How Much Knowledge is too Little?

When a Lack of Knowledge Becomes a Barrier to Comprehension

Citation:

O'Reilly, T., Wang, Z., & Sabatini, J. (2019). How much knowledge is too little? When knowledge becomes a barrier to comprehension. *Psychological Science*, 1-8.
https://journals.sagepub.com/doi/10.1177/0956797619862276

Published and peer reviewed by: *Psychological Science*, the flagship journal of the Association for Psychological Science, is the leading peer-reviewed journal publishing empirical research spanning the entire spectrum of the science of psychology.
https://journals.sagepub.com/home/pss

Abstract

Have you ever found it difficult to read something due to your lack of knowledge on the topic? We investigated this phenomenon with a sample of 3,534 high school students who took a background knowledge test before working on reading comprehension tasks on the topic of ecology. Broken-line regression revealed a knowledge threshold such that below the threshold the relationship between comprehension and knowledge was weak (β =.18), but above the threshold a strong and positive relation emerged (β =.81). Further analyses indicated that certain topically relevant words (e.g. ecosystem, habitat) were more important to know than others when predicting the threshold, and these key words could be identified using natural language processing techniques. Collectively, these results may help identify who is likely to have a problem comprehending information on a specific topic, and to some extent, what knowledge is likely required to comprehend information on that topic.

Key words: background knowledge, reading comprehension, knowledge threshold hypothesis, broken-line regression, content area reading

Acknowledgement: The research reported here was supported by the U.S. Department of Education Institute of Education Sciences, Award No. R305A150176 & R305F100005, to Educational Testing Service. The opinions expressed are those of the authors and do not represent views of the U.S. Department of Education or Educational Testing Service.

How Much Knowledge is too Little? When a Lack of Knowledge Becomes a Barrier to Comprehension

Although research has shown that background knowledge can facilitate reading comprehension (e.g. Ahmed et al., 2016; Elbro & Buch-Iversen, 2013; Gough, Hoover, & Peterson, 1996; Ozuru, Dempsey, & McNamara, 2009; Shapiro, 2004), much less is known about precisely how much knowledge is necessary to understand a text, and whether there is a specific amount of knowledge required before understanding is compromised. In this study, we explore whether we can quantitatively identify a point below which a lack of students' background knowledge impedes their understanding, and above which background knowledge starts to facilitate comprehension. We call this point the *knowledge threshold*. We also explore if it is possible to predict whether students fall below or above the threshold from their basic knowledge of topically-relevant keywords, many of which do not appear in the texts they read. If successful, this approach would be useful for helping identify who may have difficulty understanding a text on a particular topic.

Why Knowledge Matters? Theoretical Perspectives

Background knowledge is critical in many models of reading (Cromley & Azevedo, 2007). In the Construction Integration Model (Kintsch, 2004), knowledge is necessary to form a *situation model*, the reader's interpretation of the text, by integrating background knowledge and the text contents. Activation models such as the resonance model (Myers & O'Brien, 1998) highlight the importance of background knowledge for reading comprehension. The text not only activates words and concepts in the readers' mind that are directly mentioned in the text, but also words and concepts that are not directly mentioned, but highly relevant to the concepts in the text. The activation of knowledge can have immediate impact on comprehension, especially when the text is closely related to what one already knows (Cook & O'Brien, 2014). When a reader knows more about a topic, reading texts on the topic would result in more activation of related knowledge (or knowledge schema) compared to a reader who knows less, and this contributes to differences in comprehension between high and low knowledge readers through mechanisms such as inference making (McNamara & O'Reilly, 2009).

The Existence of Thresholds in the Literature

Researchers have suggested that readers need a minimal level of decoding and word recognition skills (Duncan et al., 2013; Juul, Poulsen, & Elbro, 2014; Wang, Sabatini, O'Reilly, & Weeks, 2018) or general vocabulary (Hsueh-Chao & Nation, 2000; Laufer, 1989; Schmitt, Jiang, & Grabe, 2011) to read and understand text at a sufficient level. It is estimated that people need to know the meaning of between 95-98% of the words in text in order to comprehend it well (Hsueh-Chao & Nation, 2000; Laufer, 1989; Schmitt et al., 2011). Research has shown that English language learners need to have a vocabulary size of over 3000 in order to achieve acceptable performance on a comprehension test (Laufer, 1992). The consequences of falling below a threshold are striking. Fifth through tenth grade students who fell below a decoding threshold showed little growth (< .05SD per year) compared to peers who were above the decoding threshold (about 0.2SD per year; Wang et al., 2018).

It is important to note that the above studies focused on exploring whether there are specific decoding and vocabulary thresholds for reading comprehension. A lack of knowledge, as reflected by not knowing important keywords on a topic, will likely create difficulties in understanding topical texts similar to those caused by a lack of general vocabulary, thus resulting in a threshold that is domain specific, that is, the knowledge threshold.

The idea of a knowledge threshold is implied in the stage Model of Domain Learning (Alexander, 1997, 2003; Alexander, Jetton, & Kulikowich, 1995), which provides a framework to understand how people gain expertise in a domain as they accumulate domain knowledge. On the path to achieve expertise, the learner needs to construct a knowledge framework that provides the scaffold for subsequent learning. In doing so, they move from the acclimating stage to the competent stage (Alexander, 1997). Specifically, their knowledge transitions from fragmented to well-structured, which helps the learner become skilled in attending to important information for more efficient learning. This transition implies that the role of background knowledge undergoes a qualitative change as students develop knowledge in a domain. Because a large part of domain learning relies on reading and comprehending new materials, it also implies that the relation between background knowledge and reading comprehension might change during this transition.

Measuring Knowledge with a Topical Vocabulary Task

While previous studies have mostly used factual statements in the form of multiple-choice questions to evaluate background knowledge, the development of these items often requires expertise in the topic and is time-consuming. For example, Cromley and Azevedo (2007) had to read the passages in a reading comprehension test and identify the background knowledge that could be important for students to know in order to understand the content. They then developed items around the identified background knowledge. Finally, they also developed distractor items. Even after all these procedures, not all items turned out to be usable because some items were too easy or too difficult. These complications limit the assessment of background knowledge.

In this paper, we propose a different method to measure background knowledge. According to the resonance model, one’s knowledge on a topic can be evaluated by examining one’s knowledge activation when introduced to a topic. To elicit knowledge activation, we present students with a list of words--some related to the topic and others unrelated--and have them decide whether each word is related to the topic to be read. The selection of topically related key words was achieved through using a natural language processing database, which was generated by calculating co-occurrences of words in a corpus of over one billion words of natural texts (Deane, 2012). The database provides a topical association index for each keyword in a given topic. Words that occur more often in a topic generally have higher topical association index. We also selected as distractors a similar number of topically irrelevant words that matched in general word frequency (in the overall English language) of the topically relevant words. Consistent with the resonance model, which posits a connection between knowledge activation and comprehension (Myers & O'Brien, 1998), we predict that students' performance in such a keyword recognition task would be correlated to comprehension.

In short, we ask two research questions: first, whether we could identify a knowledge threshold below which comprehension is limited and not predicted by knowledge, and above which the two constructs are correlated; second, whether knowledge of a few topically-related keywords identified through natural language processing could be used to identify who is below the knowledge threshold.

Method

Participants

Participants were 3,534 grade 9-12 students from 37 schools in two states in the West and Midwest of the United States. The data was collected as part of a separate, multi-school study conducted by colleagues of ours in another organization. Our organization was responsible for the design, administration, scoring, and psychometric analyses of the measures. Consequently, the sample size was determined by the needs of this allied study. Due to agreements with the schools, we do not have individual demographic information available. However, we were able to obtain demographic information for the whole recruitment pool from which our participants were drawn. For the whole recruitment pool of 14,747 students, 49% were female; 14% were English language learners; 56% were eligible for free or reduced-price lunch and 61% were nonwhite students.

Materials

Comprehension was measured with a scenario-based assessment on the topic of ecosystems (O'Reilly, Weeks, Sabatini, Halderman, & Steinberg, 2014). The reading comprehension test had 34 comprehension items (34 total points). Reliability as reflected by Cronbach's alpha calculated from the current sample was .88. Items measured single text understanding such as the ability to recognize and provide accurate paraphrases, the ability to summarize text, and the ability to recognize opinions and incorrect information. The reading comprehension test also contained items that measured students' ability to apply what they read across multiple texts and reason about scientific content. This included items that required students to interpret data, apply classifications to given scientific abstracts, and apply scientific definitions to given vignettes. Thus, the measure included some traditional style reading items, as well as items that measured students' ability to reason and apply the information they read. The length of the two primary content passages in the reading comprehension test were 814 and 304 words, with Flesch Kincaid (Kincaid, Fishburne Jr, Rogers, & Chissom, 1975) grade levels of 9.8 and 15.4 respectively. The respective text complexity grade level estimates using the TextEvaluator system (Sheehan, Kostin, Napolitano, & Flor, 2014) were 10 and 12 which was within the grade range of this sample.

Before students worked on the reading comprehension tasks, they were also given a background knowledge measure with two types of items. The first was a topical vocabulary task (44 items). Students saw a list of keywords and they were asked to indicate whether each keyword was either related or unrelated to the topic of ecology. Only 9 of the 26 topical words in the topical vocabulary task were explicitly mentioned in the texts. A topical association index was obtained for these keywords from the natural language processing database provided by Deane (2012). The second type of item tested students' factual knowledge related to the topic *ecosystems*, in the form of a multiple-choice test (13 items). The analyses below were performed with both topical vocabulary and factual multiple-choice items as the background knowledge measure. Using topical vocabulary items alone as the background knowledge measure yielded similar results. For both item types, students were told that their performance on these items would not count towards their final score, and that they were allowed to select an "I don't know" option, if they decided that they did not know the answer to that question. For the purposes of this paper, the "I don't know" option was scored as incorrect. Reliability (Chronbach's α) of background knowledge items was .91.

Because the topical vocabulary task we used is a vocabulary measure, questions arise regarding whether the knowledge threshold identified with this measure is actually a result of a general vocabulary threshold as discovered by Laufer (1989) as opposed to a threshold that is realted to the topic of the texts. To deal with this, a small subset of all the students (n=303) also completed a history vocabulary test that included 44 items. In the test, students saw lists of words and they needed to indicate if each word was related to the topic of the history of U.S. immigration in the 19th and early 20th centuries. Thus, the format of this history vocabulary test was exactly the same as the topical vocabulary test used on the topic of ecology. Reliability of these history vocabulary items calculated from the 303 students was .88.

Procedure

Students took the reading comprehension test along with the background knowledge section online during their regular 55-minute class period. Selected response items were automatically scored, and constructed responses were manually scored by two trained raters following a rubric developed for these items. When the two raters disagreed in their initial scoring, they discussed to reach an agreement before providing a final score. There were four constructed response items, including three summary items and one paraphrase item. Each of the three summary responses was scored on a 0-3 scale. The paraphrase response was given a binary score of 0 or 1. To evaluate inter-rater agreement, 300 student responses on each of the four constructed response items were independently scored by the two raters. For the summary items, 72% of the 300 responses were given the exact score by both raters, and on another 21% of the responses the two raters only had 1 point score difference, thus adjacent agreement was 93%. For the paraphrase item, the two raters provided the exact score on 86% of all responses. For the purpose of final score calculation, each summary score was rescaled to a one-point scale by dividing the manual score by three, so that each item in the reading comprehension test task was worth one point.

Analysis

To answer our first research question about the identification of a knowledge threshold, we used broken-line regression (Adams, 2014; Muggeo, 2008). Broken-line regression is a statistical method that identifies a changepoint in linear regression and it provides significance level and confidence interval for the changepoint (i.e. threshold). Instead of estimating one regression slope as in linear regression, broken-line regression estimates two regression slopes, divided by the identified changepoint. This method has recently been used in educational research (Wang et al., 2018) and could be useful for making future binary decisions (e.g., teach background knowledge before reading or not).

To answer our second research question regarding whether students' knowledge threshold status could be determiend by their recognition of topical keywords, we selected the six keywords that had the highest natural language processing topical association index and used performance on these six keywords to predict students' knowledge threshold status using logistic regression. This was to show the specificity of the topical knowledge and the utility of using natural language processing topical assocation index to select keywords to test students' knowledge threshold status.

Results

Students' background knowledge and reading comprehension were scored separately. Mean score on the background knowledge questions was 38, *SD*=10, range [0, 54]. Mean score on the reading comprehension section was 15, *SD*=7, range [1.33, 34]. To improve the interpretability of results, these scores were also transformed to Z-score (mean=0, SD=1) before we performed broken-line regression.

Broken-line regression (Adams, 2014; Muggeo, 2008) confirmed that the relation between background knowledge and reading comprehension was affected by a knowledge threshold at a background knowledge score (standardized scores in parentheses) of 33.5 (-.40), $p<.01$, with 95% confidence interval [29, 36] ([-.79, -.16]). When predicting reading comprehension with background knowledge, the regression slope was relatively flat $B= .12$ (β

=.18) with 95% confidence interval [.09, .15] ([.13, .23]) for students having a knowledge score below the threshold, and became significantly steeper *B*= .56 (β =.81) with 95% confidence interval [.51, .61] ([.73, .89]) for students whose knowledge score was above the threshold (Figure 1). Eighty-seven percent of students in the below threshold group (n=835) had a comprehension score lower than 15, which was equivalent to the grand mean comprehension score across all students in this sample. On the other hand, 91% of students (n=1356) whose comprehension score was above the mean also scored above the knowledge threshold. These results supported the idea that knowledge does not play the same facilitative role to comprehension when it was below vs. above the threshold.

To evaluate whether the threshold was specific to knowledge related to the reading topic, i.e., ecology terms, the broken-line relation was replicated with the subset of 303 students who also took the history topical vocabulary test, controlling for this history vocabulary. For this subsample, performance on the background knowledge section was *M*=38, *SD*=11; performance on the comprehension section was *M*=15, *SD*=7, almost identical to the whole sample, thus representative of the whole sample. Using this subsample, a knowledge threshold was identified at background knowledge (using ecosystems background knowledge) score of 30 (-.73), $p<.01$, with 95% confidence interval [22, 40] ([-1.51, .16]). When students' background knowledge was below the threshold, their background knowledge failed to predict comprehension *B*=0 (β =0) with 95% confidence interval [-.13, .13] ([-.26, .26]); when students' background knowledge was above the threshold, background knowledge was positively related to comprehension *B*=.51 (β =.79 with 95% confidence interval [.37, .64] ([.63, .85]). Importantly, this broken-line relation remained significant even after controlling for the effect of history vocabulary: the threshold remained at background knowledge=30 (-.73), $p<.01$, with 95% confidence interval [21,41] ([-1.52, .25]). Thus, the threshold has a topic specific component.

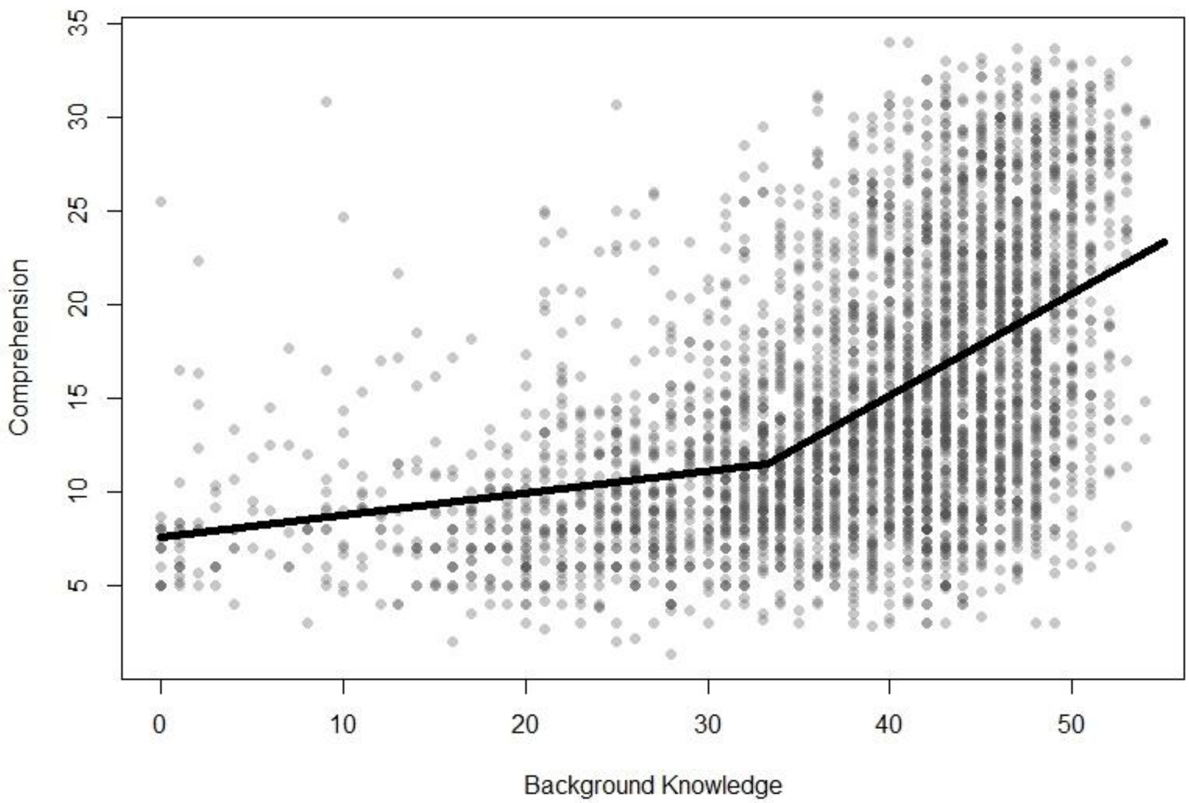

Figure 1. Non-linear relation between background knowledge and reading comprehension; shading of the dots reflects the number of overlapping cases.

As another test of the specificity of the threshold to relevant ecology knowledge, we identified which keywords were most predictive of students' knowledge status (above vs. below threshold). We calculated the correlation between student's performance on each keyword and the student's knowledge threshold status. The keywords differed in their predictability to students' knowledge threshold status (Figure 2). The most predictive keyword was *ecosystems,* which explained almost 30% of variance in students' knowledge threshold status. In contrast, other keywords were less predictive of students' threshold status. For example, the recognition of *densities* or *fauna* explained less than 3% of variance in students' threshold status.

Not surprisingly, many of the highest ranking topical keywords were also mentioned in the texts of the reading comprehension task (Figure 2, marked with asterisk). The correlation between students' comprehension and performance on the keywords that were mentioned in the reading texts was $r(3532) = .36$, $p<.01$, with 95% confidence interval [.33, .39]; the correlation between students' comprehension and performance on the keywords that were not mentioned in the reading texts was $r(3532)=.38$, $p<.01$, with 95% confidence interval [.36, .41].

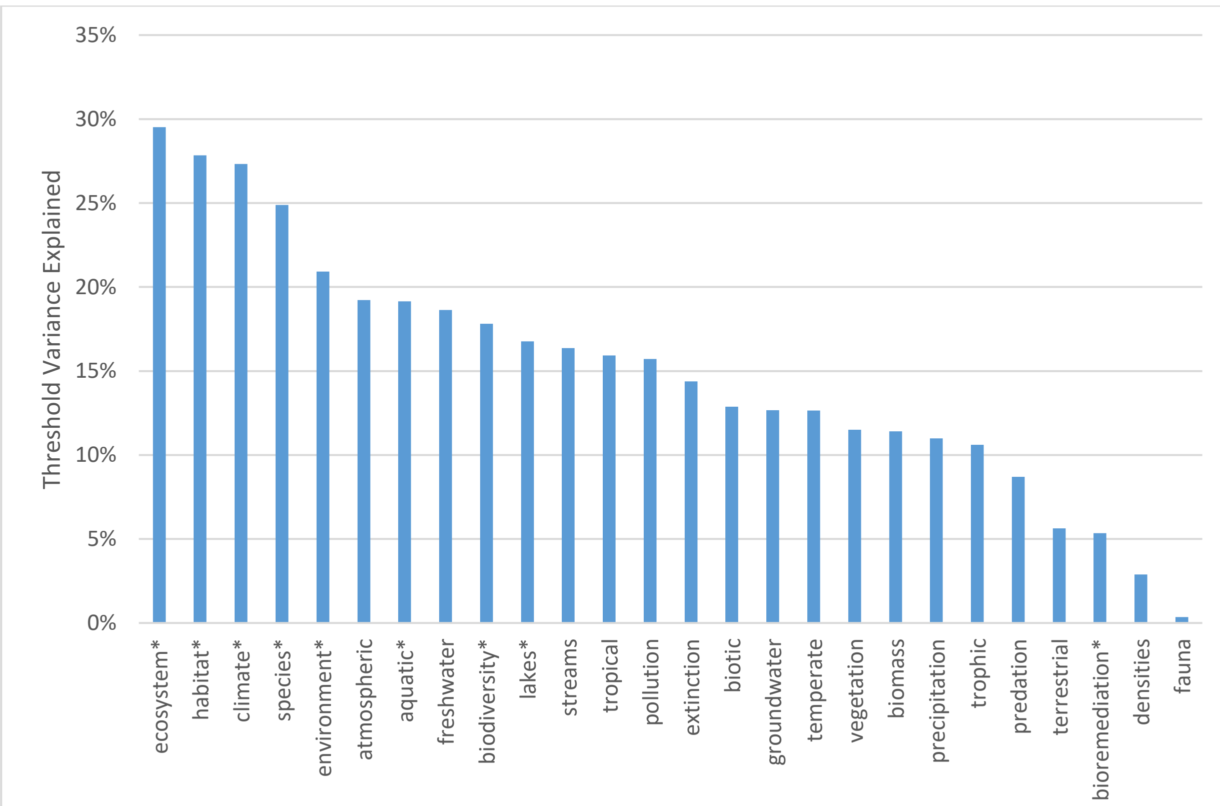

Figure 2. Threshold status variance explained by topical keywords. Keywords marked with an asterisk (*) appeared in texts in the reading comprehension task.

Interestingly, the ranking of keywords based on how much variance of threshold status they explained (Figure 2) converged well with the ranking of these words based on how likely these keywords occur in natural texts on the topic, the latter of which was reflected by the topical association index provided by Deane's (2012) database. The correlation of the two rankings (Spearman correlation) were $r(23)=.65$, $p<.01$, with 95% confidence interval [.34, .83]. After controlling for general word frequency, the two rankings were still significantly correlated, $r(22)=.62$, $p<.01$, with 95% confidence interval [.29, .82]. Thus, a measure of topical association that accounts for how frequent a word is in a given topic, is more predictive of threshold status than a measure based on how frequent a word appears in the general language. In other words, the threshold has a topic specific component beyond general vocabulary.

To evaluate the utility of the topical association index in identifying "must-know" keywords, we further compared both threshold groups' (above vs. below knowledge threshold) performance on a few keywords that had the highest topical association index to determine how many keywords we would need to reliably identify students who might be below the knowledge threshold. After some exploration by varying the number of the keywords, we found that by only using the top six keywords that had the highest topical association, we were able to correctly identify 74% of the students who were below the knowledge threshold, with 26% false alarm rate. While the above-threshold group had an average accuracy of 95% on these items (SD=10%), the below-threshold group's mean performance was only 64% (SD=33%). Thus, it appears that knowledge on these six words is critical for students to stay above the knowledge threshold.

Discussion

Our results support the knowledge threshold hypothesis. Using broken-line regression, we were able to identify a quantifiable point (59% correct on the knowledge test) at which there was a qualitative change in the relationship between background knowledge and reading comprehension. Below the threshold, the slope was relatively flat (*B*=.12), but above the threshold, increases in the level of a student's knowledge were strongly associated with increases in comprehension (*B*= .56). Importantly, the knowledge threshold seems to be specific to the knowledge on the domain of texts to be read. After controlling for the effect of history vocabulary in the subset of 303 students, a domain different from the reading comprehension texts, the knowledge threshold remains significant. The existence of this threshold suggests that students might need a minimum amount of knowledge of a topic to comprehend a text about that topic. Eighty-seven percent of the students who fell below the threshold scored below the mean on the comprehension assessment, whereas 91% of the students whose comprehension score was higher than the mean were above the knowledge threshold. Thus, there seems to be a qualitative change in the relationship between background knowledge and comprehension, this point can be quantifiably identified, and it is associated with different comprehension profiles.

We also found that some words were more predictive of exceeding the knowledge threshold than others. For instance, the words *ecosystems, habitat* and *species* are more predictive than other words such as *bioremediation, densities* and *fauna*. Interestingly, these more predictive words were also among the highest topically associated words as reflected by natural language processing-based statistics (Deane, 2012). The above threshold group achieved near ceiling performance (i.e. 95% correct) on six of the keywords that had the highest topical association index. This suggests that these words might be "must know" words for students in order to perform above the knowledge threshold. Indeed, simply using students' performance on these six words, we were able to correctly identify the great majority of students who were below the knowledge threshold with an acceptable false alarm rate (26%).

The results lend some support to activation models of reading comprehension such as the resonance model (Myers & O'Brien, 1998), which posits that the words in the text activate information described previously in the text as well as relevant background knowledge *not included* in the texts. Indeed, 17 of the 26 topical words in the knowledge measure were *not* mentioned in the text. The fact that these associated but not mentioned words were predictive of students' comprehension (r=.38, p<.01) supports the knowledge activation process of the resonance model (Myers & O'Brien, 1998). In other words, not only activation of keywords that are explicitly mentioned, but also those that are not mentioned in the texts are predictive of comprehension performance.

The current results have implications for instruction. Identifying which students may have a problem reading a given text on a particular topic is informative for teachers. While there are many reasons why a student may not comprehend, ranging from weakness in decoding (Wang et al., 2018), vocabulary (Hsueh-Chao & Nation, 2000; Laufer, 1989; Schmitt et al., 2011), or inference making (Cain, Oakhill, Barnes, & Bryant, 2001), the current study explored another possibility: limited background knowledge. Knowledge measures such as the one described here, would not take too much time (less than 3 minutes) from instruction or reading. We were able to correctly identify the great majority of students who fell below the threshold based on the six most frequently seen keywords for the topic. Having a quick measure of students' knowledge might be able to reveal the transition where knowledge starts to facilitate reading comprehension. In terms of the model of domain learning, such a transition may signal a possible shift from beginning (acclimating stage) to more capable levels of knowledge development (competent stage). More importantly, identifying those students who fall below the threshold is important as they are likely to have comprehension difficulties and should be targeted for additional instruction.

Limitations

While the results of the current study are encouraging, there are a number of limitations. First, although the comprehension assessment used in this study included a range of source texts, they were all on the same general topic (ecology). Future research is required to determine whether these results generalize to different topics or domains. Second, while we used two item types (topical vocabulary choice and factual multiple choice), future research should employ other item types as measures of knowledge to examine the generalizability and robustness of the threshold hypothesis. The position of the knowledge threshold may depend on the knowledge measure used. For example, an easier knowledge test might result in the threshold identified at a higher knowledge score, but the nonlinear relation between knowledge and comprehension should still be observed. This should be examined in future studies. Third, future research should explore whether the effects observed in this study transfer to different comprehension assessments (e.g., traditional reading comprehension test), or varying levels of text complexity. Fourth, future studies should examine the knowledge threshold hypothesis in different populations (e.g. middle school or college students). In short, while the precise numerical value of the threshold may change for different populations or materials, future research should explore whether an identifiable threshold may limit comprehension under different conditions.

Conclusion

Measuring students' background knowledge before they read a text may reveal which students are likely to have a reading comprehension problem and which may need to build additional background knowledge before reading. But how much knowledge is too little? The answer to this question is complex but is likely discernable with an empirically identifiable knowledge threshold.

References

Adams, M. (2014). lm. br: An R Package for Broken Line Regression. http://cran.pau.edu.tr/web/packages/lm.br/vignettes/lm.br.pdf

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. (1997). Mapping the multidimensional nature of domain learning: The interplay of cognitive, motivational, and strategic forces. In M. L. Maehr & P. R. Pintrich (Eds.), *Advances in motivation and achievement* (Vol. 10, pp. 213-250). Greenwich, CT: JAI Press Inc.

Alexander, P. A. (2003). The development of expertise: The journey from acclimation to proficiency. *Educational Researcher, 32*(8), 10-14.

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

Alexander, P. A., Jetton, T. L., & Kulikowich, J. M. (1995). Interrelationship of knowledge, interest, and recall: Assessing a model of domain learning. *Journal of Educational Psychology, 87*(4), 559-575.

Cain, K., Oakhill, J. V., Barnes, M. A., & Bryant, P. E. (2001). Comprehension skill, inference-making ability, and their relation to knowledge. *Memory & Cognition, 29*(6), 850-859.

Cook, A. E., & O'Brien, E. J. (2014). Knowledge activation, integration, and validation during narrative text comprehension. *Discourse Processes, 51*(1-2), 26-49.

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.

Deane, P. (2012). NLP methods for supporting vocabulary analysis. In J. Sabatini, T., O'Reilly, & L. Albro (Eds.), *Reaching an understanding: Innovations in how we view reading assessment*, (pp. 117-144). Lanham, MD: Rowman & Littlefield.

Duncan, L. G., Castro, S. L., Defior, S., Seymour, P. H., Baillie, S., Leybaert, J., . . . Porpodas, C. D. (2013). Phonological development in relation to native language and literacy: Variations on a theme in six alphabetic orthographies. *Cognition, 127*(3), 398-419.

Elbro, C., & Buch-Iversen, I. (2013). Activation of background knowledge for inference making: Effects on reading comprehension. *Scientific Studies of Reading, 17*(6), 435-452.

Gough, P. B., Hoover, W. A., Peterson, C. L., Cornoldi, C., & Oakhill, J. (1996). Some observations on a simple view of reading. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp. 1-13). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Hsueh-Chao, M. H., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language, 13*(1), 403-430.

Juul, H., Poulsen, M., & Elbro, C. (2014). Separating speed from accuracy in beginning reading development. *Journal of Educational Psychology, 106*(4), 1096-1106.

Kincaid, J. P., Fishburne Jr, R. P., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel*. Orlando, Florida: Institute for Simulation and Training.

Kintsch, W. (2004). The construction-integration model of text comprehension and its implications for instruction. In R. B. Ruddell & N. J. Unrau (Eds.), *Theoretical models* *and processes of reading* (Vol. 5, pp. 1270-1328). Newark, Delaware: International Reading Association, Inc.

Laufer, B. (1989). What percentage of text-lexis is essential for comprehension. In C. Lauren & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines* (pp. 316-323). Clevedon, UK: Multilingual Matters.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. J. L. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 126-132). London: Palgrave Macmillan.

McNamara, D. S., & O'Reilly, T. (2009). Theories of comprehension skill: Knowledge and strategies versus capacity and suppression. In A. M. Columbus (Ed.), *Progress in Experimental Psychology Research* (pp. 1-24). Hauppauge, NY: Nova Science Publishers, Inc.

Muggeo, V. M. (2008). Segmented: an R package to fit regression models with broken-line relationships. *R news, 8*(1), 20-25.

Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes, 26*(2-3), 131-157.

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.

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.

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, *95*(1), 26-43.

Shapiro, A. M. (2004). How including prior knowledge as a subject variable may change outcomes of learning research. *American Educational Research Journal, 41*(1), 159-189.

Sheehan, K. M., Kostin, I., Napolitano, D., & Flor, M. (2014). The TextEvaluator tool: Helping teachers and test developers select texts for use in instruction and assessment. *The Elementary School Journal, 115*(2), 184-209.

Wang, Z., Sabatini, J., O'Reilly, T., & Weeks, J. (2018). Decoding and reading comprehension: A test of the decoding threshold hypothesis. *Journal of Educational Psychology (Advance Online Publication)*. doi: 10.1037/edu0000302