uniquely explaining 11% of students' SBA reading comprehension performance. Finally, in Model 4, the effect of gender did not add additional

explanatory power above and beyond differences in reading skill and background

knowledge (Model 4 was not shown in these tables due to space limit).

3.3   Discussion

The goal of Study 2 was to examine the similarities and differences between the

traditional comprehension assessment and SBA, by examining how performance on

each assessment was related to basic reading skills, academic knowledge, and topic

knowledge. Correlation analyses showed that the two comprehension assessments

were similarly related to students' basic reading skills, but they diverged on how each

was related to academic knowledge and topic (non-academic) knowledge. The effect

of academic knowledge was stronger in traditional reading comprehension, whereas

the effect of topic (non-academic) knowledge was stronger in SBA. Traditional

reading comprehension was highly correlated to academic knowledge (r=.70), and in

fact, multiple regression results showed that academic knowledge was the strongest

predictor of traditional reading performance. In contrast, the strongest predictor to

SBA performance was topic knowledge.

Students' interest on the football topic positively predicted their comprehension

performance, and slightly more so for the SBA (r=.26) than the traditional test

(r=.14), although the difference failed to reach statistical significance. The effect of

interest on comprehension performance seems relatively small (r's <.30), and it

disappeared once the effect of topic knowledge was considered. Thus, this suggests

that students' interest in the test topic did not provide unique contributions to

differences between the two comprehension tests. In other words, we do not have

evidence to support the notion that one's comprehension is more affected by their interest, or more generally, motivation to interact with test contents beyond the level of students' topic knowledge.

Results of Study 2 also helped explain the gender differences in SBA observed on Study 1. It appears the gender difference in the SBA was not a consistent effect. Although Study 2 also revealed a gender difference in the SBA, $p=.036$, this effect was not robust against Bonferroni correction for multiple comparisons. Once the effect in football knowledge was controlled in multiple regression, the effect of gender on the SBA disappeared. Thus, the gender difference in SBA was likely to be a result of differences in topic knowledge.

4 General Discussion

The rapidly changing literacy environment of the 21st century poses new demands on people's reading comprehension skills (Alexander, 2012; Britt et al., 2017; Goldman et al., 2016; Leu et al., 2013). As such, evaluations of reading comprehension skill should go beyond reading single texts about esoteric topics toward evaluating authentic, goal-oriented, multiple-document reading tasks. Critically however, these scenario-based assessments may be influenced by reader's existing knowledge differently than in traditional reading tests.

The goal of this study was to examine the role of different types of prior knowledge in two reading comprehension assessments: a traditional single-passage reading comprehension assessment and a scenario-based multiple-text reading comprehension assessment. The traditional assessment was comparable to many

standardized reading comprehension assessments. By contrast, the scenario-based assessment used in this study broadly resembles aspects of the 2018 PISA Reading assessment to embrace key elements of reading in the 21st century literacy environment (OECD, 2019). To help disentangle the effects of prior knowledge with other factors related to reading comprehension performance, we developed the two comprehension assessments about a non-academic topic (American football). Students' topic knowledge, general academic knowledge and basic reading skills were also evaluated. We asked three research questions, and below we address each of these questions based on our results.

4.1 How does students' performance in a traditional single-passage reading comprehension test compare to that in a scenario-based multiple-text reading comprehension test? (RQ1)

Study 1 employed a within-subjects design to compare students' performance on the traditional reading comprehension test and on the SBA. The two reading comprehension tests had moderate correlation (r = .57), consistent with prior findings on the relation between SBA and traditional reading comprehension tests (Sabatini, O'Reilly, Weeks, & Wang, 2020). These results suggest that the two reading comprehension tests measure related but not identical constructs.

The SBA was significantly more difficult than the traditional reading comprehension test used in this study. This means that students who have satisfactory performance on the traditional reading comprehension test may experience more difficulties with the SBA. In other words, success in traditional reading comprehension assessments may not always predict success in multiple-text,

32

scenario-based reading. This supports the notion that traditional comprehension tests may not be an adequate standard for students to meet, if meeting this standard is inadequate for completing more sophisticated, complex multi-source reading tasks.

We can also draw from data in Study 2 to examine if the differences in traditional test and SBA performance are related to basic reading skill. The traditional reading comprehension tests correlated to the basic reading skills (RISE subtests on vocabulary, syntax, and morphology) between r's = .56 - .57. The correlations between basic reading skills and SBA performance were slightly lower (r's = .50 - .54). Comparison of these pairs of correlation coefficients showed that they did not statistically differ. Thus, both tests similarly required students' basic reading skills. Taken together, these findings indicate that there are meaningful differences in test performance across the two types of assessments used in this study, and the differences are likely driven by factors beyond text difficulty and lower-level reading skills.

4.2   Does the role of academic knowledge differ across the two reading comprehension tests? (RQ2)

Reading comprehension is critical for knowledge acquisition, and vice versa (Kendeou & O'Brien, 2016; Pearson & Hamm, 2005). Given that reading comprehension is fundamental across the curriculum, more skilled readers are likely to be more successful in their academic content courses. Thus, we expected that students' reading comprehension performance would be correlated with academic knowledge (general science and history knowledge), even when this knowledge per

se was not directly related to the reading topic.

Our results provide support for this hypothesis. Both reading comprehension tests were focused on American football, a non-academic topic, yet performance on these tests correlated with students' knowledge in science and history. On the traditional comprehension test, academic knowledge was strongly correlated with comprehension test performance (r = .70). Academic knowledge was also moderately to strongly correlated to SBA performance (r's range .53-.57).

The finding that academic knowledge was more correlated with traditional reading comprehension than with SBA performance seems consistent with the different emphases of the two types of comprehension tests. Traditional comprehension tests are designed to focus on the application of specific skills on a text, independent of how much one knows about the topic of that text, and thus it may reflect on how well one would understand new content and acquire new knowledge (Pearson & Hamm, 2005). Consequently, performance on traditional comprehension tests strongly predicts the level of academic knowledge students have acquired. In contrast, SBA focuses on how to apply both existing and new knowledge to solving a problem in a multiple-text reading scenario (Rouet, Britt, et al., 2017). Therefore, it not only requires comprehension and knowledge acquisition, but also requires multiple-text comprehension and knowledge application. The additional processes associated with knowledge application potentially reduced the strength of relation between SBA and students' academic knowledge level. In other words, one might have acquired a good deal of academic knowledge but still not

know how to apply the knowledge within a novel scenario in a topic with which one

is unfamiliar.

4.3 Does the role of topic knowledge differ across the two reading comprehension tests? (RQ3)

To help separate topic knowledge from general academic knowledge, we

selected a non-academic topic and developed the reading assessments. The

correlation between football knowledge and academic knowledge was below .25,

and the correlation between football knowledge to basic reading skills was

below .29. These results are highly consistent with prior research. For example,

McNamara and McDaniel (2004) reported that knowledge on a non-academic topic

(i.e. baseball) correlated to basic reading skills at r=.31, and to general knowledge at

r=.24. These results support the notion that academic knowledge and skills (i.e. basic

reading skills) are separable from nonacademic knowledge.

Based on the differences in how the SBA and traditional comprehension

assessment treated topic knowledge, we predicted that SBA should be more

dependent on topic knowledge. This hypothesis was supported. Football knowledge

was as correlated with SBA performance as was academic knowledge and basic

reading skill (Table 2). That is to say, the knowledge that students had about the topic

at hand was similarly important for comprehension as were the role of foundational

skills in the SBA. In the traditional reading comprehension assessment, by contrast,

football knowledge was a much weaker (r = .25) predictor of comprehension as

compared to basic reading skill and general academic knowledge.

These patterns were further clarified in multiple regression analyses. After

35

controlling for the effect of basic reading skills and academic knowledge, football knowledge explained no more than 1% of variance of traditional comprehension. In contrast, football knowledge still explained around 10% of variance in SBA comprehension after the effects of reading skills and academic knowledge were accounted for.

Several differences between the SBA and the traditional comprehension assessment might have contributed to the differences in the role of topic knowledge. First, the SBA requires the processing of multiple texts that provide different perspectives around a topic. Prior research has demonstrated that topic knowledge helps students integrate information to achieve comprehension and knowledge acquisition (Kendeou & O'Brien, 2016; McNamara, 2001; Rouet, 2003). In the case of the SBA, topic knowledge may facilitate the integration of information across passages to form a coherent multilayered mental model. This suggests that students with higher topic knowledge might have an advantage in integrating information from multiple sources, thus making them better at the SBA. Second, the SBA not only requires students to understand the content, but also to be able to apply their understanding to solving problems and achieve the reading goal of a scenario. Prior research has demonstrated the benefit of topic knowledge in more demanding tasks (Bråten, Gil, & Strømsø, 2011; Carson, 2007; Ericsson & Kintsch, 1995; Gil, Bråten, Vidal-Abarca, & Strømsø, 2010), partly by freeing up the working memory demands from lower level processing (i.e. to gather information from reading the texts) to be used in higher level tasks (i.e. to solve problems and achieve the reading goal).

Alternatively, readers with greater topic knowledge automatically activate more relevant information while reading which leads to more rapid decay of irrelevant information and in turn, a more coherent and stable mental representation of the text content (McNamara & McDaniel, 2004).

4.4   Implications

To summarize the key findings of this study, students' performance in the traditional single-passage reading comprehension test and that in the scenario-based reading comprehension test converge on the dependence on basic reading skills, but they diverge on the relation to different types of knowledge.

Given that traditional comprehension and SBA are similarly dependent on basic reading skills, existing instructional activities that prove successful in training students' basic reading skills will probably remain relevant. On the other hand, the fact that SBA is likely more dependent on topic knowledge than is traditional comprehension reveals new demands for how we might improve reading comprehension in the new literacy environment. It seems necessary to find ways to help students gain the necessary topic knowledge before they can do well in more complex comprehension tasks that require deeper processing of the texts. Indeed, a number of comprehension strategies have been shown effective to help students compensate for knowledge deficits when reading materials that are unfamiliar to them (McNamara, 2017).

Over three decades ago, Johnston (1984) provided compelling results to demonstrate that a comprehension test can be biased if students have different

levels of topic knowledge. He, along with many other researchers, showed that

individuals who know more about the reading topic had an advantage in

comprehension performance. Based on findings like this, researchers and test

designers have tried to reduce the effect of knowledge on reading comprehension,

by using a variety of passages on different topics hoping to eliminate the knowledge

bias by maintaining everyone's topic knowledge comparably on balance. Our design

of the traditional comprehension assessments has used a similar rationale. It appears

that this effort to remove knowledge bias is successful, since we found low

correlation between football knowledge and performance on the traditional

comprehension assessment (*r's* around .25). However, Johnston (1984) further found

that when the effect of prior topic knowledge was mitigated by mixing a variety of

reading topics, a general intelligence bias emerged: those who had higher general

intelligence did better in such a comprehension test. Consistent with Johnston's

(1984) finding, we found a related pattern. While the correlation to topic knowledge

was low, traditional comprehension performance on the topic of American football

was strongly correlated to students' general academic knowledge (*r's*=.70), which

could be considered a facet of crystalized intelligence (Cattell, 1963). Furthermore,

regression analysis showed that general academic knowledge was the strongest

predictor of traditional comprehension, even masking the effect of basic reading

skills and topic knowledge. It seems that when reading comprehension is measured

this way, it highly overlaps with a general knowledge test.

Johnston (1984) has also provided a potential solution for removing this topic

knowledge bias. He proposed that students' topic knowledge should be measured along with their comprehension on the topic, and he demonstrated that after the effect of prior topic knowledge was statistically removed, the residual score of reading comprehension was free from both topic knowledge bias and general intelligence bias. This method can be potentially used in the evaluation of reading comprehension with SBA. It is possible that after removing the effect of prior knowledge in SBA performance, the residual SBA score reflects the reading comprehension ability without influence of topic knowledge nor general knowledge. We leave this question to future research.

4.5   Limitations and Future Directions

In this study we compared students' performance in one SBA comprehension test and one traditional comprehension test. While the development of the SBA is largely consistent with aspects of the PISA 2018 Reading Assessment Framework (OECD, 2019), the generalizability of our findings needs to be examined in future studies with more SBA tests. For researchers who are interested in developing more SBAs, in addition to the PISA 2018 Reading Assessment Framework, we invite them to read publications on the theoretical foundation of SBA (O'Reilly & Sabatini, 2013; Sabatini et al., 2014) and practical considerations for instantiating an SBA (Sabatini et al., 2018; Sabatini et al., 2020).

In order to separate the effects of academic knowledge and topic knowledge, we focused on a non-academic topic – American football. Although we attempted to find a topic that was relatively balanced across demographics, there was still evidence of

gender differences. Female students had less football knowledge than their male counterparts. They also performed less well on the football SBA. There were no gender differences in the traditional comprehension test. These findings contradict most of the existing literature, which almost always reveals that female students outperform males on reading tests (Stoet & Geary, 2013). Thus, in order to disentangle the effects of gender and knowledge (O'Reilly & McNamara, 2007), future research should replicate this study with other topics.

Additionally, the between-subjects design of Study 2 prevented us from running advanced models such as structural equation modeling, which could help reveal more nuanced relations among basic reading skills, academic knowledge, topic knowledge, traditional comprehension and SBA comprehension. For example, using a complete within-subjects design, one would be able to examine the relation between SBA and traditional comprehension assessment after removing the effect of topic and general academic knowledge, which would provide evidence to evaluate the residual score solution proposed by Johnston (1984).

Finally, the passages that we used in the two reading comprehension assessments were not identical. Part of this is because of the inherent nature of the two assessments: in the SBA passages are connected by a reading scenario, whereas in the traditional assessment passages are isolated. Future experimental studies that systematically vary the text content of SBA and traditional assessments will further clarify the source of differences between the two types of comprehension assessments. For example, we found that the SBA was more difficult than the

traditional comprehension test used in this study. While this finding is consistent with the notion that SBAs require deeper comprehension (O'Reilly, Sabatini, & Wang, 2018), one can certainly revise our traditional comprehension test and make it more difficult than the SBA. Future research should examine the sources of difficulty of SBAs.

5   Conclusion

The rapidly changing literacy environment in which large amounts of information becomes easily accessible in many different forms poses new demands on people's reading comprehension ability. Today, people read in a variety of scenarios to achieve certain goals, and they process information from multiple sources to acquire knowledge and then apply it to new situations. Our comparison of students' performance on a traditional reading comprehension test with that on a modern scenario-based multiple-text reading comprehension test illustrates the varying demands underlying reading comprehension in terms of skills and knowledge. While the role of basic reading skills remains largely the same in the two types of comprehension tasks, and those who know more in general often achieve better comprehension (especially for traditional reading), knowledge on the specific reading topic may become more important in scenario style reading environments. Knowing the level students' prior topic knowledge will help understand their comprehension performance in modern scenario-based multiple-text reading.

References

Alexander, P. A. (2012). Reading into the future: Competence for the 21st century. *Educational Psychologist, 47*(4), 259-280. http://doi.org/10.1080/00461520.2012.722511

Alexander, P. A., & Kulikowich, J. M. (1991). Domain knowledge and analogic reasoning ability as predictors of expository text comprehension. *Journal of Reading Behavior, 23*(2), 165-190. https://doi.org/10.1080%2F10862969109547735

Alexander, P. A., Kulikowich, J. M., & Schulze, S. K. (1994). How subject-matter knowledge affects recall and interest. *American Educational Research Journal, 31*(2), 313-337. https://doi.org/10.3102/00028312031002313

Anderson, R. C. (1978). Schema-directed processes in language comprehension. In A. Lesgold, J. Pelligreno, F. Fokkema, & R. Glaser (Eds.), *Cognitive Psychology and Instruction* (pp. 67-82). New York, NY: Plenum.

Bloom, B. S. (Ed.) (1956). *Taxonomy of Educational Objectives. Handbook 1: Cognitive Domain.* New York, NY: David McKay Company, Inc.

Bråten, I., Gil, L., & Strømsø, H. I. (2011). The role of different task instructions and reader characteristics when learning from multiple expository texts. In M. Mccrudden, J. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 95-122). Greenwich, CT: Information Age Publishing.

Britt, M. A., Rouet, J.-F., & Durik, A. M. (2017). *Literacy beyond text comprehension: A theory of purposeful reading*. New York, NY: Taylor & Francis.

Cain, K., Oakhill, J. V., Barnes, M. A., & Bryant, P. E. (2001). Comprehension skill, inference-making ability, and their relation to knowledge. *Memory & cognition, 29*(6), 850-859. https://doi.org/10.3758/BF03196414

Carson, J. (2007). A problem with problem solving: Teaching thinking without teaching knowledge. *The Mathematics Educator, 17*(2).

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*(1), 1-22. https://psycnet.apa.org/doi/10.1037/h0046743

Cook, A. E., & O'Brien, E. J. (2014). Knowledge activation, integration, and validation during narrative text comprehension. *Discourse Processes, 51*(1-2), 26-49. https://doi.org/10.1080/0163853X.2013.855107

Cunningham, A. E., & Stanovich, K. E. (1998). What reading does for the mind. *American Educator, 22*, 8-17.

Dochy, F., Segers, M., & Buehl, M. M. (1999). The relation between assessment practices and outcomes of studies: The case of research on prior knowledge. *Review of educational research, 69*(2), 145-186. https://doi.org/10.3102%2F00346543069002145

Dunn, L. M., & Markwardt, F. C. (1970). *Peabody individual achievement test*. Circle Pines, MN: American Guidance Service.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*(2), 211-245. https://doi.org/10.1037/0033-295X.102.2.211

Gil, L., Bråten, I., Vidal-Abarca, E., & Strømsø, H. I. (2010). Summary versus argument

tasks when working with multiple documents: Which is better for whom?

Contemporary Educational Psychology, 35(3), 157-173.

https://doi.org/10.1016/j.cedpsych.2009.11.002

Goldman, S. R., Britt, M. A., Brown, W., Cribb, G., George, M., Greenleaf, C., . . .

Project READI. (2016). Disciplinary literacies and learning to read for

understanding: A conceptual framework for disciplinary literacy. *Educational

Psychologist, 51*(2), 219-246.

https://doi.org/10.1080/00461520.2016.1168741

González, N., Moll, L. C., & Amanti, C. (2006). *Funds of knowledge: Theorizing

practices in households, communities, and classrooms*. Mahwah, NJ:

Lawrence Erlbaum Associates, Inc.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during

narrative text comprehension. *Psychological Review, 101*(3), 371.

https://doi.org/10.1037/0033-295X.101.3.371

Griffin, T. D., Jee, B. D., & Wiley, J. (2009). The effects of domain knowledge on

metacomprehension accuracy. *Memory & Cognition, 37*(7), 1001-1013.

https://doi.org/10.3758/MC.37.7.1001

Hahnel, C., Schoor, C., Kroehne, U., Goldhammer, F., Mahlow, N., & Artelt, C. (2019).

The role of cognitive load in university students' comprehension of multiple

documents. *Zeitschrift für Pädagogische Psychologie, 33*, 105-118.

http://doi.org/10.1024/1010-0652/a000238

Hall, V. C., & Edmondson, B. (1992). Relative importance of aptitude and prior

domain knowledge on immediate and delayed posttests. *Journal of*

Educational Psychology, 84(2), 219-223. https://doi.org/10.1037/0022-
0663.84.2.219

Johnston, P. (1984). Prior knowledge and reading comprehension test bias. *Reading*

Research Quarterly, 19(2), 219-239. https://doi.org/10.2307/747364

Kendeou, P., & O'Brien, E. J. (2014). The knowledge revision components (KReC)

framework: Processes and mechanisms. In D. N. Rapp & J. L. G. Braasch (Eds.),

Processing inaccurate information: Theoretical and applied perspectives from

cognitive science and the educational sciences (pp. 353-377). Cambridge, MA:

MIT Press.

Kendeou, P., & O'Brien, E. J. (2016). Prior knowledge: acquisition and revision. In P.

Afflerbach (Ed.), *Handbook of individual differences in reading: Reader, text,*

and context (pp. 151-163). New York, NY: Routledge.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A

construction-integration model. *Psychological Review, 95*(2), 163.

https://doi.org/10.1037/0033-295X.95.2.163

Kintsch, W. (1994). Text comprehension, memory, and learning. *American*

Psychologist, 49(4), 294-303.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: NY:

Cambridge University Press.

Leslie, L., & Caldwell, J. S. (2016). *Qualitative reading inventory*. New York: NY:

Pearson.

Leu, D. J., Kinzer, C., Coiro, J., Castek, J., & Henry, L. (2013). New literacies and the new literacies of online reading comprehension: A dual level theory. In N. Unrau & D. Alvermann (Eds.), *Theoretical models and process of reading (6th ed.)* (pp. 1150-1181). Newark, DE: International Reading Association.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., Dreyer, L. G., & Hughes, K. E. (2000). *Gates-MacGinitie reading tests- 4th Ed. (GMRT-4)* Rolling Meadows, IL: Riverside Publishing.

Magliano, J. P., McCrudden, M. T., Rouet, J.-F., & Sabatini, J. (2018). The modern reader: Should changes to how we read affect research and theory? In M. F. Schober, D. N. Rapp, & M. A. Britt (Eds.), *The Routledge handbook of discourse processes* (pp. 343-361). New York, NY: Routledge.

Mayer, R. E. (2002). Rote versus meaningful learning. *Theory Into Practice, 41*(4), 226-232. https://doi.org/10.1207/s15430421tip4104_4

McCrudden, M. T., & Schraw, G. (2007). Relevance and goal-focusing in text processing. *Educational psychology review, 19*(2), 113-139. https://doi.org/10.1007/s10648-006-9010-7

McNamara, D. S. (2001). Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 55*(1), 51-62. https://doi.org/10.1037/h0087352

McNamara, D. S. (2017). Self-explanation and reading strategy training (SERT) improves low-knowledge students' science course performance. *Discourse*

Processes, 54(7), 479-492. https://doi.org/10.1080/0163853X.2015.1101328

McNamara, D. S., Jacovina, M. E., & Allen, L. K. (2015). Higher order thinking in

comprehension. In P. Afflerbach (Ed.), *Handbook of individual differences in*

reading: Text and context (pp. 164-176). New York, NY: Rouledge.

McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior

knowledge and text coherence. *Discourse processes, 22*(3), 247-288.

https://doi.org/10.1080/01638539609544975

McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of

comprehension. In B. H. Ross (Ed.), *Psychology of learning and motivation*

(Vol. 51, pp. 297-384). Burlington, MA: Academic Press.

McNamara, D. S., & McDaniel, M. A. (2004). Suppressing irrelevant information:

Knowledge activation or inhibition? *Journal of Experimental Psychology:*

Learning, Memory, and Cognition, 30(2), 465-482.

https://doi.org/10.1037/0278-7393.30.2.465

McNamara, D. S., & O'Reilly, T. (2009). Theories of comprehension skill: Knowledge

and strategies versus capacity and suppression. In A. M. Columbus (Ed.),

Progress in experimental psychology research (pp. 1-24). Hauppauge, NY:

Nova Science Publishers, Inc.

Moll, L. C., Amanti, C., Neff, D., & González, N. (1992). Funds of knowledge for

teaching: Using a qualitative approach to connect homes and classrooms.

Theory Into Practice, 31(2), 132-141.

https://doi.org/10.1080/00405849209543534

Murphy, P. K., & Alexander, P. A. (2002). What counts? The predictive powers of

  subject-matter knowledge, strategic processing, and interest in domain-

  specific performance. *The Journal of Experimental Education, 70*(3), 197-214.

  https://doi.org/10.1080/00220970209599506

National Assessment Governing Board. (2017). *Reading Framework for the 2017*

  National Assessment of Educational Progress. Washington, DC: U.S.

  Government Printing Office Retrieved from

  https://www.nagb.gov/content/nagb/assets/documents/publications/frame

  works/reading/2017-reading-framework.pdf.

Nielsen. (2014). 2013 Year in sports media report.   Retrieved from

  https://www.nielsen.com/wp-content/uploads/sites/3/2019/04/year-in-

  sports-media-report-2013.pdf

O'Reilly, T., & Sabatini, J. (2013). *Reading for understanding: How performance*

  moderators and scenarios impact assessment design (RR-13-31). Princeton,

  NJ: Educational Testing Service.

O'Reilly, T., & McNamara, D. S. (2007). The impact of science knowledge, reading skill,

  and reading strategy knowledge on more traditional "high-stakes" measures

  of high school students' science achievement. *American Educational Research*

  Journal, 44(1), 161-196. https://doi.org/10.3102/0002831206298171

O'Reilly, T., Sabatini, J., & Wang, Z. (2018). Using Scenario-Based Assessments to

  Measure Deep Learning. In K. Millis, D. Long, J. Magliano, & K. Weimer (Eds.),

  Deep learning: Multi-disciplinary approaches (pp. 197-208). New York, NY:

Routledge.

O'Reilly, T., Wang, Z., & Sabatini, J. (2019). How Much Knowledge Is Too Little? When

a Lack of Knowledge Becomes a Barrier to Comprehension. *Psychological*

Science, 30(9), 1344-1351. https://doi.org/10.1177/0956797619862276

OECD. (2019). *PISA 2018 Assessment and Analytical Framework*. Paris, France: OECD.

Paul, X. J., & Christopher, G. (2017). Teaching Reading Skill Using Heuristic Technique

Through Schema Theory. *IUP Journal of English Studies, 12*(1), 11-15.

Pearson, P. D., & Hamm, D. N. (2005). The assessment of reading comprehension: A

review of practices—past, present, and future. In S. G. Paris & S. A. Stahl

(Eds.), *Children's reading comprehension and assessment* (pp. 13-69).

Mahwah, NJ: Lawrence Erlbam Associates.

Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D. (2017). nlme: linear and nonlinear

mixed effects models.   Retrieved from http://CRAN.R-

project.org/package=nlme

Rapp, D. N., & McCrudden, M. T. (2018). Relevance before, during, and after

discourse experiences. In J. Strassheim & H. Nasu (Eds.), *Relevance and*

irrelevance: Theories, factors and challenges (pp. 141-160). Berlin, Germany:

Walter de Gruyter GmbH & Co. KG.

Rouet, J.-F. (2003). What was I looking for? The influence of task specificity and prior

knowledge on students' search strategies in hypertext. *Interacting with*

computers, 15(3), 409-428.

Rouet, J.-F., Britt, M. A., & Durik, A. M. (2017). RESOLV: Readers' representation of

reading contexts and tasks. *Educational Psychologist, 52*(3), 200-215.

https://doi.org/10.1080/00461520.2017.1329015

Rouet, J.-F., Durik, A. M., & Britt, M. A. (2017). *Literacy beyond text comprehension: A theory of purposeful reading*. New York, NY: Routledge.

Rupp, A. A., Ferne, T., & Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: A cognitive processing perspective. *Language testing, 23*(4), 441-474.

https://doi.org/10.1191/0265532206lt337oa

Sabatini, J., O'Reilly, T., Halderman, L. K., & Bruce, K. (2014). Integrating scenario-based and component reading skill measures to understand the reading behavior of struggling readers. *Learning Disabilities Research & Practice, 29*(1), 36-43. https://doi.org/10.1111/ldrp.12028

Sabatini, J., O'Reilly, T., Wang, Z., & Dreier, K. (2018). Scenario-based assessment of multiple source use *Handbook of Multiple Source Use* (pp. 447-465). New York, NY: Routledge.

Sabatini, J., O'Reilly, T., Weeks, J., & Wang, Z. (2020). Engineering a twenty-first century reading comprehension assessment system utilizing scenario-based assessment techniques. *International Journal of Testing, 20*, 1-23.

https://doi.org/10.1080/15305058.2018.1551224

Sabatini, J., Weeks, J., O'Reilly, T., Bruce, K., Steinberg, J., & Chao, S.-F. (2019). *SARA Reading Components Tests, RISE forms: Technical adequacy and test design, 3rd edition (Research ReportNo. RR-19-36)*. Princeton, NJ: Educational Testing

Service.

Salkowitz, R. (2018). Data Shows Women Paid More Attention To Super Bowl LII Than

Men.   Retrieved from

https://www.forbes.com/sites/robsalkowitz/2018/02/05/data-shows-women-

paid-more-attention-to-super-bowl-lii-than-men/#34b5dc445c4a

Schneider, W., Körkel, J., & Weinert, F. E. (1989). Domain-specific knowledge and

memory performance: A comparison of high-and low-aptitude children.

Journal of Educational Psychology, 81(3), 306-312.

https://doi.org/10.1037/0022-0663.81.3.306

Shapiro, A. M. (2004). How including prior knowledge as a subject variable may

change outcomes of learning research. *American Educational Research*

Journal, 41(1), 159-189. https://doi.org/10.3102/00028312041001159

Sheehan, K. M., Kostin, I., Napolitano, D., & Flor, M. (2014). The TextEvaluator tool:

Helping teachers and test developers select texts for use in instruction and

assessment. *The Elementary School Journal, 115*(2), 184-209.

https://doi.org/10.1086/678294

Spilich, G. J., Vesonder, G. T., Chiesi, H. L., & Voss, J. F. (1979). Text processing of

domain-related information for individuals with high and low domain

knowledge. *Journal of verbal learning and verbal behavior, 18*(3), 275-290.

https://doi.org/10.1016/S0022-5371(79)90155-5

Stanovich, K. E. (1993). Does reading make you smarter''literacy and the

development. In H. W. Reese (Ed.), *Advances in child development and*

behavior (Vol. 24, pp. 133-180). San Diego, CA: Academic Press.

Stoet, G., & Geary, D. C. (2013). Sex differences in mathematics and reading achievement are inversely related: Within-and across-nation assessment of 10 years of PISA data. *PloS One, 8*(3), e57988.

https://doi.org/10.1371/journal.pone.0057988

Talwar, A., Tighe, E., & Greenberg, D. (2018). Augmenting the Simple View of Reading for struggling adult readers: A unique role for background knowledge. *Scientific Studies of Reading, 22*(5), 351-366.

https://doi.org/10.1080/10888438.2018.1450410

Voss, J. F., Vesonder, G. T., & Spilich, G. J. (1980). Text generation and recall by high-knowledge and low-knowledge individuals. *Journal of Verbal Learning and Verbal Behavior, 19*(6), 651-667. https://doi.org/10.1016/S0022-5371(80)90343-6

Wiederholt, J. L., & Bryant, B. R. (1992). *Gray oral reading tests: GORT-3*. Austin, TX: PRO-ED.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Itasca, IL: Riverside Publishing.
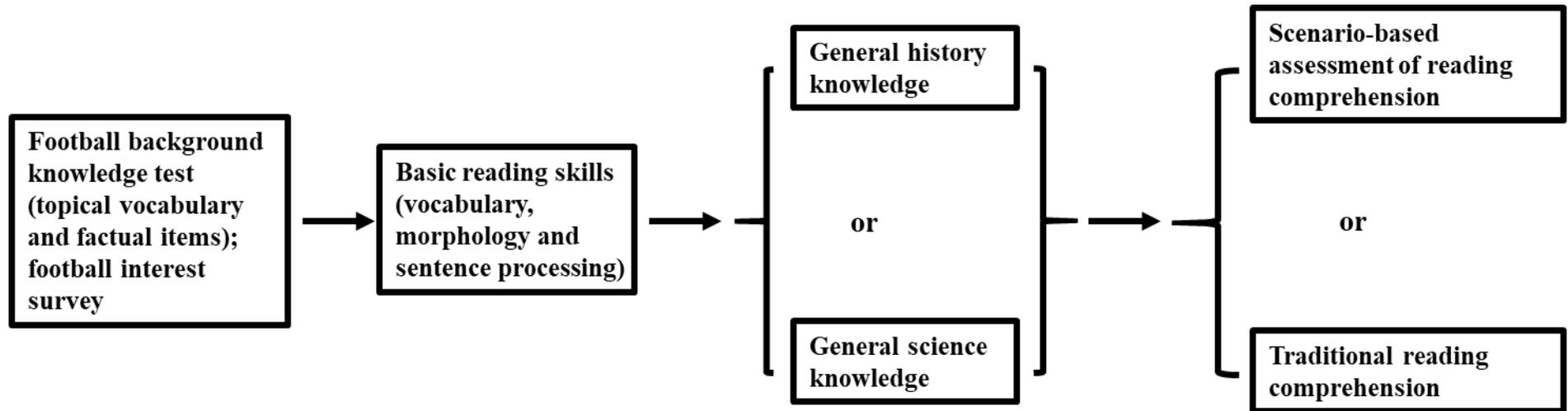
Figure 1. Tests taken by participants of Study 2.

Table 1.
Performance on all measures of Study 2.

	M	*SD*	Range	*N*	# Items	Reliability
Vocabulary	18.2	5.2	[3,24]	1085	24	.88
Morphology	7.2	2.4	[0,9]	1085	9	.84
Sentence Processing	3.7	1.6	[0,5]	1085	5	.79
Football knowledge	38.5	14.0	[6,70]	1085	70	.94
Football interest	20.8	10.4	[1,42]	766	7	.93
History Knowledge	8.2	3.9	[1,21]	570	22	.72
Science Knowledge	10.6	5.0	[0,25]	515	25	.81
RC SBA	11.5	5.6	[1,24]	399	24	.86
RC Traditional, Form A	22.1	11.1	[6,41]	221	41	.94
RC Traditional, Form B	23.4	11.3	[7,40]	223	40	.95
RC Traditional, merged	22.8	11.2	[4,41]	444	81	NA

Note. RC=reading comprehension. SBA=scenario-based assessment.

Table 2.
Correlation among measures (lower left) and corresponding sample sizes (diagonal and upper right).

	Traditional RC	SBA	Football knowledge	Football interest	History knowledge	Science knowledge	VOC	MOR	SP
Traditional RC	786	342	432	402	214	218	432	432	432
SBA	.57	741	389	364	204	185	389	389	389
Football knowledge	.25	.50	1085	766	570	515	1085	1085	1085
Football interest	.14	.26	.44	766	394	372	766	766	766
History knowledge	.70	.53	.20	.07 (N.S.)	570	0	570	570	570
Science knowledge	.70	.57	.25	.02 (N.S.)	--	515	515	515	515
Vocabulary (VOC)	.56	.54	.29	<.01(N.S.)	.50	.57	1085	1085	1085
Morphology (MOR)	.57	.50	.26	.02 (N.S.)	.43	.52	.74	1085	1085
Sentence Processing (SP)	.57	.53	.26	<.01 (N.S.)	.46	.57	.69	.77	1085

Note: Except for those marked as N.S., all $p's<.01$; RC=reading comprehension; SBA=scenario-based assessment.

Table 3.
Multiple regression predicting traditional reading comprehension performance using reading skills, science knowledge and football knowledge

Variable	B (SE)	β	t	sr^2	Adjusted R^2	ΔR^2
Model 1					.40	.40**
Vocabulary	.70 (.18)	.32	3.79**	.06		
Morphology	.98 (.43)	.20	2.26*	.02		
Sentence Processing	1.4 (.54)	.19	2.50*	.03		
Model 2					.54	.14**
Vocabulary	.20 (.17)	.09	1.18	.01		
Morphology	.96 (.37)	.19	2.56*	.03		
Sentence Processing	.39 (.48)	.06	.80	<.01		
Science Knowledge	1.10 (.13)	.51	8.42**	.25		
Model 3					.55	.01*
Vocabulary	.17 (.17)	.08	1.00	<.01		
Morphology	.96 (.37)	.19	2.57*	.03		
Sentence Processing	.35 (.48)	.05	.73	<.01		
Science Knowledge	1.09 (.13)	.50	8.27**	.24		
Football Knowledge	.09 (.04)	.10	2.08*	.02		

*p<.05; **p<.01.

Table 4.

Multiple regression predicting traditional reading comprehension performance with reading skills, history knowledge and football knowledge

Variable	B (SE)	β	t	sr^2	Adjusted R^2	ΔR^2
Model 1					.39	.39**
Vocabulary	.42 (.17)	.20	2.51*	.02		
Morphology	.68 (.45)	.15	1.52	.01		
Sentence Processing	2.58 (.65)	.35	3.99**	.05		
Model 2					.59	.20**
Vocabulary	.09 (.14)	.04	.61	<.01		
Morphology	.47 (.37)	.10	1.27	<.01		
Sentence Processing	1.97 (.53)	.27	3.69**	.03		
History Knowledge	1.50 (.15)	.52	1.15**	.20		
Model 3					.59	<.01
Vocabulary	.07 (.14)	.03	.50	<.01		
Morphology	.47 (.37)	.10	1.27	<.01		
Sentence Processing	1.97 (.53)	.27	3.68**	.03		
History Knowledge	1.49 (.15)	.51	9.99**	.19		
Football Knowledge	.03 (.04)	.04	.88	<.01		

*p<.05; **p<.01; reading comprehension=reading comprehension.

Table 5.

Multiple regression predicting SBA reading comprehension performance using basic reading skills, science knowledge, and football knowledge

Variable	B (SE)	β	t	sr^2	Adjusted R^2	ΔR^2
Model 1					.30	.30**
Vocabulary	.32 (.11)	.30	2.98**	.03		
Morphology	.16 (.26)	.08	.63	<.01		
Sentence Processing	.75 (.37)	.22	2.02*	.02		
Model 2					.38	.08**
Vocabulary	.21 (.10)	.20	1.98*	.01		
Morphology	.14 (.25)	.06	.57	<.01		
Sentence Processing	.28 (.36)	.08	.77	<.01		
Science Knowledge	.40 (.08)	.37	4.89**	.08		
Model 3					.48	.10**
Vocabulary	.15 (.10)	.14	1.56	.01		
Morphology	-.06 (.23)	-.03	-.28	<.01		
Sentence Processing	.35 (.33)	.10	1.07	<.01		
Science Knowledge	.38 (.07)	.36	5.16**	.08		
Football Knowledge	.14 (.02)	.35	6.12**	.11		

*p<.05; **p<.01; SBA=scenario-based assessment.

Table 6. Multiple regression predicting SBA reading comprehension performance using basic reading skills, history knowledge, and football knowledge

Variable	B (SE)	β	t	sr^2	Adjusted R^2	ΔR^2
Model 1					.33	.33**
Vocabulary	.30 (.10)	.28	3.00**	.03		
Morphology	.15 (.22)	.07	.66	<.01		
Sentence Processing	.98(.34)	.28	2.84**	.03		
Model 2					.39	.06**
Vocabulary	.16 (.10)	.15	1.54	.01		
Morphology	.20 (.22)	.09	.91	<.01		
Sentence Processing	.74 (.33)	.21	2.22*	.01		
History Knowledge	.39 (.09)	.30	4.38**	.06		
Model 3					.50	.11**
Vocabulary	.14 (.09)	.13	1.52	.01		
Morphology	.05 (.20)	.02	.23	<.01		
Sentence Processing	.82 (.30)	.24	2.70**	.02		
History Knowledge	.32 (.08)	.25	3.95**	.04		
Football Knowledge	.15 (.02)	.35	6.68**	.11		

*p<.05; **p<.01; SBA=scenario-based assessment.