Hammill Institute on Disabilities

Journal of Learning Disabilities
2023, Vol. 56(1) 43–57

DOI: 10.1177/00222194221087387
journaloflearningdisabilities.sagepub.com
SAGE

Exploring Thresholds in the Foundational Skills for Reading and Comprehension Outcomes in the Context of Postsecondary Readers

Joseph P. Magliano, PhD[1], Amani Talwar, PhD[1*], Daniel P. Feller, MS[1], Zuowei Wang, PhD[2], Tenaha O'Reilly, PhD[2], and John Sabatini, PhD[3]

Abstract

There is a range of reasons why college students may be underprepared to read, but one possibility is that some college students are below a threshold of proficiency in the component skills of reading. The presence of thresholds means that when students fall below that threshold, their proficiency in that component skill of reading is not sufficient for there to be a relationship with comprehension performance. The present study assessed (a) whether there were thresholds in proficiencies in foundational skills, (b) whether students falling below the thresholds were disproportionately in developmental literary programs (i.e., institutionally designated as underprepared), and (c) the implications of being below the thresholds on engaging in strategic processing during reading. College students were administered assessments of foundational literacy skills, text comprehension, and strategic processing of texts. The sample included students who were enrolled in developmental literacy programs and students who were not. Thresholds were found in the foundational skills associated with word-, sentence-, and discourse-level processing. Participants below these thresholds were represented disproportionately by students determined to be underprepared for college and assigned to developmental literacy programs. Finally, students falling below the thresholds demonstrated lower reading strategy scores than students above the threshold.

Keywords

developmental education, reading, literacy skills, comprehension

Higher education faces a crisis in that a number of students do not read at a proficiency level necessary to be successful in their courses (National Assessment of Educational Progress [NAEP], 2015; National Center for Education Statistics [NCES], 2019; Perin, 2020) and ultimately fail to graduate with a degree (Bailey et al., 2010). College readiness to read certainly has an impact on the successful completion of reading literacy assignments associated with coursework. More broadly, it may also affect the development of disciplinary-specific, higher-level literacy skills important for college and career success that are often acquired through participation in college courses (Goldman et al., 2016; Shanahan & Shanahan, 2008). This problem is a particular challenge in open access institutions (i.e., institutions with no admissions criteria; consequently, all students are admitted) where a large percentage of students may need supplemental support in reading literacy (Bailey, 2009; Perin, 2020). Unfortunately, developmental literacy programs designed to support reading literacy outcomes have a poor track record of success (Bailey et al., 2016; Crisp & Delgado, 2014) and consequently are being defunded or are undergoing dramatic changes in how their support is being provided (Cormier & Bickerstaff, 2020).

We have argued that finding ways to help struggling college readers, including those referred to developmental programs, requires research directed at understanding the psychological factors that are associated with success in a variety of different literacy activities that occur in college (Feller et al., 2020; Magliano et al., 2020; Perin, 2020). We have distinguished between the foundational skills that support the act of reading (word recognition, morphological knowledge, vocabulary, and sentence processing) and

[1]Georgia State University, Atlanta, USA
[2]Educational Testing Service, Princeton, NJ, USA
[3]The University of Memphis, TN, USA

*Amani Talwar is now affiliated with American Institutes of Research, Washington, DC, USA.

Corresponding Author:
Joseph P. Magliano, Department of Learning Sciences, Georgia State University, P.O. Box 3978, Atlanta, GA 30302, USA.
Email: jmagliano@gsu.edu

higher-level inferences skills (bridging between texts constituents, elaborating based on relevant prior knowledge). Although both lower- and higher-level skills account for variance in performance on literacy tasks, it is arguably surprising how much variance in performance is accounted for by lower-level foundational skills (Ari, 2016; Halldórsdóttir et al., 2016). Ideally, by the time students reach college they should be proficient readers, which would provide the requisite skills to support the development of discipline-specific literacy skills (Shanahan & Shanahan, 2008). Although students are assumed to be proficient in foundational skills such as decoding and word recognition by fifth grade, this is often not the case (e.g., Wang, Sabatini, O'Reilly, & Weeks, 2019), and such problems could persist into college (Perin, 2020).

The present study was inspired by the Reading Systems Framework (RSF; Perfetti & Stafura, 2014), which is a theoretical framework that provides a basis for understanding the coordination of the skills that support comprehension. The RSF assumes that a lack of proficiency in the foundational skills that support reading at word (decoding, lexical access) and sentence (syntactic and semantic processing) processes has negative consequences for higher-order inference skills that support comprehension. Consistent with this perspective, Wang, Sabatini, O'Reilly, and Weeks (2019) showed that there is a threshold in decoding and word recognition proficiency for students in Grades 5 through 10 that may limit students' comprehension. Students who were above this threshold showed normal comprehension development in subsequent years, but students below the threshold did not. The present study extends the work of Wang, Sabatini, O'Reilly, and Weeks (2019) and explores whether there are thresholds in foundational skills that may limit comprehension for college students. Motivated by the RSF, we were specifically interested in assessing foundational skills of word recognition, access to word knowledge (vocabulary, morphological knowledge), and the computation of sentence-level semantics (i.e., proficiency in constructing the meaning of sentences). We also explored whether the existence of thresholds has implications on inference strategies.

Underprepared College Readers in the United States

According to the Programme for the International Assessment of Adult Competencies (PIAAC), an estimated 36.2 million millennials (47% of millennials; the most frequent age band for college students) performed at or below Level 2 in literacy (i.e., the ability to successfully read relatively long texts and engage tasks that require integrating, comparing, and contrasting information), while more than 10.4 million individuals performed below Level 1 (the ability to successfully read short texts and execute a single goal-directed action, such as literal comprehension and searching for information; Sands & Goodman, 2018). Adults in this ability range were more likely to show difficulties with component reading skills, that is, foundational skills, as measured in the PIAAC survey (Grotlüschen et al., 2016). This is consistent with previous trends of low word and passage reading fluency levels among U.S. adults at these levels (Baer et al., 2009).

There is evidence that college students also demonstrate low proficiencies in foundational reading skills (Ari, 2016; Halldórsdóttir et al., 2016; Macaruso & Shankweiler, 2010; Perin, 2020). For example, Macaruso and Shankweiler (2010) assessed proficiency in many of the components of reading skills from word processing to auditory language comprehension of students enrolled in general education courses at a community college. Although these students were in community college, their actual reading level only ranged from Grades 6 to 10. Thus, there is converging evidence of widespread reading difficulties in young adults, with additional evidence pointing to some college students requiring support for the foundational skills of reading. Research is clearly warranted that is directed at understanding the implications of challenges in foundational skill on (a) performance on comprehension tasks that are representative of college expectations and (b) higher-level comprehension processes, such as inference.

The Reading Systems Framework

The RSF provides a theoretical basis for understanding struggling college readers. Although it is recognized that reading involves word-, sentence-, and discourse-level processes, such as those measured as reading components in the PIAAC survey (Organisation for Economic Co-operation and Development [OECD], 2012, 2013), the RSF provides a framework for understanding how they are coordinated (Perfetti & Stafura, 2014). The RSF emphasizes that the quality of word and sentence processes has important implications for discourse processes. Specifically, there are direct connections between the systems that support word identification (i.e., recognition), word knowledge (i.e., vocabulary, morphology), sentence-level processes (i.e., syntactic processing, constructing accurate semantic representations of sentences), and discourse-level processes (i.e., establishing connections between sentences via inference generation, comprehension). The quality of the output of word- and sentence-level processes has implications for the quality of discourse-level processes. Perfetti and Stafura (2014) used the metaphor of "pressure points" in foundational skills to convey the possibility that the quality of higher-order processes is contingent on the quality of the output of lower-level processes.

Magliano et al. (2020) provided evidence for the interrelatedness of these processes in college readers. They measured college students' proficiency in foundational reading skills, their propensity to engage in inference processes, and their comprehension proficiency. The sample ($N = 434$) came from a university ($n = 263$) and community college ($n = 171$) and was diverse with respect to gender (58% female, 36% male, 8% no response) and race (25% White, 41% Black, 16% Latino, 12% Asian, 2% other reported races, 5% nonresponding). Approximately 58% of these students were enrolled in a developmental literacy course based on admissions criteria. The foundational skills measured included proficiencies in word recognition and decoding, morphological processes, vocabulary knowledge, and accuracy in sentence-level semantics (i.e., accurately representing the meaning of sentences). Two types of inference processes deemed important for comprehension were assessed by the Reading Strategies Assessment Tool (RSAT; Magliano et al., 2011): bridging inferences that establish how sentences in a text are semantically connected and elaborative inferences that integrate relevant world knowledge into a mental model for a text (McNamara & Magliano, 2009). Finally, the researchers also measured the proficiency of students in (a) close comprehension of a text (i.e., accurately representing the explicit content in a text and generating inferences supported by it) and (b) their ability to successfully complete literacy tasks that required complex problem-solving beyond understanding the explicit content of the texts used. Consistent with the RSF, the researchers found that the relationship between proficiency in foundational skills and performance on both comprehension tasks (i.e., close comprehension and the complex problem-solving assessments) was partially mediated by the propensity to engage in bridging and elaborative inferences. Important for the purposes of the present study, they found that the mediation effect did not differ by enrollment in a developmental program.

Thresholds in Word-Level Processes

The RSF implies that there may be proficiency thresholds in foundational skills of reading such that falling below these thresholds will have catastrophic implications for discourse-level processes. Consistent with this perspective, Wang, Sabatini, O'Reilly, and Weeks (2019) analyzed the decoding and reading comprehension performance of a sample of more than 30,000 students in Grades 5 through 10 students. They wanted to determine whether there was a threshold in decoding skills with respect to performance on the comprehension assessment. The presence of a threshold means there is a critical point in a distribution of X, such that an X–Y relationship exists above that point but not below it (or vice versa). This type of relationship can be statistically tested using a method called "broken-line regression" (Adams, 2014; Knowles & Siegmund, 1989; Siegmund & Zhang, 1994). Wang, Sabatini, O'Reilly, and Weeks (2019) found evidence of a threshold in decoding skill on comprehension performance. Although there was not a significant relationship between decoding and comprehension below the threshold, the distribution above the threshold was such that magnitude of positive correlation between decoding and comprehension was consistent with the results of other studies exploring this relationship (García & Cain, 2014). Moreover, the threshold value was found to be stable across the six grade levels, indicating that older students' comprehension could be similarly affected by poor decoding compared with younger students' comprehension.

Wang, Sabatini, O'Reilly, and Weeks (2019) conducted longitudinal modeling to examine how poor decoding affected reading comprehension development, with four waves of performance data collected over 3 years. They tracked the magnitude of reading comprehension development based on students' initial decoding threshold status. The results showed that while students above the decoding threshold improved their reading comprehension by 0.02 *SD* per year, students below the threshold showed little improvement in reading comprehension.

Importantly, the decoding threshold was identified in relatively older students, namely, students in Grades 5 through 10, an age group for which foundational reading skills are no longer mandated by the Common Core State Standards (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). Without explicit instruction on foundational reading skills, students are unlikely to catch up on their own. Indeed, in a follow-up study, Wang, Sabatini, O'Reilly, and Weeks (2019) discovered that poor decoders consistently engaged in maladaptive decoding practices that had hindered their decoding development. They spent less time trying to decode words that were unfamiliar to them, which in turn predicted slower decoding development. Thus, it is reasonable to believe the effect of poor decoding on reading comprehension may persist into adulthood, which applies to the target population of the current study.

Overview of Present Study and Research Questions

The present study used both the sample collected from Magliano et al. (2020) and an additional sample of participants; therefore, we used the same measures of word processing proficiency (word decoding/recognition, vocabulary knowledge, morphological knowledge), sentence processing proficiency, inference processing (bridging and elaborative inferences), and comprehension proficiency (close comprehension, complex comprehension) as that study. In addition to exploring thresholds in word- and sentence-level processing for the comprehension tests, the presence

of a threshold for the close comprehension test on the complex comprehension test was also explored (Pearson et al., 2020). The extent to which the presence of thresholds affected inference processes was also explored. The RSF framework would predict that participants falling below the thresholds for word- and sentence-level processes would have lower bridging and elaborations scores than those above the thresholds. Finally, we explored whether participants enrolled in developmental courses disproportionately fell below thresholds. We addressed these issues by answering the following research questions:

Research Question 1 (RQ1): What is the prevalence of inadequate word, sentence, and discourse skills among college students?

Research Question 1a (RQ1a): Are there thresholds at the word, sentence, and discourse levels?

Research Question 1b (RQ1b): Do these thresholds vary as a function of the nature of the literacy task (close comprehension, complex problem-solving)?

Research Question 2 (RQ2): What proportion of students assigned to a developmental literacy program fall below the thresholds in foundational skills?

Research Question 3 (RQ3): How are comprehension strategies (paraphrasing, bridging, elaboration) different for students below the thresholds compared with students above the thresholds?

Method

Participants

As part of a larger project, a total of 624 participants were recruited from a large, 4-year institution in the U.S. Midwest as well as from a 2-year institution in the Southcentral United States. Participants were compensated with gift cards, cash, or course points/credit, depending on the institution. From this larger sample, we retained only those participants who were native English speakers and had complete data for our measures of interest. In addition, two participants were dropped from the analyses because they were extreme outliers on one measure. This procedure resulted in a final sample of 402 participants, which included 327 students (81%) from the 4-year institution and 75 students (19%) from the 2-year institution. The sample was diverse with respect to gender (58% female, 36% male, 8% no response) and race (25% White, 41% Black, 16% Latino, 12% Asian, 2% other reported races, 5% nonresponding). The majority of participants ranged from 18 to 22 years of age (89%).

In the final sample, 244 (61%) participants were enrolled in developmental literacy courses (55 in the 2-year institution and 189 in the 4-year institution) and 158 (39%) participants were not (20 in the 2-year institution and 138 in the 4-year institution). At the 4-year institution, enrollment in a developmental reading course was based on performance on an entrance examination (i.e., Reading Comprehension Placement Instrument) and ACT/SAT scores. This course was also available as a general education course for students not enrolled in the developmental program. At the 2-year institution, a state-specific version of the Accuplacer was used to evaluate whether students needed remedial coursework. Students scoring below institutional standards were required to complete two 8-week courses prior to enrolling in introductory courses. Students were recruited from a developmental writing or reading course, but unfortunately, we were not given permission to record any institutional information (e.g., standardized scores for assessments) about the students.

Measures

All measures were computer-administered without the need of a live test administrator. Test instructions were presented on a computer and scoring was computer-based.

Study Aid and Reading Assessment (SARA). The SARA (Sabatini et al., 2015, 2019) is a web-based measure composed of subtests intended to assess foundational component reading skills. The SARA takes approximately 30 to 35 min to complete. Scores from five of the subtests were used in the current study: *Decoding/Word Recognition*, in which participants determined whether a stimulus was a word, nonword, or pseudo-homophone; *Vocabulary*, in which participants selected the appropriate synonym or topically related word to match a target word; *Morphology*, in which participants read sentences with a missing word and were asked to fill in the blank by selecting a morphologically appropriate word from three choices that have the same root but different affixes; *Sentence Processing*, in which participants read sentences with a missing word and were asked to fill in the blank by selecting the appropriate word from three choices; and *Reading Comprehension*, in which participants read short passages and answered multiple-choice questions. The Study Aid and Reading Assessment–Reading Comprehension (SARA-RC) subtest served as our measure of close comprehension. The comprehension questions required locating key ideas and details as well as making inferences across portions of a single text. Each of the subscales has been shown to have good reliability across large samples (all Cronbach's α estimates $> .80$). Prior to the current study, the assessment has been used as a screener diagnostic and progress-monitoring tool. It has also been used in the context of evaluating the effectiveness of different reading interventions (Goldman et al., 2019; Kim et al., 2017). In addition, there is evidence of concurrent validity given SARA's ability in predicting state test scores (O'Reilly et al., 2012). For instance, the SARA predicted as much

41% of the variance in English language arts state test scores for students who were classified as below proficient (O'Reilly et al., 2012). The reliability (Cronbach's α) of the five subtests calculated from the current sample was .89 (Decoding), .83 (Vocabulary), .86 (Morphology), .86 (Sentence), and .84 (Reading Comprehension).

Global Integrated Scenario-Based Assessment (GISA). The GISA (Sabatini et al., 2020) is a web-based assessment designed to measure participants' ability to engage in a complex literacy task. Each GISA form takes about 45 min to complete. The measure has been used to evaluate the effectiveness of different reading interventions because of its alignment with the cognitive literature (Goldman et al., 2019). GISA measures the extent to which participants are able to reason with and beyond texts and taps into various theoretically grounded aspects of higher-order comprehension, including the ability to evaluate, integrate, and synthesize information to reach a reading goal (see O'Reilly & Sabatini, 2013; Sabatini et al., 2014; Sabatini et al., 2020). GISA provides a reading scenario (one scenario per form) wherein students interact with various simulated agents, including a professor and classmates. Participants are given multiple thematically related texts and are asked to use them to solve a specific problem. For example, in one of the forms used in the current study, participants were given the task of updating a wiki page on Leonardo da Vinci's famous Mona Lisa painting. Participants were made aware of a debate surrounding the figure depicted in the painting and were instructed to read various historical accounts to arrive at a conclusion about what should and should not be included on the wiki page. Participants completed only one such form, or scenario, and assignment of participants to forms was counterbalanced. All texts and tasks are presented to help participants reach their final goal. The GISA provides a variety of items, including multiple-choice questions, graphic organizer, summarizing and paraphrasing, and so forth. In brief, GISA simulates an enriched academic learning context and requires students to reason with and beyond various sources to complete tasks leading to a culminating final goal.

Prior research has examined GISA's reliability and validity. Studies among elementary, middle school, and high school populations have demonstrated that GISA has good internal consistency and test–retest reliability (Cronbach's $\alpha > .80$; Sabatini et al., 2014; $r = .87$; Sabatini al., 2020, respectively). In the current study, the reliability of the forms was .83. The GISA predicts other related assessments, including perspective taking (correlations ranging from $r = .23$ to $r = .38$), complex reasoning (correlations ranging from $r = .50$ to $r = .51$), academic language (correlations ranging from $r = .67$ to $r = .68$; LaRusso et al., 2016), English language arts state test scores (correlations ranging from $r = .52$ to $r = .68$; Sabatini et al., 2014), and the Gates-MacGinitie Reading Test reading comprehension test ($r = .80$; Sabatini et al., 2020). For additional information on the psychometric properties of GISA, see Sabatini et al. (2020).

Reading Strategies Assessment Tool. The RSAT (Magliano et al., 2011) is a computer-based assessment tool that utilizes think-aloud responses to evaluate the extent to which readers engage in three types of processes: paraphrasing, bridging, and elaboration. Within RSAT, participants read texts, presented one sentence at a time, on a computer screen. At target locations throughout the text, participants are asked to "think-aloud" by typing their thoughts into a textbox. The text is not visible at the time participants are asked to think-aloud. The RSAT is not a time test and the version used in this study takes approximately 30 min to complete.

RSAT uses algorithms to score participants' written responses (i.e., verbal protocols). Using keyword matching, RSAT assesses the extent to which content words (i.e., nouns, verbs, adjectives, and adverbs) from participants' responses overlap with content words from the text. A paraphrasing score is generated by comparing the number of content words from verbal protocols with content words found in the sentence read immediately prior to the think-aloud prompt. A bridging score is generated by comparing the number of content words from verbal protocols with content words found in the prior text (but not in the sentence directly preceding the prompt). Last, an elaboration score is generated based on the number of content words from verbal protocols that do not overlap with content words found anywhere in the text. Scores from each verbal protocol are aggregated across the text to produce an average score for each text. The RSAT scores have been shown to be reliable and valid (Magliano et al., 2011). For instance, there is evidence that RSAT scores have good construct validity in that the correlation between RSAT scores and human judgments of these same processes is moderate to high ($r = .75$ for paraphrasing, $r = .74$ for bridging, $r = .48$ for elaboration; Magliano et al., 2011). RSAT scores are also correlated with other measures of comprehension, such as the ACT and the Gates-McGinitie Reading Test (rs ranging from .51 to .55; Magliano et al., 2011). Last, there is evidence suggesting that RSAT has good test–retest reliability, especially given the open-ended nature of the assessment (rs = .79 for bridging and elaboration scores; Magliano et al., 2011).

In the present study, participants read and produced constructed response to two texts, one history text ("Louis XVI and the French Revolution," 19 sentences) and one science text ("The Power of Erosion," 22 sentences). The texts were presented in a random order. Think-aloud responses were produced at six locations in the history text and seven locations in the science text. Participants completed one computer-guided practice text with two think-aloud prompts before beginning the assessment. If participants produced

Table 1. Performance by Study Participants on Measures for Determining Thresholds.

Measure	Full sample ($N = 402$)			DL students ($n = 244$)			Non-DL students ($n = 158$)			Effect size: Hedges's g
	M	*SD*	Range	*M*	*SD*	Range	*M*	*SD*	Range	
GISA	16.60	5.49	2.00–27.00	14.7	5.18	2.00–27.00	19.53	4.61	6.00–27.00	−0.97
SARA Reading Comprehension	11.99	4.62	1.00–19.00	10.55	4.61	1.00–19.00	14.22	3.68	3.00–19.00	−0.86
SARA Decoding/Word Recognition	38.59	9.09	9.00–52.00	36.12	9.31	9.00–52.00	42.40	7.25	16.00–52.00	−0.73
SARA Vocabulary	27.09	5.94	6.00–35.00	25.24	6.16	6.00–35.00	29.94	4.21	16.00–35.00	−0.86
SARA Morphology	28.81	7.63	4.00–37.00	26.91	7.98	4.00–37.00	31.75	5.97	6.00–37.00	−0.67
SARA Sentence Processing	19.8	4.96	4.00–25.00	18.6	5.24	4.00–25.00	21.65	3.82	6.00–25.00	−0.64
RSAT Bridging	1.57	0.97	0.00–7.08	1.49	0.92	0.00–7.08	1.69	1.04	0.08–6.92	−0.20
RSAT Elaboration	2.92	1.50	0.00–9.00	2.80	1.54	0.00–9.00	3.09	1.44	0.69–7.69	−0.19
RSAT Paraphrasing	1.14	0.64	0.00–4.15	1.08	0.57	0.00–3.00	1.25	0.72	0.00–4.15	−0.27

Note. DL = developmental literacy; GISA = Global Integrated Scenario-Based Assessment; SARA = Study Aid and Reading Assessment; RSAT = Reading Strategies Assessment Tool.

Table 2. Bivariate Correlations Among Measures Used for Determining Thresholds.

Measure	1	2	3	4	5	6	7	8
1. GISA	—							
2. SARA Reading Comprehension	.64***	—						
3. SARA Decoding/Word Recognition	.54***	.56***	—					
4. SARA Vocabulary	.64***	.67***	.71***	—				
5. SARA Morphology	.52***	.62***	.67***	.75***	—			
6. SARA Sentence Processing	.57***	.68***	.58***	.64***	.75***	—		
7. RSAT Paraphrasing	.20***	.26***	.31***	.29***	.22***	.23***	—	
8. RSAT Bridging	.21***	.30***	.23***	.27***	.21***	.18***	.68***	—
9. RSAT Elaboration	.28***	.31***	.26***	.30***	.24***	.19***	.05	.30***

Note. GISA = Global Integrated Scenario-Based Assessment; SARA = Study Aid and Reading Assessment; RSAT = Reading Strategies Assessment Tool.
***$p < .001$.

responses that were too short (less than five words), the following prompt was given: "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." Participants were then asked to write a longer response. Upon completing the practice text, participants read the two experimental texts.

Procedure

This study was conducted at two locations over the course of 3 years. The study consisted of two sessions. Session 1 was primarily completed in a computer lab with trained study administrators. Some participants completed Session 1 during class time while others completed it outside of class time. Session 2 was completed outside of class time by all participants. Session 2 was either administered by trained study administrators in small-group sessions or self-administered, depending on the year and location. Measures were accessed through weblinks and instructions for each measure were provided on the websites.

In Session 1, participants completed SARA followed by RSAT. This session took between 60 and 90 min to complete. In Session 2, participants completed the GISA, a demographic questionnaire, and other self-report assessments of metacognition and motivation that were not utilized in the present study. Session 2 took between 60 and 90 min to complete. The demographic survey was given last at the end of Session 2.

Data Accessibility

The data and R scripts for this study can be found on Open Science Framework (https://osf.io/5pgrc/).

Results

Students' performance on measures is summarized in Table 1. Bivariate correlations among measures are reported in Table 2.

Table 3. Thresholds for Performance on Reading Comprehension Assessments.

Reading comprehension assessment	Skill	Threshold cutoff score	Threshold *p*-value	Above threshold			Below threshold		
				n	% DL students	% non-DL students	*n*	% DL students	% non-DL students
SARA-RC	Decoding/word recognition	20.00	.016	384	60	40	13	92	8
SARA-RC	Vocabulary	14.00	.076						
SARA-RC	Morphology	14.20	.004	368	58	42	34	85	15
SARA-RC	Sentence processing	12.26	.001	352	57	43	50	86	14
GISA	Decoding/word recognition	19.00	.056						
GISA	Vocabulary	17.24	.001	369	58	42	33	94	8
GISA	Morphology	13.00	.079						
GISA	Sentence processing	12.93	.001	352	57	43	50	86	14
GISA	SARA-RC	3.89	.027	383	59	41	19	95	5

Note. For the threshold in decoding/word recognition, five students were not included in the above-threshold and below-threshold groups because their decoding/word recognition score was exactly the same as the cutoff score. DL = developmental literacy; SARA-RC = Study Aid and Reading Assessment–Reading Comprehension subtest; GISA = Global Integrated Scenario-Based Assessment.

Research Question 1

This question pertained to the prevalence of inadequate word, sentence, and discourse skills among college students and identified by the presence of thresholds in foundational skills (RQ1a) and whether these varied as a function of comprehension assessment (RQ1b). In line with the approach by Wang, Sabatini, O'Reilly, and Weeks (2019), we used broken-line regression to explore thresholds in foundational skills. Broken-line regression analyses were conducted using the lm.br package (Adams, 2017) in the R statistical environment (R Core Team, 2020). Instead of estimating one slope as in ordinary least squares regression, broken-line regression estimates two slopes and a change point that connects the two slopes. Least squares regression can be understood as a special case of broken-line regression: When the two slopes estimated in broken-line regression are the same, results of broken-line regression are the same as least square regression. The significance test of a potential broken-line relation utilizes likelihood ratio to examine whether the broken-line relation (i.e., two separate slopes) fits the empirical data better than the ordinary least squares regression (i.e., one single slope). Furthermore, broken-line regression identifies a threshold point that separates the two slopes by maximizing the model fit to the empirical data.

Separate models were estimated for each bivariate relationship between a foundational skill (decoding/word recognition, vocabulary, morphology, or sentence processing) and an academic literacy task (SARA-RC or GISA). In each model, the academic literacy task was the dependent variable and the foundational skill was the sole independent variable. In addition, a separate model was estimated to explore a threshold in SARA-RC with respect to GISA performance.

Altogether, nine broken-line regression models were estimated and six significant thresholds were identified. These results are summarized in Table 3. For academic literacy as measured by SARA-RC, there were significant thresholds in decoding/word recognition ($p = .016$), morphology ($p = .004$), and sentence processing ($p < .001$). For academic literacy measured by GISA, there were significant thresholds in vocabulary ($p < .001$), sentence processing ($p = .001$), and SARA-RC performance ($p = .027$). In all six cases, the majority of students (>87%) performed above the threshold.

Research Question 2

This question pertained to determining the proportion of students assigned to a developmental literacy program that fall below the thresholds in foundational skills. The distribution of developmental students above and below each significant threshold is reported in Table 3. Participants below the thresholds were also disproportionately students who were enrolled in developmental literacy programs. Approximately 85% to 95% of the participants who fell below the thresholds were developmental students, as compared with 57% to 60% of the participants above the thresholds designated as developmental students. For performance on SARA-RC, chi-square tests of independence confirmed that developmental students were overrepresented below the decoding/word recognition threshold, $\chi^2(1) = 4.34$, $p = .037$; the morphology threshold, $\chi^2(1) = 8.33$, $p = .004$; and the sentence processing threshold, $\chi^2(1) = 14.14$, $p < .001$. Similarly, for performance on GISA, chi-square tests of independence confirmed that developmental students were overrepresented below the vocabulary threshold, $\chi^2(1) = 15.17$, $p < .001$; the sentence processing threshold, $\chi^2(1) = 14.14$, $p < .001$; and the SARA-RC threshold, $\chi^2(1) = 8.25$, $p = .004$.

Table 4. RSAT Performance Above and Below Thresholds.

Threshold		RSAT paraphrasing		RSAT bridging		RSAT elaboration	
Reading comprehension assessment	Skill	Above threshold *M* (*SD*)	Below threshold *M* (*SD*)	Above threshold *M* (*SD*)	Below threshold *M* (*SD*)	Above threshold *M* (*SD*)	Below threshold *M* (*SD*)
SARA-RC	Decoding/word recognition	1.16 (0.64)	0.92 (0.44)	1.57 (0.94)	1.74 (1.79)	2.96 (1.51)	1.96 (1.29)
SARA-RC	Morphology	1.17 (0.65)	0.83 (0.43)	1.60 (0.99)	1.21 (0.69)	2.97 (1.52)	2.33 (1.19)
SARA-RC	Sentence processing	1.17 (0.65)	0.92 (0.46)	1.61 (0.99)	1.34 (0.80)	3.01 (1.52)	2.28 (1.26)
GISA	Vocabulary	1.17 (0.65)	0.87 (0.42)	1.58 (0.94)	1.49 (1.24)	2.99 (1.50)	2.09 (1.26)
GISA	Sentence processing	1.17 (0.65)	0.92 (0.46)	1.61 (0.99)	1.34 (0.80)	3.01 (1.52)	2.28 (1.26)
GISA	SARA-RC	1.16 (0.64)	0.73 (0.42)	1.60 (0.98)	1.04 (0.68)	2.95 (1.51)	2.19 (1.26)

Note. RSAT = Reading Strategies Assessment Tool; SARA-RC = Study Aid and Reading Assessment–Reading Comprehension subtest; GISA = Global Integrated Scenario-Based Assessment.

Research Question 3

This question pertained to assessing whether comprehension strategies (paraphrasing, bridging, elaboration) were different for students below the thresholds compared with students above the thresholds. Performance on measures of comprehension strategies above and below each threshold is reported in Table 4. We used logistic regression to explore the relationships between thresholds and comprehension strategies. For each threshold, a logistic regression model was estimated, with RSAT Paraphrasing, Bridging, and Elaboration as the independent variables in each model. The dependent variable for each model indicated whether a student was above or below a threshold; this was a dichotomous variable with possible values of 0 (below threshold) and 1 (above threshold).

The logistic regression results are summarized in Table 5. For the decoding/word recognition threshold for SARA-RC performance, the model explained approximately 14% of the variance (Nagelkerke R^2) and correctly classified 97.2% of cases, $\chi^2(3) = 12.97$, $p = .005$. All three comprehension strategies emerged as significant predictors but exhibited different relationships with the threshold outcome. Higher scores on paraphrasing and elaboration were associated with a greater likelihood of being above the threshold, whereas higher scores on bridging were associated with a lower likelihood of being above the threshold. This beta weight should be interpreted with caution as bridging scores were highly and positively correlated with paraphrasing scores ($r = .68$), and its direction could be the result of a suppression effect. For the morphology threshold for SARA-RC performance, the logistic regression model explained approximately 9% of the variance (Nagelkerke R^2) and correctly classified 91.7% of cases, $\chi^2(3) = 15.72$, $p = .001$. Significant predictors were limited to paraphrasing and elaboration, with higher scores associated with a greater likelihood of being above the threshold. For the sentence processing threshold for SARA-RC performance, the logistic regression model explained approximately 9% of the variance (Nagelkerke R^2) and correctly classified 87.7% of cases, $\chi^2(3) = 19.76$, $p < .001$. Similar to the results for the morphology threshold, the significant predictors were limited to paraphrasing and elaboration, with higher scores associated with a greater likelihood of being above the threshold. For the vocabulary threshold for GISA performance, the logistic regression model explained approximately 15% of the variance (Nagelkerke R^2) and correctly classified 92.0% of cases, $\chi^2(3) = 27.32$, $p < .001$. Similar to the results for the morphology threshold, the significant predictors were limited to paraphrasing and elaboration, with higher scores associated with a greater likelihood of being above the threshold. All three comprehension strategies emerged as significant predictors but exhibited different relationships with the threshold outcome. Higher scores on paraphrasing and elaboration were associated with a greater likelihood of being above the threshold, whereas higher scores on bridging were associated with a lower likelihood of being above the threshold. For the sentence processing threshold for GISA performance, the logistic regression model explained approximately 9% of the variance (Nagelkerke R^2) and correctly classified 87.7% of cases, $\chi^2(3) = 19.76$, $p < .001$. Significant predictors were limited to paraphrasing and elaboration, with higher scores associated with a greater likelihood of being above the threshold. These results were identical to those reported for the sentence processing threshold for SARA-RC performance because the cutoff scores for the two thresholds were extremely close (see Table 3). For the SARA-RC threshold for GISA performance, the logistic regression model explained approximately 12% of the variance (Nagelkerke R^2) and correctly classified 95.2% of cases, $\chi^2(3) = 15.42$, $p = .001$. Paraphrasing was the sole significant predictor, with higher scores associated with a greater likelihood of being above the threshold.

Table 5. Logistic Regression for Thresholds.

Predictor	*B*	*SE*	Wald's χ^2	*p*	Odds ratio	95% CI
Decoding/word recognition threshold for SARA-RC (Nagelkerke $R^2 = .136$)						
RSAT Paraphrasing	1.46	0.70	4.36	.037	4.31	[1.09, 16.98]
RSAT Bridging	−0.74	0.35	4.35	.037	0.48	[0.24, 0.96]
RSAT Elaboration	0.65	0.27	5.60	.018	1.91	[1.12, 3.27]
Morphology threshold for SARA-RC (Nagelkerke $R^2 = .089$)						
RSAT Paraphrasing	1.08	0.45	5.80	.016	2.95	[1.22, 7.12]
RSAT Bridging	−0.06	0.29	0.05	.825	0.94	[0.53, 1.66]
RSAT Elaboration	0.34	0.16	4.70	.030	1.41	[1.03, 1.91]
Sentence processing threshold for SARA-RC (Nagelkerke $R^2 = .092$)						
RSAT Paraphrasing	0.92	0.36	6.65	.010	2.51	[1.25, 5.05]
RSAT Bridging	−0.21	0.22	0.91	.340	0.81	[0.52, 1.25]
RSAT Elaboration	0.43	0.13	10.15	.001	1.53	[1.18, 2.00]
Vocabulary threshold for GISA (Nagelkerke $R^2 = .152$)						
RSAT Paraphrasing	1.57	0.47	11.26	.001	4.82	[1.92, 12.09]
RSAT Bridging	−0.71	0.29	5.92	.015	0.49	[0.28, 0.87]
RSAT Elaboration	0.62	0.18	12.25	<.001	1.87	[1.32, 2.65]
Sentence processing threshold for GISA (Nagelkerke $R^2 = .092$)						
RSAT Paraphrasing	0.92	0.36	6.65	.010	2.51	[1.25, 5.05]
RSAT Bridging	−0.21	0.22	0.91	.340	0.81	[0.52, 1.25]
RSAT Elaboration	0.43	0.13	10.15	.001	1.53	[1.18, 2.00]
SARA-RC threshold for GISA (Nagelkerke $R^2 = .119$)						
RSAT Paraphrasing	1.31	0.62	4.42	.036	3.70	[1.09, 12.51]
RSAT Bridging	0.13	0.45	0.09	.766	1.14	[0.47, 2.77]
RSAT Elaboration	0.40	0.22	3.27	.070	1.49	[0.97, 2.29]

Note. SARA-RC = Study Aid and Reading Assessment–Reading Comprehension subtest; RSAT = Reading Strategies Assessment Tool; GISA = Global Integrated Scenario-Based Assessment.

Informed by the logistic regression results, we examined the correlations between comprehension strategies and academic literacy. The RSAT Paraphrasing, Bridging, and Elaboration scores exhibited weak to moderate correlations with SARA-RC and GISA, with coefficients ranging from .20 to .31 (see Table 2). We recomputed these correlations after excluding students with inadequate foundational skills (i.e., students who scored below at least one of the six significant thresholds; $n = 73$). After this procedure, performance on SARA-RC was correlated with RSAT Paraphrasing at $r = .20$, RSAT Bridging at $r = .35$, and RSAT Elaboration at $r = .26$. Performance on GISA was correlated with RSAT Paraphrasing at $r = .14$, RSAT Bridging at $r = .23$, and RSAT Elaboration at $r = .24$. A comparison of these coefficients with those reported in Table 2 suggests that the exclusion of below-threshold students does not dramatically alter the magnitude of the associations between academic literacy and comprehension strategies. However, it is worth noting that the correlations involving bridging slightly increased in magnitude and those involving paraphrasing and elaboration slightly decreased in magnitude.

Discussion

This study was motivated by research suggesting that many students come to college with inadequate reading skills (NAEP, 2015; Perin, 2020), with some proportion of students lacking proficiencies in the foundational skills that support reading (Ari, 2016). In addition, research questions were motivated by the RSF (Perfetti & Stafura, 2014) and recent research that suggests inadequate foundational skills can be revealed by a threshold relationship to reading comprehension (Wang, Sabatini, & O'Reilly, 2019; Wang, Sabatini, O'Reilly, & Weeks, 2019).

Major Findings

Three sets of research questions were addressed. The RQ1 pertained to the presence of thresholds at word, sentence, and discourse levels of processing for the two literacy tasks (RQ1a) and whether they differed as a function of the comprehension assessments (RQ1b). For the SARA-RC test (close comprehension), we found significant thresholds in decoding/word recognition, morphological knowledge, and sentence processing. The decoding/word recognition

threshold is a direct replication of the study by Wang, Sabatini, O'Reilly, and Weeks (2019), who used an adolescent sample, suggesting that foundational reading skill problems may even persist into college. For the GISA comprehension test (complex comprehension), we found thresholds in vocabulary, sentence processing, and SARA-RC. Although the percentage of college students below these thresholds appears to be small (i.e., <13%), the effect is robust across foundational skills and comprehension tasks. Poor foundational reading skills are still a limiting factor for some college students' reading comprehension (RQ1a). These results clearly illustrate the importance of proficiency in processing at the word, sentence, and discourse (the only evidence at this level pertains to GISA) levels of language in college literacy tasks (Ari, 2016; Feller et al., 2020). The results are consistent with the assumption of the RSF that proficiencies in foundational skills of reading have important implications on discourse comprehension outcomes (Perfetti & Stafura, 2014).

Given that the bivariate correlations between foundational skills and performance on the two literacy assessments are comparable, we are hesitant to overinterpret the differences in the thresholds found at the word level across them (RQ1b). Nonetheless, it is worth considering those thresholds in terms of differences between the comprehension assessments and the literacy skills they are intended to assess. As we have discussed, SARA-RC is intended to measure proficiency in the close comprehension of a text, which involves accurately representing the content conveyed in the text and generating inferences that are important for comprehending the texts (e.g., drawing bridging inferences across sentences). The decoding/word recognition and morphological subscales reflect the basic processes that enable one to accurately read words that comprise texts and are critical for close comprehension (Feller et al., 2020; Magliano et al., 2020; Martino & Hoffman, 2002; Perfetti, 2007). In contrast, GISA is intended to reflect an authentic academic task that requires complex reasoning and problem-solving, and these extratextual processes and inferences rely relatively heavily on prior knowledge (O'Reilly et al., 2019; OECD, 2012). Vocabulary knowledge reflects the intersection between word and world knowledge (Perfetti, 2007; Perfetti & Stafura, 2014). GISA covers topics that may require relatively more prior knowledge to be successful and this may be one reason why there is a threshold in vocabulary knowledge. For instance, O'Reilly et al. (2019) found that success on the GISA was dependent upon the level of student's background knowledge. In particular, they found evidence of a knowledge threshold for GISA when using a measure of topical vocabulary as a part of the background knowledge test. By extension, it is possible that the vocabulary threshold in this study is related to, or is a proxy for, the general level of student academic knowledge. The fact that there is a threshold on the GISA for close comprehension makes sense, as the ability to problem-solve is dependent upon one's ability to first understand the gist before the information can be applied. While the results regarding the thresholds at the various levels are intriguing, they should be replicated.

RQ2 pertained to determining whether participants falling below the thresholds were disproportionately developmental students. There were indeed disproportionately more students assigned to developmental literacy programs that fell below the thresholds for both assessments. At one level, this is not surprising given that we found evidence of thresholds. Students were placed in these programs based on performance on standardized assessments of reading proficiency (e.g., Accuplacer). However, these results raise an important question as to *why* these students fall below these thresholds. The data available do not afford a definitive answer to this question. It is possible that (a) college students who fall below these thresholds have diagnosed or undiagnosed reading and learning disabilities (Deacon et al., 2012; Joshi & Bouck, 2017; Metsala et al., 2019) and (b) helping these students become successful in college would require developing compensatory mechanisms to overcome those disabilities (Taymans, 2009). It is also possible that these students have not developed effective decoding practices in their daily reading activities. For example, Wang, Sabatini, and O'Reilly (2019) found that poor decoders (i.e., below the decoding threshold) from Grades 6 through 8 tended to spend significantly less time trying to decode words that were unfamiliar to them, and the time spent decoding unfamiliar words predicted decoding growth in the subsequent years. It is possible that poor decoders in college also engage in less effective decoding practices, and if so, some decoding intervention aimed at changing their decoding habits might be helpful.

College students with inadequate foundational skills represent the higher end of a continuum of struggling adult readers, a population that includes approximately 19% of adults in the United States (NCES, 2019) and demonstrates deficits in word-level and/or higher-level competencies (Talwar et al., 2020). These basic literacy needs are typically addressed in adult education programs, which are unfortunately affected by low funding, a lack of evidence-based practices, and insufficient student progress (National Research Council, 2012). Even those adult education students who achieve high school equivalency are not necessarily prepared for postsecondary studies (Perin, 2020) and may exhibit skill levels similar to those of below-threshold students in the current sample.

RQ3 addressed assessing whether participants who fell below the thresholds had lower strategy scores. We found evidence that falling below some of the thresholds had implications for the propensity to engage in the comprehension strategies of paraphrasing, bridging, and elaborating.

Again, however, we are cautious about overinterpreting these findings. Importantly, paraphrasing, bridging, and elaboration are associated the strategy self-explaining texts during reading (McNamara, 2004), which both naturally occur during reading (Chi et al., 1998) and can be trained (McNamara et al., 2004). Moreover, bridging and elaboration are assumed to be important for comprehension across multiple theories of text comprehension (McNamara & Magliano, 2009). The positive coefficients for paraphrasing and elaboration suggest that falling above the threshold led to an increase in these processes, whereas the negative coefficients associated with bridging inferences suggest that falling below the threshold led to a decrease in bridging inferences. Paraphrasing and elaboration require deliberate processes (Magliano et al., 1999), and it is likely that college students need to be proficient readers to devote resources to them. The finding of negative coefficients with respect to bridging inferences is consistent with an assumption of the RSF that bridging processes are heavily supported by foundational skills, in particular those operating at the word level (Perfetti & Stafura, 2014).

Implications for Practice, Research, and Theory

The results of this study are consistent with prior research indicating that a subset of developmental students does not have sufficient foundational reading skills upon entering college (Ari, 2016; Halldórsdóttir et al., 2016). It is no wonder that postsecondary institutions have spent an enormous amount of money to support students who struggle as readers (e.g., Crisp & Delgado, 2014). Many of the extant developmental programs have not been broadly successful in helping students progress into credit-bearing courses or increasing retention or graduation rates (Bailey, 2009; Jaggars & Stacey, 2014). On the contrary, advocates of these programs argue for their necessity and the need for continued research (Boylan & Trawick, 2015). The results of this study indicate the need for more research on how to support underprepared college students who are below the threshold in some foundational skills to develop the higher-level skills that are necessary to complete the literacy tasks required in college courses.

Some struggling college readers may have diagnosed (and underreported) or undiagnosed reading and learning disabilities, as do many students who have challenges with foundational skills (Chevalier et al., 2017; Sparks & Lovett, 2009; Swanson & Hsieh, 2009). Because reporting of disabilities is optional for adults attending college, it is difficult to accurately assess the extent that college students or participants in this sample have learning or reading disorders. However, it is quite possible that many of the participants who fell below the thresholds have systemic reading disabilities. This study underscores the importance of supporting students with reading disabilities, as it is very likely that many may fall below the threshold on foundational skills, which will have negative implications for the extent that they engage in the higher-order literacy processes and strategies that are essential for the range of literacy tasks they will encounter in college.

This study was conducted in 4- and 2-year open-access institutions. The sample sizes at each location do not afford an exploration of whether the presence and cutoff points for thresholds are different across institutions. However, it is important to explore the extent that the thresholds are stable across institutions, given there are considerable differences in the admission criteria for 4- and 2-year institutions. It is likely the case that institutions need to tailor their support for the literacy needs of first-year students. Current trends in student support involve allowing students to take credit-bearing courses and to provide support concurrently. The present study does not yield insights into the nature of that support. Instead, it indicates that there are likely students below thresholds in some of the foundational skills of reading and that this has implications on the inference processes that support comprehension. More research is needed to devise ways to support low-skilled college readers and we suspect that personalized, technology-based support will be part of a strategy to help these students be successful (e.g., Graesser et al., 2020).

Finally, the current study illustrated the utility of the RSF as a theoretical framework for guiding research on struggling college readers. It is important to note that the RSF is not a formal theory and was proposed to help motivate research questions regarding the relationships between the component skills of reading (Perfetti & Stafura, 2014). Perfetti and Stafura argued that there are "pressure points" in some of these systems such that the quality of the output of earlier processes affects the quality of later processes, which is consistent with the existence of thresholds. Finding ways to support students who fall below these thresholds is no simple task, but demonstrating their presence is the first step in the process.

Limitations

There are several limitations of the present study that must be acknowledged. First, the sample is relatively small for studies of this nature. As such, this certainly has implications for the extent that the results can be generalized beyond this sample. The current study illustrates the need to conduct future studies with a large sample that affords exploring the extent that these thresholds are stable or different across 2- and 4-year institutions, as well as institutions that service students in a variety of socioeconomic contexts. Second, and related, given the study's exploratory nature, we chose not to correct for alpha for multiple

comparisons (albeit four of the thresholds had an α less than .005). Third, a decision was made in this study exploring thresholds to restrict the sample to native English speakers. However, future studies should include a more diverse sample of students, such as second-language learners. Many community college districts provide postsecondary educational opportunities to a relatively high proportion of students who are English language learners but who also vary in proficiency in English (Bergey et al., 2018). As such, it is important to understand the nature of thresholds in postsecondary English language learners. In addition, while the present study focused on foundational skills of reading, there are other factors, such as working memory and prior knowledge, that could be explored. In fact, O'Reilly et al. (2019) found evidence of thresholds in prior knowledge on performance on scenario-based assessment in high school students and the existence of such thresholds likely persists into college.

It is important to acknowledge some limitations of RSAT. Although RSAT allows the use of verbal protocols at a scale that would be challenging when relying on hand coding to identify comprehension processes, the algorithms are imperfect. They give a proxy of the propensity to engage in bridging and elaborative processes, but not directly the quality of those processes. Moreover, the measure of elaborative inferences is considerably worse than bridging inferences. The computer-based detection of elaborative inferences is a challenge, in part given their idiosyncratic nature (Millis et al., 2007). Nonetheless, RSAT has been shown to be predictive of a variety of comprehension outcomes (Magliano et al., 2011, 2020) and successfully used to show differences in development and nondevelopmental readers that were confirmed with hand coding (Feller et al., 2020). All this said, more research is needed to improve the classification algorithms of RSAT.

Conclusion

This study represents a first step in a larger research agenda to help uncover reasons why some college students may struggle with the literacy demands of their coursework. This study adds to the literature by demonstrating that there may be thresholds in foundational reading skills that can limit adults' comprehension on both close and applied literacy tasks (e.g., Wang, Sabatini, O'Reilly, & Weeks, 2019). Although the incidence of falling below these thresholds was related to enrollment in developmental literacy programs, students' use of key reading strategies also plays a role. While we cannot point directly to the presence or prevalence of specific learning or reading disabilities in this sample, national statistics warrant further exploration of whether and how much of this population may face specific reading/learning challenges beyond a history of inadequate reading instructional experiences. With a more complete understanding of the causes of comprehension failure, instructors are in a better position to address student weaknesses with appropriate interventions. In addition, we believe that the next step in this research is to (a) replicate this study on a larger scale and across multiple institutions and (b) assess the longitudinal implications of falling above and below the thresholds over the early college years with respect to college performance and retention. Such research should consider assessing growth in foundational skills and strategy use, as well as other factors, such as motivation to persist in courses. A study of this nature will provide insights into how to best support readers who vary in skills during their early college experience.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

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, U.S. Department of Education (R305A150193 and R305A190063). All opinions expressed in this manuscript are those of the authors and not the funding agencies.

ORCID iDs

Amani Talwar https://orcid.org/0000-0003-0338-5906
Zuowei Wang https://orcid.org/0000-0002-9832-6193

References

Adams, M. (2014). *lm.br: An R package for broken line regression*. https://cran.pau.edu.tr/web/packages/lm.br/vignettes/lm.br.pdf

Adams, M. (2017). *lm.br: Linear model with breakpoint* (R package version 2.9.3) [Computer software]. https://CRAN.R-project.org/package=lm.br

Ari, O. (2016). Word recognition processes in college-age students' reading comprehension achievement. *Community College Journal of Research and Practice*, *40*(8), 718–723. https://doi.org/10.1080/10668926.2015.1098578

Baer, J., Kutner, M., & Sabatini, J. P. (2009). *Basic reading skills and the literacy of the America's least literate adults: Results from the 2003 national assessment of adult literacy (NAAL) supplemental studies* (NCES 2006-488; Issue NCES 2006-488). National Center for Education Statistics, Institute for Education Sciences, U.S. Department of Education. https://nces.ed.gov/pubs2009/2009481.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. https://doi.org/10.1002/cc.352

Bailey, T., Jeong, D. W., & Cho, S. W. (2010). Referral, enrollment, and completion in developmental education sequences in

community colleges. *Economics of Education Review*, *29*(2), 255–270. https://doi.org/10.1016/j.econedurev.2009.09.002

Bailey, T. R., Jaggars, S. S., & Jenkins, D. (2016). *Redesigning America's community colleges: A clearer path to student success*. Harvard University Press.

Bergey, R., Movit, M., Baird, A. S., & Faria, A. M. (2018). *Serving English language learners in higher education: Unlocking the potential*. American Institutes for Research. https://files.eric.ed.gov/fulltext/ED585543.pdf

Boylan, H. R., & Trawick, A. R. (2015). Contemporary developmental education: Maybe it's not as bad as it looks. *Research & Teaching in Developmental Education*, *31*(2), 26. http://www.jstor.org/stable/45373108

Chevalier, T. M., Parrila, R., Ritchie, K. C., & Deacon, S. H. (2017). The role of metacognitive reading strategies, metacognitive study and learning strategies, and behavioral study and learning strategies in predicting academic success in students with and without a history of reading difficulties. *Journal of Learning Disabilities*, *50*(1), 34–48. https://doi.org/10.1177/0022219415588850

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, R., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*(2), 145–182. https://doi.org/10.1016/0364-0213(89)90002-5

Cormier, M., & Bickerstaff, S. (2020). Research on developmental education instruction for adult literacy learners. In D. Perin (Ed.), *The Wiley handbook of adult literacy* (pp. 541–561). Wiley-Blackwell.

Crisp, G., & Delgado, C. (2014). The impact of developmental education on community college persistence and vertical transfer. *Community College Review*, *42*(2), 99–117. https://doi.org/10.1177/0091552113516488

Deacon, S. H., Cook, K., & Parrila, R. (2012). Identifying high-functioning dyslexics: Is self- report of early reading problems enough? *Annals of Dyslexia*, *62*(2), 120–134. https://doi.org/10.1007/s11881-012-0068-2

Feller, D. P., Magliano, J. P., Sabatini, J., O'Reilly, T., & Kopatich, R. D. (2020). Relations between component reading skills, inferences, and comprehension performance in community college readers. *Discourse Processes*, *57*(5-6), 473–490. https://doi.org/10.1080/0163853X.2020.1759175

García, J. R., & Cain, K. (2014). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of Educational Research*, *84*(1), 74–111. https://doi.org/10.3102/0034654313499616

Goldman, S. R., Britt, M. A., Brown, W., Cribb, G., George, M., Greenleaf, C. D. L., 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. https://doi.org/10.1080/00461520.2016.1168741

Goldman, S. R., Greenleaf, C., Yukhymenko-Lescroart, M., Brown, W., Ko, M., Emig, J., George, M., Wallace, P., Blum, D., Britt, M. A., & Project READI. (2019). Explanatory modeling in science through text-based investigation: Testing the efficacy of the project READI intervention approach. *American Educational Research Journal*, *56*(4), 1148–1216. https://doi.org/10.3102/0002831219831041

Graesser, A. C., Greenberg, D., Olney, A., & Lovett, M. W. (2020). Educational technologies that support reading comprehension for adults who have low literacy skills. In D. Perin (Ed.), *The Wiley handbook of adult literacy* (pp. 471–493). Wiley-Blackwell.

Grotlüschen, A., Mallows, D., Reder, S., & Sabatini, J. (2016). *Adults with low proficiency literacy or numeracy. OECD Publishing*. http://dx.doi.org/10.1787/5jm0v44bnmnx-en

Halldórsdóttir, L. Y., Valdimarsdottir, H., & Arntzen, E. (2016). Effects of a multicomponent intervention to improve reading fluency in adult students. *Behavioral Interventions*, *32*(2), 3–20. https://doi.org/10.1002/bin.1452

Jaggars, S. S., & Stacey, G. W. (2014). *What we know about developmental education outcomes. research overview*. Community College Research Center. http://ccrc.tc.columbia.edu/media/k2/attachments/what-we-know-about-developmental-education-outcomes.pdf

Joshi, G. S., & Bouck, E. C. (2017). Examining postsecondary education predictors and participation for students with learning disabilities. *Journal of Learning Disabilities*, *50*(1), 3–13. https://doi.org/10.1177/0022219415572894

Kim, J. S., Hemphill, L., Troyer, M. T., Thomson, J. M., Jones, S. M., LaRusso, M., & Donovan, S. (2017). Engaging struggling adolescent readers to improve reading skills. *Reading Research Quarterly*, *52*(3), 357–382. https://doi.org/0.1002/rrq.171

Knowles, M., & Siegmund, D. (1989). On Hotelling's approach to testing for a nonlinear parameter in regression. *International Statistical Review/Revue Internationale de Statistique*, 205–220. https://doi.org/10.2307/1403794

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. https://doi.org/10.1080/19345747.2015.1116035

Macaruso, P., & Shankweiler, D. (2010). Expanding the simple view of reading in accounting for reading skills in community college students. *Reading Psychology*, *31*(5), 454–471. https://doi.org/10.1080/02702710903241363

Magliano, J. P., Higgs, K., Santuzzi, A., Tonks, S. M., O'Reilly, T., Sabatini, J., Feller, D. P., Kopatich, R. J., Ray, M., & Parker, C. (2020). Testing the inference mediation hypothesis in a postsecondary context. *Contemporary Educational Psychology*, *61*. https://doi.org/10.1016/j.cedpsych.2020.101863

Magliano, J. P., & Millis, K.K., The RSAT Development Team, Levinstein, I., & Boonthum, C.(2011). Assessing comprehension during reading with the reading strategy assessment Tool (RSAT). *Metacognition and Learning*, *6*, 131–154. https://doi.org/10.1007/s11409-010-9064-2

Magliano, J. P., Trabasso, T., & Graesser, A. C. (1999). Strategic processes during comprehension. *Journal of Educational Psychology*, *91*(4), 615–629. https://doi.org/10.1037/0022-0663.91.4.615

Martino, N. L., & Hoffman, P. R. (2002). An investigation of reading and language abilities of college freshmen. *Journal*

of Research in Reading, *25*(3), 310–318. https://doi.org/10.1111/1467-9817.00178

McNamara, D. S. (2004). SERT: Self-explanation reading training. *Discourses Processes*, *38*(1), 1–30. https://doi.org/10.1207/s15326950dp3801_1

McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iSTART: Interactive strategy trainer for active reading and thinking. *Behavioral Research Methods, Instruments, & Computers*, *36*, 222–233. https://doi.org/10.3758/BF03195567

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). Elsevier Science.

Metsala, J. L., Parrila, R., Conrad, N. J., & Deacon, S. H. (2019). Morphological awareness and reading achievement in university students. *Applied Psycholinguistics*, *40*(3), 743–763. https://doi.org/10.1017/S0142716418000826

Millis, K. K., Magliano, J. P., Todaro, S., & McNamara, D. S. (2007). Assessing and improving comprehension with latent semantic analysis. In T. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *The handbook of latent semantic analysis* (pp. 207–225). Lawrence Erlbaum.

National Assessment of Educational Progress. (2015). *National Assessment of Educational Progress (NAEP): 2015 mathematics assessment*. Institute of Education Sciences. https://www.nationsreportcard.gov/reading_math_2015/#?grade=4

National Center for Education Statistics. (2019). *Highlights of the 2017 U.S. PIAAC results web report* (NCES 2020-777). U.S. Department of Education. https://nces.ed.gov/surveys/piaac/current_results.asp

National Governors Association Center for Best Practices, & Council of Chief State School Officers. (2010). *Common Core State Standards Initiative: About the standards*. http://www.corestandards.org/about-the-standards/

National Research Council. (2012). *Improving adult literacy instruction: Options for practice and research*. The National Academies Press. https://doi.org/10.17226/13242

O'Reilly, T., & Sabatini, J. (2013). *Reading for understanding: How performance moderators and scenarios impact assessment design* (ETS Research Report Series RR-15-17). ETS. https://doi.org/10.1002/j.2333-8504.2013.tb02338.x

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. https://doi.org/10.1080/02702711.2012.631865

O'Reilly, T., Wang, Z., & Sabatini, J. (2019). How much knowledge is too little? When a lack of knowledge becomes a barrier to comprehension. *Psychological Science*, *30*(9), 1344–1351. https://doi.org/10.1177%2F0956797619862276

Organisation for Economic Co-operation and Development. (2012). *Literacy, numeracy and problem solving in technology-rich environments: Framework for the OECD survey of adult skills*. OECD Publishing. http://dx.doi.org/10.1787/9789264128859-en

Organisation for Economic Co-operation and Development. (2013). *OECD skills outlook 2013: First results from the survey of adult skills*. OECD Publishing. http://dx.doi.org/10.1787/9789264204256-en

Pearson, P. D., Palincsar, A. S., Biancarosa, G., & Berman, A. I. (2020). *Reaping the rewards of the reading for understanding initiative*. National Academy of Education. https://naeducation.org/wp-content/uploads/2020/06/Reaping-the-Rewards-of-the-Reading-for-Understanding-Initiative.pdf

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, *11*(4), 357–383. https://doi.org/10.1080/10888430701530730

Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, *18*(1), 22–37. https://doi.org/10.1080/10888438.2013.827687

Perin, D. (2020). Reading, writing, and self-efficacy of low-skilled postsecondary students. In D. Perin (Ed.), *The Wiley handbook of adult literacy* (pp. 237–260). Wiley-Blackwell.

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/

Sabatini, J., Bruce, K., Steinberg, J., & Weeks, J. (2015). *SARA reading components tests, RISE Forms: Technical adequacy and test design* (ETS RR-15-32). Educational Testing Service. https://origin-www.ets.org/Media/Research/pdf/RR-13-08.pdf

Sabatini, J., O'Reilly, T., Weeks, J., & Wang, Z. (2020). 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., 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). https://doi.org/10.1002/ets2.12269

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

Sands, A., & Goodman, M. (2018). *Too big to fail: Millennials on the margins*. ETS Center for Research on Human Capital and Education. https://www.ets.org/s/research/report/opportunity-too-big-to-fail.pdf

Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review*, *78*(1), 40–59. https://doi.org/10.17763/haer.78.1.v62444321p602101

Siegmund, D. O., & Zhang, H. (1994). Confidence regions in broken line regression. *Lecture Notes-Monograph Series*, 292–231. https://www.jstor.org/stable/i397505

Sparks, R. L., & Lovett, B. J. (2009). College students with learning disability diagnoses: Who are they and how do they perform? *Journal of Learning Disabilities*, *42*(6), 494–510. https://doi.org/10.1177/0022219409338746

Swanson, H. L., & Hsieh, C. J. (2009). Reading disabilities in adults: A selective meta-analysis of the literature. *Review of Educational Research*, *79*(4), 1362–1390. https://doi.org/10.3102/0034654309350931

Talwar, A., Greenberg, D., & Li, H. (2020). Identifying profiles of struggling adult readers: Relative strengths and weaknesses in lower-level and higher-level skills. *Reading and Writing*, *33*, 2155–2171. https://doi.org/10.1007/s11145-020-10038-0

Taymans, J. M. (Ed.). (2009). *Learning to achieve: A review of research on serving adults with learning disabilities*. National Institute of Literacy. http://lincs.ed.gov/publications/pdf/L2ALiteratureReview09.pdf

Wang, Z., Sabatini, J., & O'Reilly, T. (2019). When slower is faster: Time spent decoding novel words predicts better decoding and faster growth. *Scientific Studies of Reading*, *24*, 397–410. https://doi.org/10.1080/10888438.2019.1696347

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://doi.org/10.1037/edu0000302