Testing the Inference Mediation Hypothesis in a Post-Secondary Context

Joseph P. Magliano[1], Karyn Higgs[2], Alecia Santuzzi[2], Stephen M. Tonks[1], Tenaha O'Reilly[3], John Sabatini[4], Daniel Feller[2], Ryan D. Kopatich[2], Melissa Ray[4], & Christopher Parker[2],

1. Georgia State University
2. Northern Illinois University
3. Educational Testing Services
4. The University of Memphis
5. University of Texas, Dallas

To appear in *Contemporary Educational Psychology*

Magliano, J. P., Higgs, K., Santuzzi, A., Tonks, S. M., O'Reilly, T., Sabatini, J., Feller, D., Kopatich, R. D., Ray, M., & Parker, C. (2020). Testing the Inference Mediation Hypothesis in a Post-Secondary Context. *Contemporary Educational Psychology*. https://doi.org/10.1016/j.cedpsych.2020.101867

Contemporary Educational Psychology utilizes a peer review process involving a double blind review process. Papers deemed suitable by the editor after an initial evaluation are typically sent to a minimum of two independent expert reviewers to assess the scientific quality of the paper.

Authors' Note

The present research was supported by the Center for the Interdisciplinary Study of Language and Literacy at Northern Illinois University and a grant awarded to the first author by the Institute of Education Sciences, US Department of Education (R305A150193, R305A190063). All opinions expressed in this manuscript are those of the authors and not the funding agencies. None of the authors have a financial stake in the publication of this research and therefore there are no perceived conflicts of interest.

Abstract

The inference mediation hypothesis (IMH) assumes that individual difference factors that affect reading proficiency have direct and indirect effects on comprehension outcomes, with the indirect effects involving inference processes. The present study tested the IMH in a diverse sample of two and four-year college students in a task that emphasizes comprehension of the passage (traditional assessment) and a task that emphasizes complex problem solving (SBA). Participants were administered assessments of foundational skills that support reading, inference generation, a traditional assessment of comprehension proficiency, and a scenario-based reading assessment. The results support the IMH. However, the strength of the indirect relationships depended on the type of reading performance assessment. Coherence building inferences partially mediated the relationship for both assessments. However, elaborative inferences only partially mediated the relationship for the scenario-based assessment. The results are discussed in terms of theories of purposeful reading and implications for understanding college readiness.

Introduction

An alarming number of students entering their first year of college are not ready to be successful academic readers (Baer, Cook & Baldi, 2006; Bailey, 2009; Greene & Forster, 2003; Jenkins & Boswell, 2002; NAEP, 2015). While the actual number of underprepared college students is unknown, estimates range from 40% to a staggering 90% (Perin & Charron, 2006). These students are at risk of not completing their college degree, which in turn has implications for career success. In fact, it is well recognized that advanced literacy skills are necessary for many professions in the 21st century (Britt, Rouet, & Durik, 2018; Magliano, McCrudden, Rouet, & Sabatini, 2017). College is a critical period for the acquisition and refinement of these advanced skills because students learn to read within the expectations of their professional disciplines (Goldman et al., 2016; Shanahan & Shanahan, 2008). As such, it presents a serious problem for a high percentage of incoming college students, if they are not prepared to meet the reading expectations in college.

To effectively address this problem, one needs to understand what contributes to success in authentic academic reading tasks. This study was conducted to understand some of the aspects that support purposeful reading in academic contexts in a diverse sample of students in two- and four-year institutions. This sample included participants who were identified as not ready to read for college, based on admissions criteria, and therefore were enrolled in supplemental (developmental) programs for improving literacy (reading and writing) and study skills. This study specifically explored some of the literacy skills (e.g., foundational reading skills, inferencing) that support purposeful reading in academic contexts (Britt et al., 2018; McCrudden & Schraw, 2007). In particular, this study tested an *Inference Mediation Hypothesis* (IMH) which assumes that inferences partially mediate the relationship between the foundational skills

that support reading and performance on reading tasks (Cromley & Azevedo, 2007; Kopatich, Magliano, Millis, Parker, & Ray, 2019). We argue that testing this hypothesis with different tasks that vary in the extent that they reflect the complex literacy tasks faced in college will help gain insights into challenges faced by struggling college readers. Below, we describe the nature of purposeful reading in academic contexts and the implications for the IMH.

The Inference Mediation Hypothesis

Reading is aided by a set of skills that support the process of reading and the construction of a coherent mental model (Cromley & Azevedo, 2007; Kopatich et al., 2019; Perfetti & Stafura, 2014). In the context of the present study, we make a distinction between foundational skills and inference processes. Foundational skills range from word (lexical access, decoding) to sentence processing (syntactic processing, proposition construction). After a propositional representation for each sentence is constructed, readers potentially generate inferences that establish how these representations are related to prior discourse context or integrate relevant background knowledge into the mental model. These represent two classes of inferences emphasized by models of comprehension: *bridging inferences* establish the relationships between a given sentence and the prior discourse context (e.g., causal, temporal, and spatial relationships, anaphor resolution), whereas *elaborative inferences* establish how one's relevant background knowledge is related to the discourse content (McNamara & Magliano, 2009).

There is growing evidence that the impact of foundational skills on outcomes associated with purposeful reading are partially mediated through inference processing (Ahmed et al., 2016; Cromley & Azevedo, 2007; Cromley, Snyder-Hogan, & Luciw-Dubas, 2010; Kopatich et al., 2019), which we label the *Inference Mediation Hypothesis* (IMH). For example, Kopatich et al., (2019) had college students think aloud while reading texts. The extent that the students engaged

in bridging and elaborative inferences while thinking aloud was measured. During reading, these participants also answered opened-ended comprehension questions and completed a measure of proficiency in foundational skills. Consistent with the IMH, Kopatich et al., (2019) found that there were both direct and mediational relationships between foundational skills and performance on the comprehension questions. Both bridging and elaborative inferences partially mediated the relationship between foundational skills and comprehension performance, but the relationship was more robust for bridging than for elaboration.

We contend that the nature of the mediational relationship between inferencing and reading comprehension will vary based on the nature of the task associated with reading. In Kopatich et al., (2019) the answers to comprehension questions were in the prior discourse, and the extent to which readers could access that information was likely related to successfully generating bridging inferences. However, when a literacy task requires integrating or reasoning with information beyond the current text, generating elaborative inferences may be of greater importance. In the present study, we explored this possibility by giving college students comprehension tests that were qualitatively different with respect to purposeful reading. In doing so, we can assess the extent that the nature of inference mediation varies across task.

The Nature of Purposeful Reading

Contemporary perspectives of academic reading (and reading that occurs outside of academic contexts) construe it as purposeful and goal-directed; thus, it can also be viewed as a problem-solving activity (Britt et al., 2018; McCrudden & Schraw, 2007; OECD, 2018; Rouet, 2006; Snow, 2002). First, reading is a goal-directed behavior, and as such there is always a purpose behind a decision to read (Graesser, Singer, & Trabasso, 1994), even if that purpose is relatively vague (reading to become familiar with the content of a chapter prior to a lecture,

versus reading to answer specific questions about that chapter for homework). Second, virtually all academic reading activities (i.e., using information in texts, whether in print or electronic format) are grounded in instructor-assigned or self-selected tasks (e.g., preparing for a quiz/test/group discussion, answering questions, writing a paper, performing an assigned project, etc.; McCrudden & Schraw, 2007), and readers have to deploy different strategies to successfully accomplish different tasks (Britt et al., 2018). Third, even when readers are given the same task, they may adopt very different strategies for accomplishing that task (Farr, Prichard, & Smitten, 1990; McCrudden, Magliano, & Schraw, 2010).

Consider a situation in which students are asked to find websites on the internet that help them provide an explanation for why tsunamis are destructive, or a different task in which students are asked to identify steps that can be taken to minimize damage from tsunamis in populated areas. The texts that students find may be written for very different purposes than that of the task at hand, and they will have to extract the information relevant to accomplishing the task (Britt et al., 2018; Goldman, 2011; Goldman, Braasch, Wiley, & Brodowinska, 2012; Goldman, & Scardamalia, 2013; Magliano, et al., 2017). These alternate tasks may require the students to think differently about the text content in ways the author did not originally intend. Thus, reading in an academic context can require a student to understand what a text is about (the author's intended message), but also to determine what information is relevant to their goals, and to process that information in a manner consistent with achieving their ultimate aims (alignment and usefulness with reader goals). The reading strategies that college students adopt can vary as a function of the nature of the task and the instructions (Linderholm & van den Broek, 2002; Narvaez, van den Broek, & Ruiz, 1999; van den Broek, Lorch, Linderholm, & Gustafson, 2001). As such, it is not surprising that success in academic reading tasks has been shown to be

profoundly impacted by how effective students are in applying various reading strategies and comprehension processes (Britt et al., 2018; Cerdán & Vidal-Abarca, 2008; Cerdán, Vidal-Abarca, Martinez, Gilabert, & Gill, 2009; Goldman & Durán, 1988; Ozuru, Best, Bell, Witherspoon, & McNamara, 2007; Pressley & Afflerbach, 1995; Rouet, 2006; Wiley & Voss, 1999).

Assessing Purposeful Reading. We distinguish between two approaches to assess purposeful reading, specifically traditional standardized assessments and scenario-based assessments (SBA). The typical purpose of a traditional standardized assessment is to assess how proficient students are at comprehending the intended messages of texts. In contrast, the purpose of an SBA is to assess students' ability to use texts to solve authentic problems that they may encounter in academic contexts (Sabatini, O'Reilly, Halderman & Bruce, 2014a). The development of SBAs arose in response to the recognition that academic reading tasks often require processes beyond those required to understand a single text in isolation and require skills, such as evaluating, integrating and synthesizing information from multiple sources to make decisions or solve problems (Gordon Commission, 2013; NGA & CCSSO, 2010; McCrudden et al., 2010; Partnership for 21st Century Skills, 2008, Sabatini, et al., 2014a). Table 1 shows how these approaches differ in terms of contexts, tasks, goals, and texts. By contexts, we refer to the extent to which the texts and items are situated within an assessment. In a traditional test, there is typically no context specified beyond instructions to read and answer questions, and as such, the specification of context is minimal, and not related to the activities that students engage in beyond taking standardized tests. In contrast, SBAs provide a more elaborated context that contains characters (teachers, students), a problem that the test taker is given to solve that links all texts and questions, simulated social exchanges between characters, and finally the

assessment ends with items that reflect the ultimate outcome of the task (e.g., problem that the test taker is trying to solve). For example, in the SBA used in the present study, the test takers are given the task of correcting a Wiki on a topic (the historical person that was subject in the painting, The Mona Lisa) by a character who is a college instructor. The instructor and student agents introduce the tasks of reading the texts and answering questions that progress towards completing the primary task. At times the test taker is asked to respond to open-ended items that ask them to reflect on why they were asked to read a particular text. These items are part of the context and intended to increase metacognitive thinking about the texts and items in the assessment form and are not scored.

Table 1

Dimensions of Variation Between Traditional and Scenario-Based Reading Assessments

Dimension	Traditional	Scenario-based
Goals	Answer Question	Complex Problem
Context	Minimal	High
Items	Multiple choice	Variety of types
Texts	Unrelated	Related

By goals, we mean the goal of the test taker. While taking a traditional test, the student can have both local and global level goals. At a local level, the purpose could be dictated at each item on an assessment. That is, each question provides a local task, and across items, the nature of those tasks will differ depending on the knowledge and processes required by those items. For

example, some items may require the test taker to identify a close paraphrase of the content of a text segment (e.g., sentence, paragraph, or entire text); some may require the identification of an inference warranted by a text segment; some may require using content to reason about a topic not explicitly discussed in the texts (Magliano, Millis, Ozuru, & McNamara, 2007). In this case, there are multiple local purposes for reading. At a global level, there is a general purpose for taking an assessment, such as reading to get a high score (Rupp, Ferne, & Choi, 2006). Test takers will adopt strategies at both the local and global levels (Cerdán, Gilabert, & Vidal-Abarca, 2011; Vidal-Abarca, Mañá; & Gil, 2010). In either case, the items on a traditional standardized reading test are typically designed to sample the student's understanding in relation to the author's intended purpose for writing the text. As such, traditional standardized tests of reading comprehension are typically intended to assess a student's ability to closely understand texts, albeit some items may require reasoning beyond the texts. For example, Magliano et al. (2007) did an analysis of the processes required to answer questions in two commonly used tests of comprehension proficiency (Nelson-Denny test of comprehension and the Gates-MacGinitie test of reading comprehension) and found that the vast majority of questions required verifying the meaning of words in sentence contexts, identifying accurate paraphrases, and generating inferences that were closely supported by the texts.

The goals of students taking an SBA can similarly be characterized at local and global levels. However, the intention is that students adopt the goal to accomplish the task that is part of an item's context. Of course, students understand that they are taking an assessment, and may choose to do well on it. Students are asked to embrace the problem that they are given to solve and the intent is that the global goal of doing well on the test becomes secondary to solving the task.

By items, we mean the specific questions that students have to answer as they progress through the test. In a traditional test, the items are typically in a multiple-choice format. There are varying number of questions associated with a series of texts. There is typically no explicit rationale or order to the items associated with the texts, or to the ordering of the texts. In contrast, the progression of texts and items in an SBA are carefully crafted such that they lead to the completion of the task. SBAs contain a variety of item types, such as multiple-choice questions, open-ended questions, and summarization of texts. At different points during the assessment, students are also required to evaluate the relevance of information in relation to the goal for reading and state what evidence would strengthen or weaken claims. These items are intended to help the student engage in the context and adopt goals associated with it.

Finally, traditional tests typically contain a sequence of unrelated texts on a topic for which students are likely to be unfamiliar (with the intent of reducing the impact of prior knowledge on test performance). In contrast, the texts in the SBAs are all related in different ways. Some texts may provide contradictory information that varies in reliability, whereas others may be convergent. As such, SBAs reflect the multiple documents situation that is inherent in many literacy activities within and outside of academic contexts (Britt et al., 2018; Rouet, Britt, & Durik, 2017; Sabatini, O'Reilly, Halderman, & Bruce, 2018).

In the present study, we used two standardized tests that reflected qualitatively different types of purposeful reading. The first assessment was similar to a traditional reading comprehension assessment in that it required test takers to answer questions related to the meaning of a single text. This included questions about key ideas, details, and inferences that connected key information. For this assessment there was no globally stated purpose for reading. It was expected that students would construct mental models of the single texts that were in line

with the author's intended meaning. That is, the assessment was intended to include items that reflect the extent that readers can accurately represent content and generate inferences afforded by the texts. The second assessment was a scenario-based assessment on the Mona Lisa topic discussed above. It requires the evaluation and integration of multiple sources to uncover who was the model in the Mona Lisa painting. In line with prior findings that different types of reading purposes result in different types and degrees of inferential activity (Linderholm & van den Broek, 2002; Narvaez, et. al,, 1999; van den Broek, et al., 2001); we suspected that the type of inference processes demanded by each type of comprehension assessment might be different, and as described below, may differentially mediate the relationship between foundational reading skill and comprehension outcomes for the two types of assessments.

Overview of the Current Study and Research Questions

The goal of the present study was to test the IMH in a diverse sample of two- and four-year college students, and specifically in a task that emphasizes comprehension of the passage (traditional assessment) and a task that emphasizes complex problem solving (scenario-based assessment). Participants completed assessments of foundational reading skills (i.e., word recognition and decoding, vocabulary, morphological knowledge and sentence processing), inferencing, a traditional assessment of reading comprehension, and a scenario-based assessment of reading comprehension. To test the IMH, we posed the following research questions:

RQ 1: Are foundational skills differentially predictive of traditional and scenario-based assessments of comprehension skill?

RQ 2: Are bridging and elaborative inference strategies differentially predictive of traditional and scenario-based assessments of comprehension skill?

RQ 3: Does level of foundational skills indirectly relate to traditional and scenario-based reading comprehension outcomes through inferencing strategies?

RQ 1 and 2 are preliminary to RQ3, which tests the IMH with a traditional assessment and an SBA. However, the answers to both preliminary questions are interesting and important in their own right. With respect to RQ1, given that both traditional assessments and SBAs require test takers to read texts, one possible answer is that proficiency in foundational skills will account for similar variance in both types of assessments. However, the SBAs were designed to assess processing skills that go beyond reading and responding to items. If SBAs require complex problem-solving behaviors that go beyond those that are typically employed when taking standardized test, as intended by the test maker, then foundational skills may account for less variance in an SBA than a traditional assessment. In support of this possibility, Sabatini et. al., (2014a) found that for middle school students, low levels of foundational skills limited performance on an SBA, however, higher levels of foundational skills did not necessarily lead to higher performance on the SBA. Thus, foundational skills may be necessary, but not sufficient for skilled performance on the complex literacy tasks assessed by the SBA (Sabatini et al., 2014a). With respect to RQ2, as previously discussed in the context of the IMH, it is possible that elaborative processes may be more important in contexts that require problem solving and reasoning beyond basic text comprehension. As such, elaboration may be more strongly correlated with SBA performance than with the traditional assessment. (See also LaRusso et al., 2016 for evidence of cognitive skills beyond basic text comprehension in SBAs).

It is important to emphasize that 58% of the sample of students in this study were designated as not ready for the literacy demands of college and were enrolled in a developmental educational program intended to improve college literacy readiness. Students were recruited

from these programs to ensure that the sample in this study reflected a range of college readiness to read. However, exploratory analyses were conducted to assess if enrollment in these programs moderated the paths tested in the final model test in RQ3, which would indicate that the associations between foundational skills and performance on the reading assessments might vary depending whether or not college students are designated as struggling readers, as indicated by their enrollment in a developmental course.

Methods

Participants

A total of 434 students from a large, 4-year institution in the Midwest, a community college in the Southwest, and a community college in the Northeast participated in at least one of the two study sessions. See Table 2 for demographics. Fourteen students were dropped from hypothesis tests because they were missing data on the SBA

In the full sample, there were 263 students from the four-year institution, and 171 students from a two-year institution. The majority of participants were first year students and included participants who were enrolled in a developmental literacy program and those who were not. Across all institutions, 58% (n = 251) of students were designated as needing additional support in the form of a developmental literacy program. At the four-year midwestern university, 141 students were enrolled in one of two courses intended to support college reading and college study strategies. These participants were required to take one or both of these courses as part of their enrollment in a program that admits students who do not meet the criteria for traditional admission to the university. For admittance to this program students were required to have a minimum high school grade point average of a 2.0 and a minimum ACT composite test score of 17 (composite score); SAT composite of 910 or a percentage rank of 70 percentile or

higher in their graduating class. All students in the non-traditional admittance program were then administered the Accuplacer test (College Board, 2019) and placed into the developmental reading courses based on scores on that test.

Table 2
Demographic Information for Participants in the CFA

Participant Information	Total	Proportion
Participant count	434	
Developmental enrollment (DE)	251	0.58
DE	251	0.58
not DE	149	0.34
no info	34	0.08
School Type		
2 year	171	0.39
4 year	263	0.61
Sex		
Female	245	0.56
Male	155	0.36
no response	34	0.08
First Language		
English	313	0.72
Not English	99	0.23
no response	22	0.05
Race/Ethnicity		
Black/African American	179	0.41
White	109	0.25
Asian	52	0.12
Hispanic/Latino	70	0.16
American Indian/ Alaska Native	3	0.01
Native Hawaiian/ Pacific Islander	1	>0.01
No Selection	20	0.05
Age Range		
18-22	341	0.79
23-37	32	0.07
38-55	7	0.02
no response	54	0.12

Another 110 of the participants in DE courses were enrolled at the southwestern community college. This community college is an open enrollment school that utilizes the TSI (a Texas version of the Accuplacer test) to assess the need for developmental coursework. Based on their scores, students could be required to participate in a developmental reading program that consists of two eight-week courses. The developmental reading program works closely with other programs at the school (e.g., English, history, government, psychology and biology) and uses textbook examples from these courses to help students prepare for academic reading in their first English course and in other disciplines.

The sample also includes 22 students from a northeastern community college. These students were recruited from either developmental reading or writing skills courses. However, information about which course the students were enrolled in was unavailable. As such these students were coded as missing information about enrollment in developmental reading courses and were not included in analyses were DE enrollment status was used.

Compensation varied across the locations. Participants received either monetary compensation, course credit or gift certificates for participating in each session (or a combination of money and course credit across sessions).

Statement of ethics compliance

The research presented in this article was reviewed by an institutional human subjects compliance board and all participants signed an informed consent form before their participation.

Data access

The data for this study is accessible on Open Science Framework (https://osf.io/5pgrc/)

Materials

Foundational reading skills. A measure of general foundational reading skills was obtained based on the Study Aid and Reading Assessment (SARA: O'Reilly, Sabatini, Bruce, Pillarisetti & McCormick, 2012; Sabatini, Bruce, Steinberg & Weeks, 2015; Sabatini et al., 2019). This assessment measures multiple components of reading using a sequence of subtests that reflect a continuum of component reading skills. In the current study we utilized four of the six subtest scores to measure foundational skills (word recognition and decoding, vocabulary, morphology and sentence processing). The assessment has been tested with tens of thousands of students and demonstrates high reliability (five of six subtests have Cronbach's $\alpha > .88$) and has evidence of concurrent validity in predicting state test scores (O'Reilly et al., 2012; Sabatini, et al., 2015; Sabatini et al., 2019).

Inference processes. Inference processes were assessed with the Reading Strategy Assessment Tool (RSAT; Magliano, Millis, The RSAT Development Team, Levinstein, & Boonthum, 2011). RSAT is a computer-based assessment tool that provides measures of processes supporting comprehension of texts, in particular (1) bridging inferences (2) elaborative inferences.

The RSAT measures are obtained by having participants produce typed, open-ended verbal protocols using a variant of think-aloud instructions. Texts are presented one sentence at a time and participants advance to the next sentence at their own pace. Participants can see only the current sentence. After target sentences, participants see the prompt "What are you thinking now?" appear on the screen and type their responses into a text box beneath the prompt.

RSAT uses computational algorithms, based on keyword matching, to assess the extent to which words from a participant's protocol overlap with words from the text (see Magliano et al.,

2011). The bridging score is generated based on the number of content words from prior sentences. The elaboration score is generated based on the number of content words in the participant's response that were not present in the prior discourse context.

RSAT process measures have been shown to have respectable validity and reliability. RSAT bridging and elaboration scores are highly correlated with human judgments of the presence of these processes (i.e., $.50 < r < .78$; Magliano et al., 2011). RSAT processing scores are also correlated with the Gates-MacGinitie, and the comprehension portion of the ACT (*r*'s ranging from .51-.55), roughly to the same extent the two measures correlate with one another (*r* = .59; Gilliam, Magliano, Millis, Levinstein & Boonthum, 2007; Magliano et al., 2011). Finally, test-retest reliability of the automated scores is high, particularly when considering the open-ended nature of the assessment (*r*'s = .79 for bridging and elaboration scores).

In the current study, participants read two texts in RSAT, presented in a randomized order. Participants read a history text ("Louis XVI and the French Revolution", 19 sentences) and produced verbal protocols at 6 locations, and a science text ("The Power of Erosion", 22 sentences) in which they produced protocols at 7 locations.

Traditional measure of reading comprehension. The traditional assessment of reading comprehension was provided by the Reading Comprehension subtest of SARA (Sabatini et al., 2019). This test involved answering 22 multiple choice questions associated with three texts. The reading comprehension subtest of the SARA is designed to measure students' basic understanding of a single text (i.e., there are no cross-passage items). Some items require the test taker to locate key ideas and important details in the text. Successful performance on these items may require a test taker to be able to recognize paraphrases. The second class of items requires the test taker to draw inferences. These item types include local or bridging inferences (e.g.,

resolve an anaphoric referent across adjacent sentences), global inferences (connecting information across multiple distant sentences) and some knowledge-based inferences (requiring a connection to general background knowledge).

Scenario based measure of reading comprehension. Scenario-based reading was assessed using a form of the Global, Integrated, Scenario-based Assessment (GISA) (O'Reilly & Sabatini, 2013; Sabatini, O'Reilly & Deane, 2013; Sabatini, O'Reilly, Weeks, & Zang, 2019) developed for high school students, but adapted for this study. In the GISA, items are grounded in an academically authentic task; students are provided with a global purpose for reading a collection of thematically related texts (e.g., the need to correct a wiki on a historical topic). Simulated teacher and student agents contextualize each item in the task, help to structure and scaffold the tasks, as well as provide test takers an opportunity to identify and correct errors expressed by the simulated students. Unlike many off-the-shelf reading assessments that measure the piecemeal understanding of single texts, the GISA provides test takers with a realistic, domain-specific purpose for reading a collection of sources and materials. This allows for the measurement of skills associated with higher-level comprehension such as knowledge of text structure, evaluation, application, perspective taking and integration of information in service of completing a goal (see, Bennett, 2011; O'Reilly & Sabatini, 2013; O'Reilly & Sheehan, 2009; Sabatini et al., 2013; Sabatini, et al., 2018). The GISA has been shown to be reliable in elementary through high school populations as evidenced by good internal consistency (Cronbach's $\alpha > .80$; O'Reilly, Weeks, Sabatini, Halderman, & Steinberg, 2014) and test-retest reliability ($r = .87$; Sabatini, O'Reilly, Halderman, & Bruce, 2014b). Additionally, the GISA has robust correlations with other reading measures such as English language arts state test scores ranging from .52 to .68 (O'Reilly et al., 2014) and correlates with measures of deep

understanding including academic vocabulary, complex reasoning, and perspective taking (LaRusso et al., 2016). The items cover a broad range of difficulty with no apparent floor or ceiling effects when used with intended populations (see McCarthy et al., 2018; O'Reilly et al., 2014; Sabatini, Halderman, O'Reilly, & Weeks, 2016; Sabatini, et al., 2014b).

The version of the GISA used in the current study involved a scenario in which students were asked to update and correct a wiki about the Mona Lisa. Through interaction with various texts and the GISA agents, participants are tasked with identifying the problem with a wiki (i.e., conflicting theories about the identity of the person depicted in the painting of the Mona Lisa), and suggest how to update the wiki.

Students completed sections of the test that included multiple-choice (MC), constructed-response (CR), and graphic organizer (GO) items. More specifically, the GISA form used in this study required the use of a host of skills including: identifying evidence to support a theory; identifying contradictions across sources; perspective taking; identifying evidence that may question the credibility of a source; identifying problems with a theory; identifying a relevant web source; identifying missing evidence that would strengthen or weaken a theory; categorizing evidence to support two different theories and providing feedback about the accuracy of a blog post.

Procedure

The study consisted of two sessions. All measures were computer-based and accessed via web links. Instructions for each measure were provided on the websites. All participants completed session one in a computer lab with trained study administrators. At the four-year institution, all participants completed Session 1 outside of class in either a small group or individual session. At the community colleges, some participants completed Session 1 during

class time, and some completed it outside of class time. In all locations, Session 2 took place outside of class. At the four-year institution, Session 2 was administered by trained personnel in either small group or individual sessions. At both community colleges, Session 2 was self-administered with students completing the session on their own.

Session 1 took students between 60-90 minutes to complete. During the session, participants first completed the SARA, followed by RSAT. In RSAT, participants were given instructions and then engaged in a practice text to familiarize themselves with the presentation format and responding to the prompt. Participants were instructed that when they saw the prompt "What are you thinking now?", they were to type their thoughts about their understanding of what they had just read in terms of what they had already read and what they know about the topic. Participants then engaged in the practice text. During the practice, participants were given feedback when their responses were less than five words (i.e., "We are interested in your thoughts about the texts, in your responses to the prompts, please tell us more about your understanding of what you are reading."). After the practice, participants read the two experimental texts in a randomized order and responded to the prompts. No feedback was provided during the experimental texts.

Session 2 took participants between 60-90 minutes. During the session, participants first completed the GISA. This was followed by several measures not utilized in the current analyses including a situational motivation measure grounded in the GISA and additional metacognitive and motivational measures. The final measure completed was a demographic survey.

Analysis

The research questions were tested using path models in Mplus v. 8.3. An aggregate latent factor representing foundation skills was created from four SARA subscales (decoding and

word recognition, vocabulary, morphology, and sentence processing). When this factor was included in the models for testing, the overall model fit was evaluated using the model Chi-square, Root Mean Square Error of Approximation (RMSEA), and Confirmatory Fit Index (CFI). For RMSEA, values less than 0.08 will be evaluated as good fit, values between 0.08 and 0.10 indicate mediocre fit, and values above 0.10 indicate poor fit (Steiger, 2007). For CFI, values greater than 0.95 will be evaluated as good fit, values between 0.90 and 0.95 show mediocre fit, and values below 0.90 demonstrate poor fit (Hu & Bentler, 1999).

To test RQ1, foundational skills (latent) was specified to predict reading performance on both the traditional reading comprehension measure and scenario-based assessment. Including both outcomes in the same model allowed us to compare the strength of the associations between the two outcomes while accounting for the shared variance between the measures. One model allowed each association between foundational skills and reading performance to be estimated freely (Free Model). A second model (Constrained Model) fixed the associations between foundational skills and each outcome to be equal to each other. If the model fit for the Constrained Model decreased significantly compared to the Free Model (using a Chi-square comparison test), this would provide evidence that the associations with each measure of reading are not equal (i.e., should not be constrained to be equal). RQ2 was examined with two separate path models that specified either bridging or elaboration as a predictor of the two reading performance outcomes. Finally, RQ3 was initially examined with separate models for each mediating variable. The models specified pathways from foundational skills through either bridging or elaboration to both traditional reading comprehension scores and scenario-based assessment scores. This enabled an assessment of the IMH for each type of inference. This was followed by a parallel indirect effects model in which foundational skills (latent) was specified as

the predictor of both reading outcomes through bridging and elaboration (parallel mediators), which enabled an assessment of the relative contributions of the two inference types on the two literacy tasks. Moreover, it afforded testing whether there are differences in the mediational relationships as a function of inference type and the nature of the literacy task. The direct effect between foundational skills and both reading outcomes was included in the model. The indirect effect estimate was computed for foundational skills to each outcome via each of the two inference processes (thus, four possible indirect paths). The indirect effect estimates were tested for statistical significance.

Results

Descriptive statistics for the measures are shown in Table 3 and bivariate correlations between the measures are shown in Table 4. The results are divided into four sections. Specifically, the first section presents a preliminary specification and testing of a formative latent variable for foundational skills, followed by three sections addressing the research questions.

Preliminary Findings

A formative latent variable was specified and tested to determine if the subtests of SARA created an aggregate latent factor representing foundational skills. The results of the model supported the construct validity for the foundational skills as a formative latent construct. Each of the four subtests significantly contributed to the formative construct ($p < .01$): word recognition (.79), vocabulary (.55), morphology (.39), and sentence processing (.15)[1]. As such, they confirmed the validity of using the aggregate of SARA subtests to represent foundational skills in subsequent analyses.

[1] Note that a formative measurement model does not have fit indices as it is a saturated model (df = 0).

RQ1: Are foundational skills differentially predictive of traditional and scenario-based assessments of comprehension skill?

Table 3

Descriptive Statistics for Measures

Measure and Sub-scores	n	M	SD	α	Std. Loading
Scenario Based Assessment (GISA)	420	16.73	5.41	--	--
Traditional Assessment (SARA-RC)	434	12.27	4.32		
Student Aide and Reading Assistant (SARA)					
Word recognition and decoding	434	37.42	9.62	.92	.83
Vocabulary	434	26.89	6.05	.87	.85
Morphology	434	29.01	7.64	.93	.89
Sentence Processing	434	20.04	4.37	.85	.79
Reading Strategies Assessment Tool (RSAT)					
Bridging score	420	1.60	1.04	--	--
Elaboration Score	420	2.99	1.87	--	--

Table 4

Bivariate Correlations of Variables

Variables	1	2	3	4	5	6	7
1. GISA (SBA)	--						
2. SARA Comp (Traditional)	.68*	--					
3. SARA Word	.54*	.57*	α=.91				
4. SARA Vocab	.64*	.65*	.74*	α=.872			
5. SARA Morphology	.54*	.57*	.72*	.75*	α=.934		
6. SARA Sentence	.57*	.61*	.63*	.64*	.74*	α=.853	
7. RSAT bridge	.27*	.35*	.24*	.27*	.23*	.22*	--
8. RSAT elaboration	.37*	.33*	.28*	.30*	.30*	.24*	.40*

Note: * indicates significance at $p < .05$. Alpha reliabilities for each SARA measure are shown on the diagonal

A model tested the predictive strength of foundational skills for each of the two types of reading outcomes ($\chi^2(6) = 24.91$, *CFI* = .98, *RMSEA* = .085, *SRMR* = .02). Parameter estimates indicated that foundational skills significantly and positively predicted scores on the traditional reading test ($\beta = .70$, $p < .001$) and scenario-based assessment ($\beta = .67$, $p < .001$).

When the relationships to both reading outcomes were constrained to be equal ($\chi^2(7) = 36.52$, *CFI* = .97, *RMSEA* = .099, *SRMR* = .06), the model fit decreased significantly (χ^2 diff (1)=11.61, $p<.001$). This suggests that the slopes between foundational skills and each reading performance outcome should not be constrained to be equal. In other words, the evidence supports the conclusion that foundational skills differently predict reading performance

depending on the assessment. As demonstrated in the unconstrained model reported above, foundational skills showed stronger prediction of traditional reading test scores than the scenario-based assessment performance.

RQ2: Are bridging and elaborative inference strategies differentially predictive of traditional and scenario-based assessments of comprehension skill?

Two separate path analysis models examined 1) the role of bridging across the two types of assessments and 2) the role of elaboration across the two types of assessments. When the role of bridging in reading outcomes was examined, bridging significantly predicted both traditional ($\beta = .35$, $p < .001$) and scenario-based performance ($\beta = .27$, $p < .001$). Note that the standardized weight is stronger for traditional than scenario-based assessment. However, the difference in beta weights was calculated using the systemfit package in R (Henningsen & Hamann, 2007; R Core Team, 2018) and there was no significant difference for bridging, $\chi^2(1) = .07$, $p = .789$.

When elaboration was tested in the prediction of reading outcomes, elaboration significantly predicted both traditional ($\beta = .33$, $p < .001$) and scenario-based reading performance ($\beta = .37$, $p < .001$). Note that the weight is stronger for the scenario-based assessment compared to the traditional assessment. This pattern for elaboration is different from that observed for bridging. However, the difference in beta weights did not reach significance, $\chi^2(1) = 3.41$, $p = .065$.

RQ3: Does level of foundational skills indirectly relate to traditional and scenario-based reading comprehension outcomes through inferencing strategies?

The IMH was tested for both reading outcomes with separate models for each mediating variable. A model specified pathways from foundational skills through each of the process

variables (bridging or elaboration) to potentially predict both traditional reading comprehension scores and performance on the scenario-based assessment.

When bridging was specified as the process variable to both types of assessments, the model fit well, ($\chi^2(9) = 26.06$, $CFI = .98$, $RMSEA = .066$, $SRMR = .02$). Foundational skills directly predicted the traditional ($\beta = .65$, $p < .001$) and scenario-based assessment ($\beta = .64$, $p < .001$). Foundational skills also predicted bridging ($\beta = .27$, $p < .001$)). Bridging predicted performance on the traditional assessment ($\beta = .17$, $p < .001$) and the scenario-based assessment ($\beta = .09$, $p = .019$). Moreover, the indirect effects for foundational skills through bridging to the traditional assessment ($ab = .05$, $p < .001$) and scenario-based assessment ($ab = .02$, $p = .029$) were significant.

When elaboration was specified as the process variable to both types of assessments, the model fit well, ($\chi^2(9) = 29.05$, $CFI = .98$, $RMSEA = .072$, $SRMR = .02$). Foundational skills directly predicted performance on the traditional ($\beta = .66$, $p < .001$) and scenario-based assessment ($\beta = .61$, $p < .001$). Foundational skills also predicted elaboration ($\beta = .32$, $p < .001$)). Elaboration predicted the traditional assessment ($\beta = .12$, $p = .001$) and the scenario-based assessment ($\beta = .18$, $p < .001$). Moreover, the indirect effects for foundational skills through elaboration to the scenario-based assessment ($ab = .06$, $p < .001$) and traditional assessment ($ab = .02$, $p = .002$) were significant.

The process variables of bridging and elaboration also were tested as parallel mediators in the same model ($\chi^2(13) = 82.47$, $CFI = .94$, $RMSEA = .11$, $SRMR = .06$). See Figure 1 for full model specification and parameter estimates. The direct effects of foundational skills were positive and statistically significant for both traditional reading comprehension scores ($\beta = .63$, $p < .001$) and the scenario-based assessments ($\beta = .61$, $p < .001$). Foundational skills positively

predicted both bridging ($\beta = .27$, $p < .001$) and elaboration processes ($\beta = .32$, $p < .001$). In this model, bridging significantly predicted higher traditional reading comprehension scores ($\beta = .15$, $p < .001$), but not scenario-based assessment performance ($\beta = .03$, $p = .44$). By comparison, elaboration predicted scenario-based assessment performance ($\beta = .17$, $p < .001$), but not traditional comprehension ($\beta = .07$, $p = .08$).

While the path models assessing bridging and elaborative inferences separately suggest that both partially mediate the relationship between foundational skills and performance on both assessments, the final model testing both inference processes as parallel mediators suggests that the relative strength of this relationship may vary by inference type and the nature of the task. Foundational skills may be more strongly related to traditional reading comprehension scores through bridging processes, or they may be more strongly related to scenario-based reading performance through elaboration processes. Results from the indirect effects analysis supported the viability of both pathways. Bridging provided a significant indirect route from foundational skills to reading comprehension on the traditional test ($ab = .04$, $p = .001$). Elaboration provided an indirect route from foundational skills to scenario-based reading performance ($ab = .05$, $p = .001$). Although small in magnitude, these indirect effects provide some support for bridging and elaboration as mechanisms to success depending on the type of reading assessment.

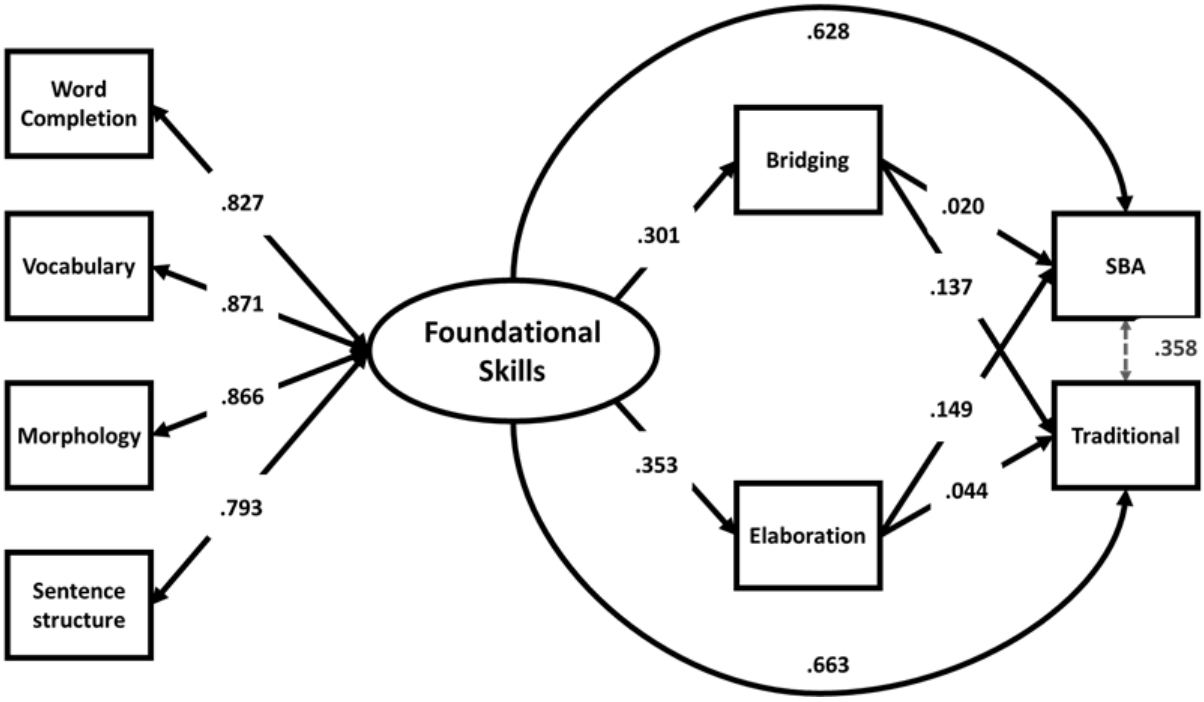

Figure 1. Inference Mediation Hypothesis Model

Exploratory analyses were conducted to assess if developmental status moderated the paths in the final model. While enrollment in developmental reading programs is based on an assessment of foundational skills (i.e., performance on the Accuplacer test or the Texas variant of it), other factors associated with students in these programs may account for variance in inferences processes, performances on the two task (Feller, Magliano, O'Reilly, Sabatini, & Kopatich, in press), and the mediation paths.

The model with both bridging and elaboration as parallel mediators of foundational skills to the two types of assessments was modified to examine moderating effects of developmental education status on each effect in the indirect effects model. This required both developmental status (0, 1) and its interaction with predictor variables to be included in the model. The majority of interaction effects showed that developmental status did not change the predictions demonstrated in the original model. Developmental status did not significantly moderate the

association between foundational skills and bridging (β = -.04, p = .43), elaboration (β = .01, p = .90), or the direct effect on traditional reading assessment scores (β = .002, p = .96). There was a small but significant interaction between developmental status and the direct effect of foundational skills on the scenario-based assessment (β = -.08, p = .04), such that those enrolled in developmental education programs showed a smaller association between foundational skills and assessment performance. Note, however, that this only emerged on the direct effect, not the indirect effect paths of interest.

The association between bridging and the assessments was not significantly moderated by developmental status (traditional: β = .04, p = .56; scenario-based: β = .09, p = .30). Similarly, the association between elaboration and the scenario-based assessment scores was not moderated by developmental status (β = -.12, p = .24), nor was there significant moderation of the traditional assessment performance (β = -.17, p = .05). Thus, the effects contributing to the inference mediation effects described in RQ3 appear consistent across developmental education and other students in the sample.

Discussion

The goal of the present study was to test the IMH (Kopatich et al., 2019) in the context of two assessments that reflect different literacy tasks. The traditional assessment involved questions that required close comprehension of the text and reflect the extent that test takers had an accurate representation of text content and could identify basic inferences needed to comprehend the texts. The SBA reflected complex problem solving with texts in which the test takers had to respond to items that reflect using text content to accomplish a goal that extends beyond the content of any one text in the assessment. The IMH assumes that the relationships of foundational skills on performance on these two assessments would be partially mediated by

inference ability. To this end, we assessed these mediational relationships with a measure of bridging and a measure of elaboration.

Testing the IMH was decomposed into three research questions. The first question (RQ1) pertained to whether foundational skills were similarly predictive of performance on the traditional assessment and the scenario-based assessment. The results indicated that the foundational skills predicted a significant amount of variance in both assessments, but that foundational skills were more strongly correlated with the traditional test than the scenario-based test. Both assessments require reading, and the requisite knowledge and skills that support reading proficiency (e.g., Sabatini et al., 2014ab). Foundational skills accounted for less variance in the SBA that required readers to go beyond comprehending a single text and to reason with and problem solve with multiple texts. However, it is also important to acknowledge that the traditional assessment was part of the same suite of assessments as the one that provided the assessment of the foundational skills, which tempers this conclusion. It is appropriate to replicate this finding with assessments of foundational skills that are independent from the traditional comprehension assessment. If the scenario-based assessment tasks require problem solving beyond demonstrating basic comprehension, then these findings should replicate.

RQ2 pertained to assessing the relationships between bridging and elaborative inferences and performance on the two assessments. Both bridging and elaborative inferences were predictive of performance on the traditional assessment and the SBA. These results make sense to the extent that comprehending the passages was necessary to answer the questions in both assessments. Theories of comprehension universally assume that these two classes of inferences are necessary for successful comprehension (McNamara & Magliano, 2009). It is interesting to note that the magnitudes of the effects were such that there was a suggestion of a stronger

relationship between elaboration and performance on the SBA than for the traditional assessment, albeit the difference between them was not significant. The pattern had interesting implications for the final model used to test RQ3.

Finally, RQ3 pertained to directly testing the IMH. Consistent with prior research testing this hypothesis, (Ahmed et al., 2016; Cromley & Azevedo, 2007; Cromley et al. 2010; Kopatich et al., 2019) we found evidence that inferences mediate the relationship between foundational skills and performance on both the traditional and SBA assessments. When assessed in isolation, there was evidence for the IMH for both bridging and elaborative inferences and with both tasks as outcomes. However, the final model suggests that nature of the mediational relationship differed for the two assessments, consistent with our prediction. Specifically, bridging inferences mediated the relationship between foundational skills and performance on the traditional assessment, whereas elaborative inferences mediated the relationship for the SBA. These results further suggest that the literacy tasks in the two assessments might be qualitatively different and are differentially supported by foundational and inference skills. The impact of reading proficiency on traditional assessments may be partially explained by the participants' ability to establish relationships between discourse constituents. Conversely, the ability to read proficiently likely frees up resources to engage in the extratextual elaboration that is required to successfully respond to the items on the scenario-based assessment. The replication of support for the IMH strongly suggest that models of reading comprehension (e.g., Graesser et al., 1994; Kintsch, 1988; 1998) and task-oriented reading (e.g., Britt et al., 2018) should be sensitive to this mediational relationship. It also lends robust support for models of reading that directly incorporate it into their assumptions, such as the Direct and Inferential Mediational Model of reading comprehension (Cromley & Azevedo, 2007).

In the current study we found support for the IMH using different measures of inference processes than other researchers who have tested it (Ahmed et al., 2016; Cromley & Azevedo, 2007; Cromley et al., 2010). For example, Cromley and Azevedo (2007) developed a multiple-choice assessment that required participants to identify appropriate inferences from a set of foils. The items assessed three types of inferences based on Oakhill and Yuill's (1996) classification, specifically, resolving anaphoric referents (e.g., the referent to a pronoun), text-to-text inferences, and background knowledge-to-text inferences. The assessment was intended to measure general proficiency in inference generation. In the present study, we adopted an approach similar to Kopatich et al. (2019), and relied on typed "think-aloud" protocols, which are sensitive to inference processes (Muñoz, Magliano, Sheridan, & McNamara, 2006). The primary difference between the present study and Kopatich et al. (2019) in terms of measuring inferences is that the present study used the computer-based scoring of the protocols and Kopatich et al. (2019) relied on human coding. Certainly, RSAT and the inference assessment measure of Cromley and Azevedo (2007) are different. RSAT bridging scores are sensitive to anaphor resolution and text-to-text inferences and elaboration scores to knowledge-to-text inference (Magliano et al., 2011). However, RSAT does not provide an assessment of the correctness of the inferences or proficiency in generating them. Rather, RSAT is sensitive to the propensity to engage in bridging and elaboration. Providing evidence for the IMH with different measures of inferencing provides robust support for it.

It is important to note it was expected that elaborative inferences would mediate the relationship for the traditional assessment, given that was the case for Kopatich et al., (2019), who also used verbal protocols to measure the tendency to engage in elaboration and bridging inferences. However, Kopatich et al. (2019) assessed comprehension in a different task, that

involved answering open-ended why and how questions. Answering these questions required both making connections across text constituents and to some extent elaborative processing. The traditional comprehension assessment had items that involved making text-to-text connections (e.g., bridging inferences), but answering these items may have required relatively little elaborative processing. Kopatich et al. (2019) and the present study suggest that the inference mediation relationship is likely complex and varies in terms of task and the extent that different inference skills are needed to complete them. Akin to a transfer-appropriate processing perspective (Morris, Bransford, & Franks, 1977), the nature of the mediational relationship may be contingent on the extent that a type of inference processing is involved in the task that provides the outcome measure.

It is important to note that the coefficients reflecting the paths relating bridging and elaborative inferences to performance on the tasks are small. This may be due to the fact that RSAT assesses the propensity to generate bridging and elaborative inferences and does not provide an assessment of the quality of those processes. Perhaps the relationship would be more robust if there were a reliable and valid assessment of quality of both types of inferences, but to our knowledge none exist at this juncture. Moreover, it is important to acknowledge that RSAT measures are more robustly correlated with tasks that requires constructing responses (e.g., open-ended short answer question) than standardized tests based on closed responses (e.g., multiple-choice questions; Magliano et al., 2011),

Although model fit statistics were evaluated as good for the models testing each inference process (bridging and elaboration) as mediators in separate models, the final model testing both processes as parallel mediators showed mediocre fit, with a CFI slightly less than .95 and RMSEA at .11. This suggests that there are potentially important covariances in the data that are

not accounted for in our model. Background knowledge certainly has an important role in predicting performance on comprehension outcomes (Cromley & Azevedo 2007; Ozuru, Dempsey & McNamara, 2009). It has been shown that the relationship between text-relevant background knowledge and performance on standardized tests is partially mediated by some foundational skills (e.g., vocabulary and word processing; Cromley & Azevedo 2007). Future tests of the IMH should include these measures and particularly in the case where DE students in supplemental support programs are involved.

Over the past two decades, there has been a substantial increase in research on the impact of task on reading processes and outcomes (e.g., Britt et al., 2018; Kaakinen & Hyönä, 2005; McCrudden, et al., 2010, McCrudden & Schraw, 2007; Rouet & Britt, 2011; van den Broek et al., 2001; Vidal-Abarca, Salmerón, & Mañá, 2011; Wiley & Voss, 1999). One possible reason for this increased interest was the Reading Comprehension Framework proposed in the influential Rand Report on reading comprehension (Snow, 2002). That framework provided an argument that literacy activities need to be contextualized as a complex interaction between the reader, text, and task. Perhaps partially in response, there have been several theories of task-oriented reading that have been proposed during this time frame (Britt et al., 2018; McCrudden & Schraw, 2007; Rouet, 2006; Rouet & Britt, 2011). Indeed, traditional models and theories of comprehension have typically been agnostic about the impact of task on processing and comprehension outcomes (McNamara & Magliano, 2009). While the present study was not designed to test theories of task-oriented reading, it certainly lends credence to the need for them. Specifically, the differences in the mediational relationships across the two assessments is consistent with arguments that task affects how processes that support comprehension are deployed (Graesser et al., 2004; Magliano, Trabasso, & Graesser, 1999). Theories of task-

oriented reading can provide explanatory mechanisms as to why, but future research is needed to support such a theory-based explanation for the present results.

Both the traditional assessments and the SBAs can be construed as engaging task-oriented reading, but they involve different problem-solving skills. Both assessments were developed with an assessment framework that was sensitive to the nature of the task within them (Sabatini, O'Reilly, Weeks, & Zang, 2019; Sabatini et al., 2019), largely because they followed an evidence-centered design approach in their development (Mislevy & Haertel, 2006; Mislevy, Steinberg & Almond, 2003; Pellegrino & Chudowsky, 2001). This approach requires the test designers to develop a cognitive framework that specifies the processes that are theoretically important for the assessment context, which is then triangulated with item design and data interpretation. An Evidence-Centered design has been deployed in a number of recent assessments, such as the Programme for International Student (PISA) and Programme for the International Assessment of Adult Competencies (PIAAC) assessment (OECD, 2018). The results of the present study underscore the importance of this approach. If tasks affect processing and outcomes, then test designers need to build assessments with this in mind, and must do so in a manner such that the assessments involve the skills and processes under consideration. Treating comprehension as a monolithic assessment construct is a practice that is unfortunately prevalent, and problematic for both research and applied contexts.

The results of this study seem to lend support for the frameworks used to develop the assessments. The traditional reading comprehension test was designed to capture elements of students' mental model of a single text that was consistent with the author's intended purpose for writing. This process involves making connections among ideas in the text. Indeed, the results suggest that foundational skills were critical for performing well on this more traditional

assessment as well as the students' ability to draw bridging inferences. In contrast, the SBA was designed to measure students' ability to integrate, evaluate and synthesize information to achieve a particular goal. While both foundational skills and bridging inferences were predictive of this more complex type of comprehension, students' ability to draw in other information (i.e., elaborative inferences) was also important. One interpretation is that the deeper comprehension required to integrate texts and solve a problem, also involves drawing upon information that is not included in any of the texts (O'Reilly, Sabatini, & Wang, 2018).

The sample in the present study contained a high proportion of students assigned to developmental education programs for literacy. These students were recruited to ensure that there was a diverse sample of reading proficiencies. The exploratory analyses did not show that developmental status moderated the mediational paths of interest, which indicates that the IMH applies to developmental and non-developmental students. However, developmental status did moderate the direct path between foundational skills and performance on the SBA, such that the relationship was weaker for developmental students than non-developmental students. We feel it is prudent to not over interpret this finding at this juncture and believe that this finding should be replicated. That said, developmental students may be less engaged when taking the SBA, which would certainly lead to a weaker relationship between foundational skills and performance on that test.

Implications for practice

At the outset of this paper, we claimed that testing the IMH in a diverse population of college students should help gain insights into what is needed to be ready to read in college. What have we learned from this study regarding this pressing issue in contemporary educational policy? First, it is well recognized that foundational skills are important for academic

performance from early stages of literacy development (Hoover & Gough, 1990) through adolescence (OECD, 2018) and for adult readers (Sabatini, 2015). In the present study, foundational skill clearly accounted for more variance than high level inference skills. At one level this is surprising because one would expect that students entering college would be relatively highly proficient readers, and as such there would be relatively little variance explained on the traditional and scenario-based assessments by an assessment of foundational skills. At another level, the present findings are clearly consistent with prior research indicating that alarming number of students are not just ill prepared to read within their disciplines (Baer, et al., 2006; Greene & Forster, 2003; Jenkins & Boswell, 2002; NAEP, 2015), but they may struggle with proficiency in even the most basic aspects of reading. Foundational skills that support academic reading (i.e., fluency) typically account for declining variance in performance on tests of basic comprehension as students progress through grade school (e.g., Vellutino, Tunmer, Jaccard, & Chen, 2007), possibly because there is less variability in these skills as students shift from learning to read to learning from reading (Hoover & Gough, 1990). It would be optimal if the same trend reliably continued through secondary education as students become prepared for academic reading in college. However, Wang, Sabatini, O'Reilly & Weeks (2019) found that students who fell below a decoding threshold, displayed little to no growth in reading comprehension across grades 5-10. Thus, foundational skills may continue to influence reading development for certain populations beyond the 4th grade and may be a key barrier to successful performance on academic reading tasks.

Proficiency in foundational skills is an underlying source of variability in adult readers (Mellard, Fall, & Woods 2010; Sabatini, Sawaki, Shore, & Scarborough, 2010; Worthy & Viise, 1996) and struggling late adolescent readers are considered part of the adult literacy spectrum

(Greenberg, 2008). We know of no study that assesses the trends in the longitudinal relationship between foundational skills and performance on academic reading tasks, but clearly the results of the present study show that in a diverse sample of college students, there is considerable variability in foundational skills. In the current study, 58% of the participants were enrolled in a developmental literacy program, and certainly these students would be considered as struggling adult readers. Moreover, the sample included students from a broad range of backgrounds, including 23% of second language learners. As such, it may not be surprising that foundational skills were predictive of both the traditional reading comprehension assessment and the scenario-based assessment. Remediating basic skills that support reading during first college experiences is going to present serious challenges.

Clearly, foundational skills are critical for academic reading as one needs to accurately process information in order to use information (OECD, 2018). Fluent readers have automaticity and expertise in foundational skills and are able to allocate attentional and memory resources to higher-level comprehension processes, including inferences (OECD, 2018, Sabatini, 2015). In contrast, readers who have less developed foundational skills must utilize more of these resources for lower-level processes (decoding, word recognition, and sentence processing), which decreases the likelihood that they can engage in inferences that support mental model construction (e.g. Perfetti, Landi, & Oakhill, 2005; Perfetti, Marron, & Foltz, 1996). This may be one reason why some college students find it challenging to read and use course material during their first college course experiences. In addition to having fewer resources for constructing a coherent and elaborated representation of texts (OECD, 2018; Sabatini, 2015; Stafura & Perfetti, 2017), readers with lower levels of foundational skills will likely have fewer resources to devote to higher-level processes that are important for purposeful academic reading, such as goal-

directed task management processes (i.e., OECD, 2018) and reasoning with and synthesizing information from multiple sources, (Britt et al., 2018). This underscores the challenges raised by underprepared college students, and the need for further research on this population.

To the extent that disciplinary reading requires students to use appropriate background knowledge to support elaborative processing (Alexander & Jetton, 2000; Goldman, 2004; Goldman et al., 2016; Lee & Spratley, 2010), students with weak foundational skills are going to be at a particular disadvantage. Weak foundational skills may inhibit comprehension (Wang et al., 2019), and consequently the construction of new knowledge. In addition, insufficient knowledge may limit comprehension on more complex tasks (O'Reilly, Wang & Sabatini, 2019).

These results raise the question of how we should best help students who are coming to their first college experiences with underdeveloped foundational skills. There is no simple answer to this question, but the results of this study illustrate that it is a pressing issue that warrants attention. Identifying students who struggle with foundational skills is an important first step and can help target interventions. Sabatini et. al., (2014a) make the case that assessments of reading comprehension should be accompanied by measures of foundational skills to help understand poor performance. This may be particularly true with SBA's where it would be less clear whether poor performance relates to a lack of foundational skills, or a lack of higher order skills needed for more complex tasks.

We would like to conclude by stating that this study shows the potential impact that research on the basic cognitive processes that support comprehension can make in terms of understanding why students struggle or succeed when reading for college. Theories of comprehension and task-oriented reading aim to describe the aspects of the reader, and in recent cases the text and task (e.g., Britt et al., 2018) that may provide the pressure points that can lead

to successful or less successful reading experiences. Understanding these pressure points is the first step in developing effective remediations. However, we strongly suspect that those interventions should occur long before students' first exposure to college courses.

References

Ahmed, Y., Francis, D. J., York, M., Fletcher, J. M., Barnes, M., & Kulesz, P. (2016). Validation of the direct and inferential mediation (DIME) model of reading comprehension in grades 7 through 12. *Contemporary Educational Psychology*, *44*, 68-82.

Alexander, P. A., & Jetton, T. L. (2000). Learning from text: A multidimensional and developmental perspective. In M. Kamil, P. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 285–310). Mahwah, NJ: Erlbaum

Baer, J. D., Cook, A. L., & Baldi, S. (2006). *The literacy of America's college students.* American Institutes for Research. Retrieved from https://www.air.org/sites/default/files/downloads/report/The20Literacy20of20Americas20College20Students_final20report_0.pdf

Bailey, T. (2009). Challenge and opportunity: Rethinking the role and function of developmental education in community college. *New Directions for Community Colleges*, (145), 11-30. Retrieved from https://files.eric.ed.gov/fulltext/ED504242.pdf

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, *18*(1), 5-25.

Britt, M. A., Rouet, J-F., & Durik, A. M. (2018). *Literacy beyond text comprehension. A theory of purposeful reading.* NY, NY: Routledge.

Cerdán, R., Gilabert, R., & Vidal-Abarca, E. (2011). Selecting information to answer questions: Strategic individual differences when searching texts. *Learning and Individual Differences*, *21*(2), 201-205.

Cerdán, R., & Vidal-Abarca, E. (2008). The effects of tasks on integrating information from multiple documents. *Journal of Educational Psychology*, *100*(1), 209-222.

Cerdán, R., Vidal-Abarca, E., Martinez, T., Gilabert, R., & Gil, L. (2009). Impact of question-answering tasks on search processes and reading comprehension. *Learning and Instruction*, *19*(1), 13-27.

College Board. (2019). ACCUPLACER. Retrieved December 2019, from http://professionals.collegeboard.com/higher-ed/placement/accuplacerhttps://accuplacer.collegeboard.org/about/get-to-know-accuplacer

Cromley, J. G., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology*, *99*(2), 311-325.

Cromley, J. G., Snyder-Hogan, L. E., & Luciw-Dubas, U. A. (2010). Reading comprehension of scientific text: A domain-specific test of the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology, 102*(3), 687–700. doi:10.1037/a0019452

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

Feller, D.P., Magliano, J.P., O'Reilly, T., Sabatini, J., & Kopatich, R.D. (in press). Relations between component reading skills, inferences, and comprehension performance in community college readers. *Discourse Processes.*

Gilliam, S., Magliano, J. P., Millis, K. K., Levinstein, I., & Boonthum, C. (2007). Assessing the format of the presentation of text in developing a Reading Strategy Assessment Tool (R-SAT). *Behavior Research Methods*, *39*(2), 199-204.

Goldman, S.R. (2004). Cognitive aspects of constructing meaning through and across multiple texts. In N. Shuart-Faris & D. Bloome (Eds.), *Uses of intertextuality in classroom and educational research* (pp. 317–352). Greenwich, CT: Information Age

Goldman, S. R. (2011). Choosing and using multiple information sources: Some new findings and emergent issues. *Learning and Instruction*, *21*(2), 238-242.

Goldman, S. R., Braasch, J. L. G., Wiley, J., & Brodowinska, K. (2012). Comprehending and learning from internet sources: Processing patterns of better and poorer learners. *Reading Research Quarterly*, *47*(4), 356-381.

Goldman, S. R., Britt, M. A., Brown, W., Cribb, G., George, M., Greenleaf, C., Lee, C. D., Shanahan, C., & Project READI (2016). Disciplinary literacies and learning to read for understanding: A conceptual framework for disciplinary literacy. *Educational Psychologist, 51*(2), 219-246.

Goldman, S. R., & Scardamalia, M., (2013). Managing, understanding, applying, and creating knowledge in the information age: Next-generation challenges and opportunities. *Cognition and Instruction*, *31*(2), 255-269.

Gordon Commission (2013). *To assess, to teach, to learn: a vision for the future of assessment*. Princeton, NJ: Author. Retrieved from: https://www.ets.org/Media/Research/pdf/gordon_commission_technical_report.pdf

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, *101*(3), 371.

Greenberg, D. (2008). The challenges facing adult literacy programs. *Community Literacy Journal, 3*(1), 39-54.

Greene, J., & Foster, G. (2003). *Public high school graduation and college readiness rates in the United States* (Education Working Paper No. 3). New York: Manhattan Institute, Center for Civic Information. Retrieved from https://files.eric.ed.gov/fulltext/ED498138.pdf

Henningsen A. & Hamann, J. D. (2007). systemfit: A package for estimating systems of simultaneous equations in R. *Journal of Statistical Software 23*(4), 1-40.

Hoover, W. & Gough, P. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal*, *2*(2), 127-160.

Hu, L.T. & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.

Jenkins, D., & Boswell, K. (2002). *State policies on community college remedial education: Findings from a national survey*. Jenkins, D., & Boswell, K. (2002). Denver, CO: Center for Community College Policy, Education Commission of the States. Retrieved from https://files.eric.ed.gov/fulltext/ED470465.pdf

Kaakinen, J. K., & Hyönä, J. (2005). Perspective effects on expository text comprehension: Evidence from think-aloud protocols, eyetracking, and recall. *Discourse Processes, 40*(3), 239-257.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*(2), 163-182.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.

Kopatich, R. D., Magliano, J. P., Millis, K. K., Parker, C. P., & Ray, M. (2019). Understanding how language-specific and domain-general resources support comprehension. *Discourse Processes, 56*(7), 530-552.

LaRusso, M., Kim, H. Y., Selman, R., Uccelli, P., Dawson, T., Jones, S., Donovan, S. & Snow, C. (2016). Contributions of academic language, perspective taking, and complex reasoning to deep reading comprehension. *Journal of Research on Educational Effectiveness*, *9*(2), 201-222.

Lee, C. D., & Spratley, A. (2010). *Reading in the Disciplines: The challenges of adolescent literacy. Final report from Carnegie Corporation of New York's Council on Advancing Adolescent Literacy.* New York: Carnegie Corporation.

Linderholm, T., & van den Broek, P. (2002). The effects of reading purpose and working memory capacity on the processing of expository text. *Journal of Educational Psychology*, *94*(4), 778–784.

Magliano, J. P., McCrudden, M. T., Rouet, J. F., & Sabatini, J. (2017). The modern reader: Should changes to how we read affect research and theory? In M. F. Schober, M. A. Britt, & D. N. Rapp (Eds), *The Routledge handbook of discourse processes* (2nd ed., pp. 343-361). New York: Taylor & Francis.

Magliano, J. P., Millis, K., Ozuru, Y., & McNamara, D. S. (2007). A multidimensional framework to evaluate reading assessment tools. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions and technologies* (pp. 107–136). New York: Lawrence Erlbaum.

Magliano, J. P., Millis, K.K., The RSAT Development Team1, Levinstein, I., & Boonthum, C. (2011). Assessing comprehension during reading with the reading strategy assessment Tool (RSAT). *Metacognition and Learning*, *6*(2), 131-154.

Magliano, J. P., Trabasso, T., & Graesser, A. C. (1999). Strategic processing during comprehension. *Journal of Educational Psychology*, *91*(4), 615-629.

McCarthy, K., Guerrero, T., Kent, K., Allen, L., McNamara, D., Chao, S., Steinberg, J., O'Reilly, T., Sabatini, J. (2018). Comprehension in a scenario-based assessment: Domain and topic-specific background knowledge. *Discourse Processes, 55*(5-6), 510-524.

McCrudden, M. T., Magliano, J. P., & Schraw, G. (2010). Exploring how relevance instructions affect personal reading intentions, reading goals and text processing: A mixed methods study. *Contemporary Educational Psychology*, *35*(4), 229-241.

McCrudden, M. T., & Schraw, G. (2007). Relevance and goal-focusing in text processing. *Educational Psychology Review*, *19*(2), 113-139.McNamara, D. S. (2004). SERT: Self-Explanation Reading Training. *Discourse Processes*, *38*(1), 1–30.

McNamara, D. S., & Magliano, J. P. (2009). Towards a comprehensive model of comprehension. In B. Ross (Ed), *The psychology of learning and motivation* (Vol. 51, pp. 297-384). New York, NY, US: Elsevier Science.

Mellard, D., Fall, E., & Woods, K. (2010). A path analysis of reading comprehension of adults with low literacy. *Journal of Learning Disabilities, 43*(2)*,* 154-195. doi:10.1177/0022219409359345

Mislevy, R. J., & Haertel, G. (2006). Implications for evidence-centered design for educational assessment. *Educational Measurement: Issues and Practice*, *25*(4), 6-20.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, *1*(1), 3-67.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Memory and Language*, *16*(5), 519-533.

Muñoz, B., Magliano, J. P., Sheridan, R., & McNamara, D. S. (2006). Typing versus thinking aloud when reading: Implications for computer-based assessment and training tools. *Behavior Research Methods*, *38*(2), 211-217.

Narvaez, D., van den Broek, P., & Ruiz, A. B. (1999). The influence of reading purpose on inference generation and comprehension in reading. *Journal of Educational Psychology*, *91*(3), 488–496.

National Assessment of Educational Progress (NAEP) (2015). *The nation's report card*. Retrieved from https://www.nationsreportcard.gov/reading_math_g12_2015/#reading

National Governors Association Center for Best Practices & Council of Chief State School Officers (2010). *Common core state standards for English language arts*. Washington, DC: Author. Retrieved from http://www.corestandards.org/assets/CCSSI_ELA%20Standards.pdf

Oakhill, J., & Yuill, N. (1996). Reading comprehension difficulties. In Cain, K. & Oakhill, J. (Eds.), *Children's Comprehension Problems in Oral and Written*, (pp. 41-75). New York, NY: Guilford Press.

O'Reilly, T., & Sabatini, J. (2013). *Reading for understanding: How performance moderators and scenarios impact assessment design* (Research Report No. RR-13–31). Princeton, NJ: Educational Testing Service

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.), D*eep comprehension: Multi-disciplinary approaches to understanding, enhancing, and measuring comprehension*, 1st Edition (pp. 197–208). NY, NY: Routledge.

O'Reilly, T., Sabatini, J., Bruce, K., Pillarisetti, S., & McCormick, C. (2012). Middle school reading assessment: Measuring what matters under an RTI framework. *Reading Psychology Special Issue: Response to Intervention, 33*(1-2), 162-189.

O'Reilly, T., & Sheehan, K. M. (2009). *Cognitively based assessment of, for and as learning: a 21st century approach for assessing reading competency* (RM-09-04). Princeton, NJ: ETS.

O'Reilly, T., Wang, Z., & Sabatini, J. (in press). How much knowledge is too little? When knowledge becomes a barrier to comprehension. *Psychological Science*.

O'Reilly T., Weeks, J., Sabatini, J., Halderman, L., & Steinberg, J. (2014). Designing reading comprehension assessments for reading interventions: How a theoretically motivated assessment can serve as an outcome measure. *Educational Psychology Review, 26*(3), 403-424.

Organisation for Economic Co-Operation and Development (2018) PISA 2018: *Draft of the analytic framework*. https://www.oecd.org/pisa/data/PISA-2018-draft-frameworks.pdf

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

Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and instruction*, *19*(3), 228-242.

Partnership for 21st Century Skills. (2008). *21st century skills map*. Washington, DC: Author. Retrieved from https://files.eric.ed.gov/fulltext/ED519502.pdf

Pellegrino, J. N. & Chudowsky, E.A. (2001). *Knowing what students know: The science and design of educational assessment*. Washington DC, National Academy Press.

Perin, D., & Charron, K. (2006). "Lights just click on every day." In T. Bailey & V. S. Morest (Eds.), *Defending the community college equity agenda* (pp. 155–194). Baltimore: The Johns Hopkins University Press.

Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227-247). Oxford: Blackwell.

Perfetti, C. A., Marron, M. A., & Foltz, P. W. (1996). Sources of comprehension failure: Theoretical perspectives. In C. Cornoldi & J. V. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp.137-165*).* Mahwah, NJ: Erlbaum.

Perfetti, C. A., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, *18*(1), 22-37.

Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale, NJ: Lawrence Erlbaum Associates.

R Core Team (2018). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

Rouet, J.F. (2006). *The skills of document use: From text comprehension to web-based learning*. Mahwah, NJ: Erlbaum.

Rouet, J. F., & Britt, M. A. (2011). Relevance processes in multiple document comprehension. In M. T. McCrudden, J. P. Magliano, & G. Schraw (Eds.)*, Text relevance and learning from text* (pp. 19-52). Greenwich, CT: Information Age Publishing.

Rouet, J.F., Britt, M. A., & Durik, A. (in press). RESOLV: Readers' representation of reading contexts and tasks. *Educational Psychologist*.

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.

Sabatini, J. (2015). Understanding the basic reading skills of U.S. Adults: Reading components in the PIAAC Literacy Survey. Report of the ETS Center for Research on Human Capital and Education. https://www.ets.org/s/research/report/reading-skills/ets-adult-reading-skills-2015.pdf

Sabatini, J., Bruce, K., Steinberg, J., & Weeks, J. (2015). *SARA reading components tests, RISE forms: Technical adequacy and test design 2nd Edition* (ETS RR-15-32). Princeton, NJ: Educational Testing Service.

Sabatini, J. P., Halderman, L. K., O'Reilly, T., & Weeks, J. P. (2016). Assessing comprehension in kindergarten through third grade. *Topics in Language Disorders*, *36*(4), 334-355.

Sabatini, J., O'Reilly, T., & Deane, P. (2013). *Preliminary reading literacy assessment framework: Foundation and rationale for assessment and system design.* (Research Report No. RR-13–30). Princeton, NJ: Educational Testing Service.

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

Sabatini, J. P., O'Reilly, T., Halderman, L., & Bruce, K. (2014b). Broadening the scope of reading comprehension using scenario-based assessments: Preliminary findings and challenges. *L'Année Psychologique*, *114*(04), 693-723.

Sabatini, J., O'Reilly, T., Wang, Z., & Dreier, K. (2018). Scenario-based assessment of multiple source use. In J. L. G. Braasch, I. Bråten & M. T. McCrudden (Eds.), *The handbook of multiple source use* (pp. 447-465). New York: Taylor & Francis/Routledge.

Sabatini, J., O'Reilly, T., Weeks, J., & Wang, Z. (2019). Engineering a 21st Century reading comprehension assessment system utilizing scenario-based assessment techniques. *International Journal of Testing*. https://doi.org/10.1080/15305058.2018.1551224

Sabatini, J. Sawaki, Y., Shore, J., & Scarborough, H. (2010). Relationships among reading skills of adults with low literacy. *Journal of Learning Disabilities, 43*(2), 122-138. doi:10.1177/0022219409359343.

Sabatini, J., Weeks, J., O'Reilly, T. Bruce, K. Steinberg, J. & Chao, Z. (2019). SARA reading components tests, RISE forms: Technical adequacy and test design, 3rd edition. *ETS Research Report Series*. Retrieved from https://onlinelibrary.wiley.com/doi/pdf/10.1002/ets2.12269

Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review*, *78*(1), 40-59.

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

Stafura, J. Z., & Perfetti, C. A. (2017). Integrating word processing with text comprehension: Theoretical frameworks and empirical examples. In K. Cain, D. L. Compton, & R. Parrila (Eds.), *Theories of reading development* (pp. 9–32). John Benjamins Publishing Company.

Steiger, J.H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences, 42*(5), 893-98.

van den Broek, P. Lorch, R. F. Jr., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition*, *29*(8), 1081–1087.

Vellutino, F. R., Tunmer, W. E., Jaccard, J. J., & Chen, R. (2007). Components of reading ability: Multivariate evidence for a convergent skills model of reading development. *Scientific Studies of Reading, 11*(1), 3-32.

Vidal-Abarca, E., Mañá, A., & Gil, L. (2010). Individual differences for self-regulating task-oriented reading activities. *Journal of Educational Psychology, 102*, 817–826.

Vidal-Abarca, E., Salmerón, L., & Mana, A. (2011). Individual differences in task-oriented reading. In M. T. McCrudden, J. P. Magliano, & G. Schraw (Eds.)*, Text relevance and learning from text* (pp. 19-52). Greenwich, CT: Information Age Publishing.

Wang, Z., Sabatini, J., O'Reilly, T., & Weeks, J. (2019). Decoding and reading comprehension: A test of the decoding threshold hypothesis. *Journal of Educational Psychology, 111*(3), 387-401. https://psycnet.apa.org/doiLanding?doi=10.1037%2Fedu0000302

Wiley, J., & Voss, J. F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology*, *91*(2), 301.

Wolfe, M. B., & Goldman, S. R. (2005). Relations between adolescents' text processing and reasoning. *Cognition and Instruction, 23*(4), 467-502.

Worthy, J., & Viise, N. M. (1996). Morphological, phonological, and orthographic differences between the spelling of normally achieving children and basic literacy adults. *Reading and Writing,* 8(2), 139-159.