



# Vocabulary, text coverage, word frequency and the lexical threshold in elementary school reading comprehension

Ulrich Ludewig<sup>1</sup> · Nicolas Hübner<sup>2</sup> · Sascha Schroeder<sup>3</sup>

Accepted: 28 October 2022  
© The Author(s) 2022

## Abstract

Vocabulary knowledge is one of the most important elements of reading comprehension. Text coverage is the proportion of known words in a given text. We hypothesize that text comprehension increases exponentially with text coverage due to network effects and activation of prior knowledge. In addition, the lexical threshold hypothesis states that text comprehension increases faster above a certain amount of text coverage. The exponential relationship between text coverage and text comprehension, as well as the lexical threshold, are at the heart of text comprehension theory and are of great interest for optimizing language instruction. In this study, we first used vocabulary knowledge to estimate text coverage based on test scores from  $N=924$  German fourth graders. Second, we compared linear with non-linear models of text coverage and vocabulary knowledge to explain text comprehension. Third, we used a broken-line regression to estimate a lexical threshold. The results showed an exponential relationship between text coverage and text comprehension. Moreover, text coverage explained text comprehension better than vocabulary knowledge, and text comprehension increased more quickly above 56% text coverage. From an instructional perspective, the results suggest that reading activities with text coverage below 56% are too difficult for readers and likely inappropriate for instructional purposes. Further applications of the results, such as for standard setting and readability analyses, are discussed.

**Keywords** Reading comprehension · Vocabulary · Text coverage · Elementary school

---

✉ Ulrich Ludewig  
ulrich.ludewig@tu-dortmund.de

<sup>1</sup> Institute of School Development and Education Research, TU Dortmund University, Vogelpothsweg 78, 44227 Dortmund, Germany

<sup>2</sup> Institute of Education, University of Tuebingen, Tübingen, Germany

<sup>3</sup> Department for Educational Psychology at the University Goettingen, Göttingen, Germany

## Introduction

Reading comprehension is a prerequisite for lifelong learning and one of the key goals of elementary education (e.g., Artlet et al., 2003). It is a multi-faceted construct that involves multiple components (e.g., Graesser et al., 2004). Vocabulary knowledge is one of the most influential determinants of reading comprehension during elementary school (e.g., McElvany et al., 2009; Quinn et al., 2015). According to the Simple View of Reading (Gough & Tunmer, 1986), reading comprehension involves two components: word recognition and language comprehension. Vocabulary knowledge is related to both language comprehension and word recognition (Duke & Cartwright, 2021). Vocabulary knowledge provides a link between phonology, orthography, and word meanings (e.g., Ehri, 2014).

Based on the hierarchical relations among reading sub-components (e.g., Kim, 2020), problems with lower-order reading components, such as word recognition and vocabulary, result in problems with higher-level components, such as inference-making. Thus, Wang et al. (2019) found a minimum level of word recognition fluency that is necessary for higher-level reading processes. Based on the Model of Lexical Quality (Perfetti, 2007), they suggested that efficient word recognition clears the way for higher-level reading processes, and therefore, problems in word recognition eventually lead to problems in higher-level processes (also Karageorgos et al., 2020). Similarly, O'Reilly et al. (2019) argued regarding vocabulary knowledge that the activation of prior knowledge only spreads properly if a critical number of known content words are present in a text. Thus, text comprehension increases above a certain level of known words in the text.

In this article, we first discuss the relationship between vocabulary knowledge (i.e., the overall number of words a person knows) and text coverage (i.e., the number of known words in a specific text). Second, we examine the linear and non-linear relationship between text coverage and text comprehension. Third, we identify thresholds that can help improve instruction and assessment of reading comprehension.

## Vocabulary knowledge and reading comprehension

Vocabulary knowledge is a multi-faceted construct (e.g., Perfetti & Hart, 2002) that is highly associated with the ability to read fluently and comprehend texts (Perfetti, 2007). Two important sub-dimensions of vocabulary knowledge are vocabulary breadth, i.e., the number of words known, and vocabulary depth, i.e., how much knowledge about semantic, orthographic, and phonological aspects of a word are available (Li & Kirby, 2015). Previous research has shown that vocabulary breadth is more strongly associated with reading comprehension than vocabulary depth (e.g., Li & Kirby, 2015; Ouellette, 2006). Additionally, semantic knowledge has a stronger association with reading comprehension than orthographic or phonological

knowledge (Richter et al., 2013). Studies generally report strong associations between vocabulary knowledge and reading comprehension (e.g., English: Quinn et al., 2015; German: Richter et al., 2013). Thus, it seems that knowing the meaning of many words increases the probability of correctly recognizing and comprehending the words in a particular text (Perfetti, 2007).

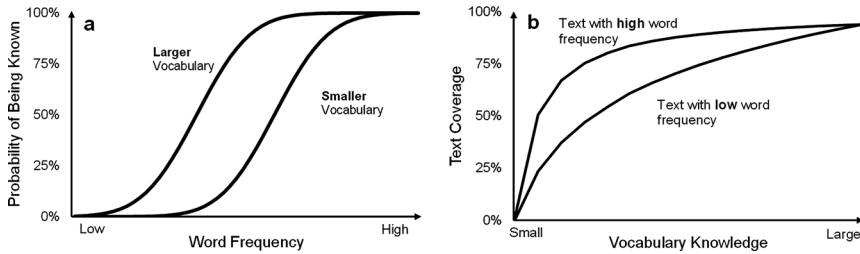
### Vocabulary knowledge and text coverage

Text coverage is usually defined as the proportion of words in a text that are known by a particular reader (Hsueh-Chao & Nation, 2000). More specifically, text coverage can be understood as the intersection between the words in a given text and a reader's vocabulary knowledge. It takes relatively few unique words to reach a relatively high text coverage in most texts (Hsueh-Chao & Nation, 2000). According to Zipf's theorem, when the words in a text are ordered according to their frequency, their probability of occurrence is inversely proportional to their place on the frequency list (Piantadosi, 2014). Thus, a small number of words occur very often and many words occur very rarely in authentic texts. Corpus analysis with large samples of texts shows that knowledge of only the 2000 most frequent words is sufficient to achieve an average text coverage of 90.6% for narrative texts and an average text coverage of 78.4% for academic texts (Nation & Waring, 1997). Text coverage for academic texts is lower because such texts include more rare words. Additionally, the relationship between text coverage and the length of the frequency ranked word list (FRWL) is logarithmic; for instance, the first 1000 most frequent words provide 72% text coverage, and the next 1000 only add 7.7 percentage points to text coverage (Nation & Waring, 1997).

To our knowledge, no previous study has examined the relationship between text coverage and readers' actual vocabulary knowledge for a representative sample of texts and/or participants. In a FRWL, the frequency of a word determines whether the word is included in the list or not. For vocabulary knowledge, this relationship is not deterministic but probabilistic, as frequent words are more likely to be known than rare words (e.g., for a review: Brysbaert et al., 2018). Overall, the correlation between the probability of knowing a word and its frequency is high (German third and fourth graders  $r=0.74$ : Trautwein & Schroeder, 2018). Therefore, the text coverage of a given FRWL and actual vocabulary may be very similar.

Figure 1 panel a illustrates probabilistic relationships between vocabulary knowledge and word frequency. Students with larger vocabularies are more likely to know more rare words compared to students with smaller vocabularies (Brysbaert et al., 2018).

Figure 1 panel b illustrates the logarithmic relationship between vocabulary knowledge and text coverage. The relationship between vocabulary knowledge and text coverage should have a logarithmic shape, similar to the relationship between a FRWL and text coverage. Additionally, given the same vocabulary knowledge, text coverage should be lower for a text with lower compared to higher average word frequency.



**Fig. 1** Diagram illustrating relationships between word frequency, vocabulary knowledge, and text coverage. *Note* Illustrative diagram (no actual data displayed). Panel a analog to Brysbaert et al., 2018. Panel b analog to Chujo and Utiyama (2005)

## Text coverage and reading comprehension

Text comprehension substantially depends on text coverage. According to the construction-integration model (Kintsch, 1988), readers' mental representation of a text is an associative network of concepts and propositions. In this network, concepts represent nodes and associations represent links. The more words are known, the more concepts and the more prior knowledge can be activated. The number of possible associations between concepts grows exponentially with the number of activated concepts. It is much easier for readers to disambiguate the meaning of a text if the words in the text immediately activate the correct concepts. Disambiguating text meaning is highly important for integrating text information with prior knowledge (Richter & Schnotz, 2018).

On one hand, readers are usually able to comprehend texts even when they contain some unknown words. This is because readers can make inferences based on contextual information to infer the meaning of unknown words if the network of associations between the known concepts is strong enough (Share & Stanovich, 1995). However, contextual inferences require additional cognitive resources or can lead to false interpretations, which makes text comprehension more challenging when text coverage is low (Cain et al., 2004). Indeed, drawing inferences from the context and building up an understanding of the text is only possible once text coverage reaches a certain level. In the lexical threshold hypothesis (Hsueh-Chao & Nation, 2000), text comprehension is assumed to be significantly impaired below a certain amount of text coverage.

## Lexical threshold hypothesis

The lexical threshold hypothesis states that text comprehension increases faster above a certain amount of text coverage (Hsueh-Chao & Nation, 2000). Relatively few and heterogeneous findings exist about the lexical threshold hypothesis. For instance, Hsueh-Chao and Nation (2000) found that individuals need to know the meaning of 98% of the words in a fictional text for comprehension in a reading for pleasure situation, where unknown words were assessed by self-report. In another study, Laufer (1989) reported that reading comprehension increased more rapidly

if individuals knew at least 95% of the words in a text. In this study, individuals were required to translate a vocabulary list in order to determine their text coverage. Laufer and Ravenhorst-Kalovski (2010) suggested two thresholds, 98% and a minimum at 95%. Their analysis was based on participants with high prior knowledge as well as a standardized test of vocabulary size and reading comprehension test. More recently, O'Reilly et al. (2019) found that the reading comprehension of ninth- to twelfth-graders increased rapidly when they knew more than 59% of the critical content words in a text. In this study, knowledge of critical content words was assessed with a multiple-choice test. By contrast, Schmitt et al. (2011) were not able to determine a clear lexical threshold in a carefully designed study with a word-nonword recognition test and a standardized reading comprehension test.

### Summary of the theoretical background

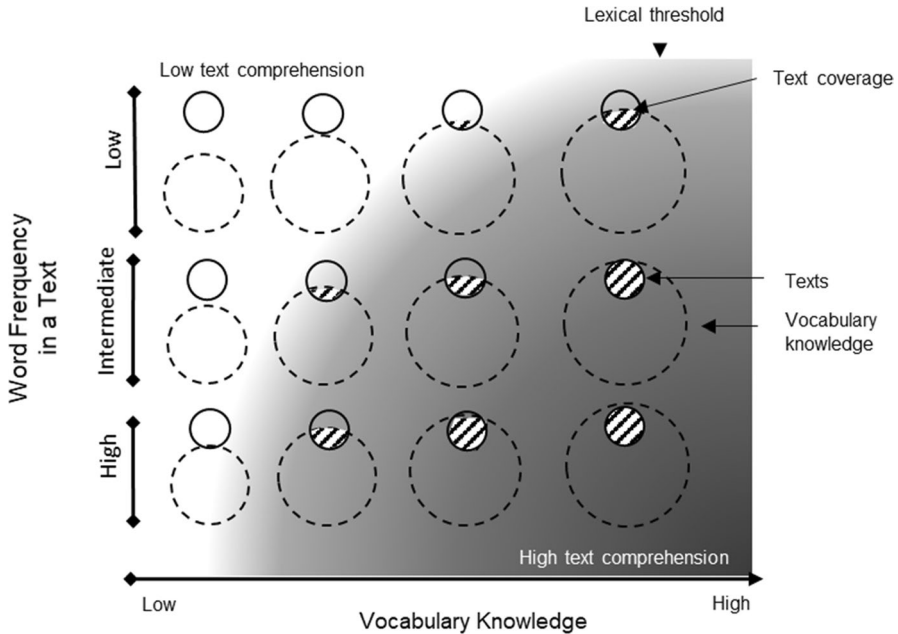
Vocabulary knowledge, text coverage, word frequency, and text comprehension are theoretically related: Vocabulary increases text coverage logarithmically, and this relationship depends on the word frequency in the text (word frequency effect: Brysbaert et al., 2011; Zipf's theorem: Piantadosi, 2014). Text comprehension theory (i.e., the construction-integration hypothesis) assumes exponential growth in connectivity and activation of prior knowledge, which means that increasing text coverage should exponentially improve text comprehension (Share & Stanovich, 1995). The lexical threshold hypothesis states that text comprehension increases faster above a certain threshold of text coverage.

Figure 2 summarizes the described relationships. The larger dashed circles represent a person's vocabulary knowledge and the smaller solid circles represent texts. The intersection between the two circles (i.e., the area with diagonal lines) is the text coverage. Texts with many rare words are more likely to be covered when persons have a larger vocabulary knowledge. The discontinuous color scale from white (upper left corner) down to almost black represents the degree of text comprehension. Text comprehension increases with more text coverage. It takes a certain amount of text coverage before comprehension increases more rapidly.

### Research question

Although the relationships between (1) vocabulary knowledge and reading comprehension, (2) vocabulary knowledge and text coverage, and (3) text coverage and reading comprehension have been investigated in separate contexts, they have rarely been researched using an integrative approach.

In this study, we used data from a vocabulary knowledge and a text comprehension test administered to a large number of fourth graders participating in a reading support program. We analyzed word frequencies from vocabulary test items and the reading comprehension texts to estimate text coverage for each participant for each text. We compare linear and non-linear models of text coverage explaining text comprehension. In addition, we investigated whether vocabulary knowledge or text coverage were better able to predict children's text comprehension. Finally, we



**Fig. 2** Diagram summarizing the relationship between word frequency, vocabulary knowledge, text coverage, text comprehension and the thresholds. *Note* Solid circles: text, dashed circles: vocabulary knowledge, area with diagonal lines: text coverage, position of the solid circle on the y-axis: number of rare words in a text, X-axis with increasing diameter of dashed circles: increase in vocabulary knowledge from left to right. Color gradient from white (left upper corner)=low text comprehension to black (right lower corner)=high text comprehension

determined potentially relevant amounts of text coverage in order to define various thresholds.

The study's three central research questions (RQ) can be summarized as follows:

RQ1: What is the shape of the relationship between text coverage and text comprehension?

We hypothesize that text comprehension increases exponentially rather than linearly, due to the effect of network connectivity on the propositional network and activation of prior knowledge.

RQ2: Does text coverage explain text comprehension better than vocabulary knowledge?

We hypothesize that text coverage better explains vocabulary knowledge because it more accurately describes the words known in a given text.

RQ3: Is there an amount of text coverage can be defined as a lexical threshold?

We hypothesize that text comprehension increases faster above a certain level of text coverage.

## Method

### Participants

The children who participated in the study attended 4th grade and were tested at the beginning of the second half of the school year. Fourth graders are typically required to comprehend texts independently, so vocabulary knowledge (i.e., knowing the meaning of words) and especially text comprehension become important. This study is a program evaluation of a project to promote language and literacy skills among fourth graders at public schools in six different German states. The program provided teachers with scientifically grounded teaching materials and handouts. Only students with parental consent were included in the present analysis. The study involved  $N_i=949$  fourth graders from  $N_c=64$  classes and  $N_s=35$  schools. About half of the participants were female, 52.05%, and children were on average  $M=10.28$  years old,  $SD=0.52$ . Overall, 64.91% of the students reported exclusively speaking German at home. The program was conducted in federal states where the share of public school students from immigrant backgrounds ranged from 50.1 to 28.9% (Stanat et al., 2017, p. 299). Thus, participants are relatively representative of these federal states. However, we conducted robustness checks to assess the impact of language background and discuss this in the limitations. We excluded 25 (2.63%) participants because they answered fewer than 50% of the items for either the vocabulary or the text comprehension test. Thus, we analyzed the test results of  $N_i=924$  participants.

### Materials

#### Vocabulary knowledge test

Vocabulary knowledge was assessed with the synonym-based vocabulary knowledge test, KFT 4–12 + R V1 (Heller & Perleth, 2000). We used this test because we considered it a good measure of ‘knowing’ the meaning of words in line with the theory that words represent nodes in an associative network. This paper–pencil test included 25 items presented in fixed order and administered under low time constraints. Thus, most students responded to all items. The items consisted of one item stem word and five response options with one key (see Fig. 3). The distractors were orthographically similar (i.e., curved versus covered) and/or semantically related (e.g., anonyms or meronyms), but not synonyms.

#### Text comprehension test

The standardized text comprehension test was the Aspects of the Learning Situation and Learning Development Test (LAU; Lehmann et al., 2002). This test includes four texts with multiple-choice (MC) items, "Mosquito" (124 words, 11 sentences, 4 items), "Candle" (106 words, 8 sentences, 7 items), "I am not blind" (206 words, 8 sentences, 7 items), and "Plastic duck" (125 words, 7 sentences, 7 items). Figure 4

## Example Item of the Vocabulary Knowledge Test

Which word has the most similar meaning to the bold word?

Item stem

(key)

**curved**

straight

round

covered

cloned

merged

**Fig. 3** Example item of the vocabulary knowledge test. *Note* Illustrative example of a typical item from the vocabulary knowledge test. This item was not actually in the test. The vocabulary knowledge test is protected by copyright

**a**

**Die ziemlich intelligente Fliege**

Eine große Spinne hatte in einem alten Haus ein schönes Netz gewoben, um Fliegen zu fangen. Jedes Mal, wenn eine Fliege sich auf dem Netz niederließ und darin hängen blieb, verzehrte die Spinne sie schleunigst, damit andere Fliegen, die vorbeikamen, denken sollten, das Netz sei ein sicherer und gemütlicher Platz.

Eines Tages schwärzte eine ziemlich intelligente Fliege so lange um das Netz herum, ohne es zu berühren, dass die Spinne schließlich hervorkroch und sagte: „Komm, ruh' dich bei mir ein bisschen aus.“ Aber die Fliege ließ sich nicht überbelpeln: „Ich setz' mich nur an Stellen, wo ich andere Fliegen sehe“, antwortete sie, „und ich sehe bei dir keine anderen Fliegen.“

Damit flog sie weiter, bis sie an eine Stelle kam, wo sehr viele Fliegen saßen. Sie wollte sich gerade zu ihnen gesellen, als eine Biene ihr zurief: „Halt, du Idiot, hier ist Fliegenleim. Alle diese Fliegen sitzen rettungslos fest.“ – „Red' keinen Unsinn!“, sagte die Fliege. „Sie tanzen doch.“ Damit ließ sie sich nieder und blieb auf dem Fliegenleim kleben wie all die anderen Fliegen.

Moral: Der Augenschein kann ebenso trügerisch sein wie die Sicherheit, in der man sich wägt.

Was tat die Spinne jedes Mal, wenn eine Fliege in ihrem Netz hängen bleibt?

a Sie reparierte das Netz sorgfältig

b Sie freute sich sehr, das es etwas zu essen gab

c Sie verzehrte die Fliege schleunigst

d Sie baute sofort ein neues Netz

**b**

**The pretty intelligent fly**

A large spider had woven a beautiful web in an old house to catch flies. Every time a fly settled on the web and got stuck in it, the spider would quickly consume it so that other flies passing by would think the web was a safe and cozy place.

One day a pretty intelligent fly buzzed around the web for so long without touching it that the spider finally crawled out and said, „Come rest with me for a while.“ But the fly would not be duped: „I only sit down in places where I see other flies,“ it replied, „and I see no other flies with you.“

With that, it flew on until she came to a place where a great many flies were sitting. It was about to join them when a bee called out to it, „Stop, you idiot, there's fly glue here. All these flies are stuck beyond repair.“ - „Don't talk nonsense,“ said the fly. „They're dancing“ with that, it settled down and stuck to the fly glue like all the other flies.

Moral: Appearances can be just as deceptive as the security into which one lulls oneself.

What did the spider do every time a fly got caught in the web?

a It repaired the web carefully

b It was very happy that there was something to eat

c It ate the fly as soon as possible

d It immediately built a new web

**Fig. 4** Example item text comprehension test. *Note* Released example item from the aspects of the learning situation and learning development test. **a** Original German and **b** english translation

shows an example item. The test was administered with low time constraints; thus, most students completed all items.

## Word frequencies

We derived word frequencies for the vocabulary test items and reading comprehension text from the ‘childLex’ corpus. The childLex corpus ([www.childlex.de](http://www.childlex.de)) includes 500 books classified as appropriate for children 6–12 years of age, including overall 9.85 million running words (i.e., “token”) and 182 thousand unique words (i.e., “types”). Normalized lemma frequencies were used for all analyses.

All words in the vocabulary test, but not all words in the reading comprehension texts, were part of the childLex corpus (see Table 1). Most of the non-included



**Table 1** Overview of word frequencies by text

Texts	Mosquito		Candle		Blind		Plastic	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Zipf <sup>1</sup> interval <sup>2</sup>								
2–3	8	6.45	0	–	21	10.40	19	15.57
3–4	13	10.48	13	12.26	30	14.85	16	13.11
4–5	20	16.13	24	22.64	32	15.84	20	16.39
5–6	23	18.55	25	23.58	27	13.37	17	13.93
6–7	41	33.06	25	23.58	52	25.74	31	25.41
7–8	19	15.32	19	17.92	40	19.80	19	15.57
Total	124	100%	106	100%	202	100%	122	100%
Not found <sup>3</sup>								
Token	2	1.61	0	0	12	5.94	12	9.83
Type	1		0	0	12		9	
<i>M</i>	5.58		5.61		5.33		5.19	
<i>SD</i>	1.72		1.28		1.44		1.80	

Token=running words in the text, types=unique words in the text, *n*=number of tokens,%=relative proportion of tokens

<sup>1</sup>Zipf value with Laplace transformation= $\log(\text{'lemma frequency} + 1) / \text{'number of unique lemma in corpus} + \text{number of words in the corpus}'$ ), <sup>2</sup>Lower boundary defined as larger than and upper boundary as smaller than or equal to, <sup>3</sup>Not-found tokens were assigned a value of 2.00 based on the Laplace transformation and were counted in the interval '2–3'

words were proper nouns or compound words, and their frequencies were interpolated using Laplace approximations (Diependaele et al., 2013). Based on the interpolated normalized lemma frequencies, so-called Zipf values were computed (Van Heuven et al., 2014). This scale is logarithmic and scaled such that a value of 3 corresponds to the frequency of a word occurring once in a million words, a value of 4 corresponds to a frequency of ten times in a million words, a value of 5 100 times in a million words, etc. The word frequencies in the vocabulary test were on average  $M=3.99$ ,  $SD=0.56$ , and ranged from 2.77 to 4.87.

The "Mosquito" and "Candle" texts had similar word frequency distributions and a similar length. The "Plastic duck" text was similarly long as these two texts but encompassed more infrequent words. The "Blind" text was longer and encompassed more infrequent words than "Mosquito" and "Candle". The differences in word frequency distributions between texts indicate that the texts had different vocabulary knowledge requirements.

## Procedure

The study was conducted in the morning hours in all classes and administered in paper–pencil format. First, the text comprehension test was administered (30 min),

followed by a decoding speed test “Würzburger Leise Leseprobe-Revision” (Würzburger silent reading test revised; WLLR; 5 min; Schneider, 2011), a word analogy test (6 min; KFT 4–12+R V3), and the vocabulary knowledge test (6 min; KFT 4–12+R V1). We only use the text comprehension test and the vocabulary knowledge test in the analysis. The tests were administered in accordance with their test manuals.

## Analysis

### Data quality

In a preparatory step, we conducted an item fit analysis because misfitting items in the text comprehension and vocabulary tests might lead to false interpretations of the test results. We applied the Rasch model (Adams & Wu, 2007) to the response data for the text comprehension and vocabulary tests using the package Test Analysis Modules (TAM; Robitzsch et al., 2021) within R (R Core Team, 2021). We identified three items in the text comprehension test with an outfit or infit below 0.7 or above 1.3 (Gustafsson, 1980). An inspection of these items suggested that they had somewhat ambiguous answers. Even readers with otherwise high reading comprehension abilities did not answer these items correctly. We decided to exclude these three items from the text comprehension test since they might not actually measure comprehension. No items were excluded from the vocabulary knowledge test based on this analysis.

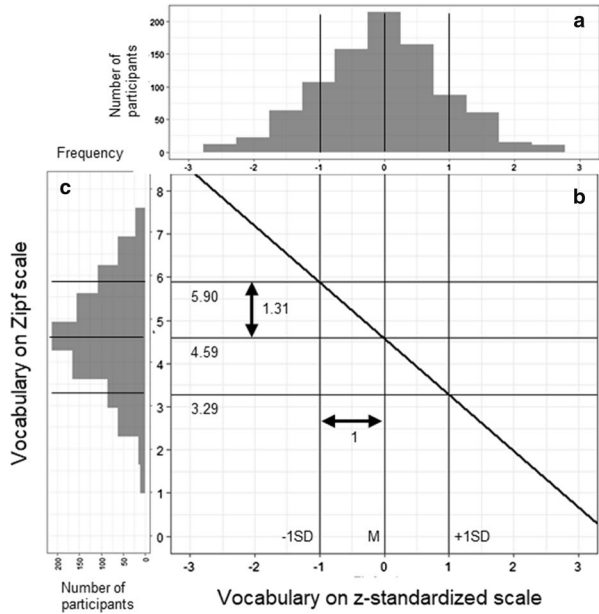
The relationship between item difficulty in the vocabulary knowledge test and the item’s word frequency was very important for the text coverage estimation. In the original 25 items, item difficulty and minimum word frequency for the synonym pair correlated only with  $r(23) = -0.33$ . However, after excluding five items with highly synonymous and orthographically similar distractors, the correlation was  $r(18) = -0.64$ . We considered this to be more consistent with previous findings on word frequency effects in German 4th graders (e.g.,  $r = -0.74$ ; Trautwein & Schroeder, 2018).

The overall rate of missing (i.e., omitted) responses was low (LAU: 2.97% and KFT: 5.65%). Missing responses were treated with the full information maximum likelihood method (FIML).

### Modeling vocabulary knowledge and text coverage

Text coverage is the intersection between a text’s words and the reader’s vocabulary knowledge. We estimated text coverage values for each child and each text. The rationale behind the estimation process was to reference children’s vocabulary knowledge test scores to the word frequency level they are likely to know and then determine which words in a text were likely to be known by each child.

**Fig. 5** Distribution of vocabulary knowledge before and after the linear transformation. *Note* Panel a shows the distribution of vocabulary knowledge before the linear transformation, panel b shows the relationship between vocabulary knowledge test scores on the z-standardized and Zipf scales, and panel c shows the distribution of vocabulary knowledge after the linear transformation

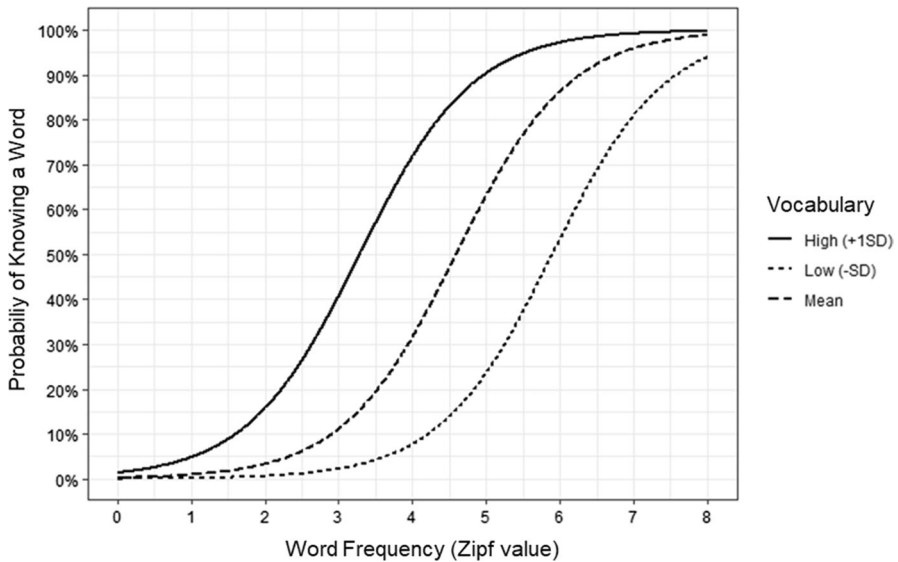


**Step 1: referencing vocabulary test score to word frequency**

The vocabulary test responses were modeled with a Rasch model using the TAM (Robitzsch et al., 2022) within R (R Core Team, 2021). Then, we regressed the minimum word frequency of the synonym pair on the item difficulty parameter  $\sigma$  ( $WF = b_0 + b_1\sigma + \epsilon$ ). The regression revealed significant regression coefficients of  $b_0 = 4.59$ ,  $SE = 1.51$ ,  $p = 0.007$  and  $b_1 = -1.31$ ,  $SE = 0.37$ ,  $p = 0.003$ . The intercept implies that an average item  $\sigma = 0$  has an expected word frequency of  $WF = 4.59$ . The slope indicates that a difficult item  $\sigma = 1$  has an expected word frequency of  $WF = 3.28$  and an easier item  $\sigma = -1$  has an expected  $WF = 5.90$ .

Based on these results, we transformed the Rasch scale,  $N(0, 1)$ , so that the item parameters were on the same scale as the expected word frequency of each item. We refer to this as the Zipf scale because this scale represents students' vocabulary knowledge as a function of word frequency,  $N(4.59, 1.30)$ .

Figure 5 panel a shows vocabulary knowledge on a Rasch scale with a mean of 0 and a standard deviation of 1. Negative values represent low vocabulary knowledge because the probability of answering a vocabulary test item correctly is low. Positive values represent high vocabulary knowledge because the probability of answering a vocabulary test item correctly is high. Figure 5 panel b shows the linear relationship between item difficulty and word frequency. An item of average difficulty has an expected word frequency of 4.59, a difficult item (i.e.,  $M - 1 SD$ ) an expected word frequency of 3.29, and an easy item (i.e.,  $M + 1 SD$ ) an expected word frequency of 5.90. Figure 5 panel c shows the distribution of vocabulary on the Zipf scale. On this scale, an average person has a value ( $\theta_p^{wf}$ ) corresponding to the expected word frequency of an item with average item difficulty.



**Fig. 6** Relationship between word frequency and the probability of knowing a word relative to vocabulary knowledge. *Note* Figure analogous to Fig. 2 in Brysbaert et al. (2018). Vocabulary: High = 1 ( $M + 1 SD$ ), Mean = 0, Low = -1 ( $M - 1 SD$ ). The probability of knowing a word is 50% when vocabulary knowledge on the Zipf scale is equal to the frequency of a word

## Step 2: probability of knowing a word

This step models the relationship illustrated in Fig. 1 panel a. The probability of a person ( $p$ ) knowing a word ( $w$ ) based on the frequency of the word ( $WF$ ) and the Zipf scale ( $\theta_p^{wf}$ ) formula F1:  $PR(\text{word known})_{pw} = \exp(b_1 * (\theta_p^{wf} - WF_w)) / (1 + \exp(b_1 * (\theta_p^{wf} - WF_w)))$ . This function was used to calculate the probability of each person knowing each word as described by Brysbaert et al. (2018) and illustrated in Fig. 6. The text coverage is the average probability of knowing each word in the text or the proportion of words estimated to be known out of the total number of words in the text.

## Modeling text coverage and text comprehension

We addressed RQ1 by comparing linear and exponential models of text coverage explaining text comprehension. We used a latent regression Rasch model (De Boeck & Wilson, 2004) that in the baseline model explains the probability of correctly solving a test item based on random effects for item difficulty and a random effect for person ability. For the explanatory models, we additionally included linear and quadratic terms for text coverage as a text-by-person covariate or vocabulary knowledge as a person covariate.

A latent regression Rasch model has the advantage that the regression coefficients represent the relationship between text coverage and measurement error-adjusted

text comprehension. This modeling approach increases interpretability, as the hypotheses relate to a text comprehension measure that is free of measurement error, and increases the reproducibility of our estimates, as imperfect reliability of the text comprehension test should bias the regression coefficients much less. The model was specified within the generalized linear mixed-effect model framework (GLMM) and fitted using the package ‘lme4’ (Bates et al., 2014) in the R environment (R Core Team, 2021).

The difference in random variance in person ability ( $\sigma_{\theta}^2$ ) between a model without text coverage (i.e., baseline model) and the explanatory models with text coverage were used to calculate the explained variance in person ability  $R_{\theta}^2$ ,  $\left(\sigma_{\theta \text{ baseline}}^2 - \sigma_{\theta \text{ text coverage}}^2\right) / \sigma_{\theta \text{ baseline}}^2$ . We used marginal  $R^2$  ( $mR^2$ ) to estimate the variance in the responses explained by the fixed effects (Nakagawa & Schielzeth, 2013).

We compared the model fits using the Bayes Information criterion (*BIC*). The *BIC* are model fit indicators (i.e., goodness of fit) that prevent overfitting by penalizing the number of variables in the model (in contrast to deviance or pseudo  $R^2$ ) and can be used to compare nested and unnested models. Lower values correspond to a better goodness of fit and fit differences of 5 – 10 points can be considered substantive (Burnham & Anderson, 2002). Additionally, we use Akaike weights ( $w_i$ ), which can be directly interpreted as conditional probabilities for each model (Wagenmakers & Farrell, 2004).

We addressed RQ2 by estimating similar linear and quadratic models using vocabulary knowledge as the predictor variable. We evaluate which models (i.e., out of all text coverage and vocabulary knowledge models) fit the data better using the same fit indicators described above.

## Modelling the lexical threshold

We addressed RQ3 using a broken-line regression (Muggeo, 2008) with average text coverage predicting the sum score on the text comprehension test. Broken-line regression is a statistical method that identifies a changepoint in a linear regression. It also provides a 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 at the identified changepoint. This method has been used in related research (O’Reilly et al., 2019; Wang et al. 2019). Based on our theoretical background regarding activation and propositional network connectivity, we expected an exponential increase and not necessarily a linear relationship with a changepoint. However, the changepoint is important from a practical perspective because decisions about factors such as text alignment are often binary (i.e., is the text too difficult for a particular student or not).

**Table 2** Results of the vocabulary and text comprehension tests

	Test	Raw score		Range <sup>4</sup>			Cor <i>r</i>
		<i>M</i>	<i>SD</i>	<i>Min</i>	<i>Max</i>	<i>Rel<sub>wlc</sub></i> <sup>3</sup>	
1	Vocabulary <sup>1</sup>	11.65	3.59	2	20	.67	.61 <sup>*</sup>
2	Text comprehension <sup>2</sup>	11.11	3.57	1	20	.69	

\*  $p < .001$

<sup>1</sup>KFT 4–12+R V1 (Heller & Perleth, 2000), <sup>2</sup>Aspects of the learning situation and learning development (LAU; Lehmann et al., 2002), <sup>3</sup>Weighted likelihood estimate reliability for vocabulary and text comprehension, <sup>4</sup>Actual range

## Results

### Vocabulary knowledge and reading comprehension

The descriptive results on raw score mean and standard deviation, range, reliability and correlations indicate that the tests for vocabulary and text comprehension worked in the intended way (Table 2). Both the vocabulary knowledge ( $Rel_{wlc}=0.67$ ) and text comprehension ( $Rel_{wlc}=0.69$ ) tests had an acceptable reliability for a large-scale assessment context. The vocabulary and text comprehension tests correlated highly with each other,  $r(922)=0.61$ ,  $t=23.2$ ,  $p < 0.001$ .

### Vocabulary knowledge and text coverage

The process of text coverage estimation might be best demonstrated with an example and a visual overview. Table 3 shows the words from one sentence of the "Mosquito" text together with the probability that each specific word will be known by children with varying levels of vocabulary knowledge. High-frequency words such as "die [the]" (WF=7.7) had a high probability of being known by both high- (100%) and low-skilled readers (92%). The probability that a low-frequency word such as "Stechmücken [mosquitoes]" (WF=2.8) would be known was 36% for a high-skilled reader and 2% for a low-skilled reader. Words with intermediate frequency such as "saugen" [suck] had a 79% probability of being known for a high-skilled reader and 11% for a low-skilled reader. Thus, the difference between children with high and low vocabulary knowledge was more pronounced for low- and average-frequency words. The text coverage is the average probability of knowing each word in a text.

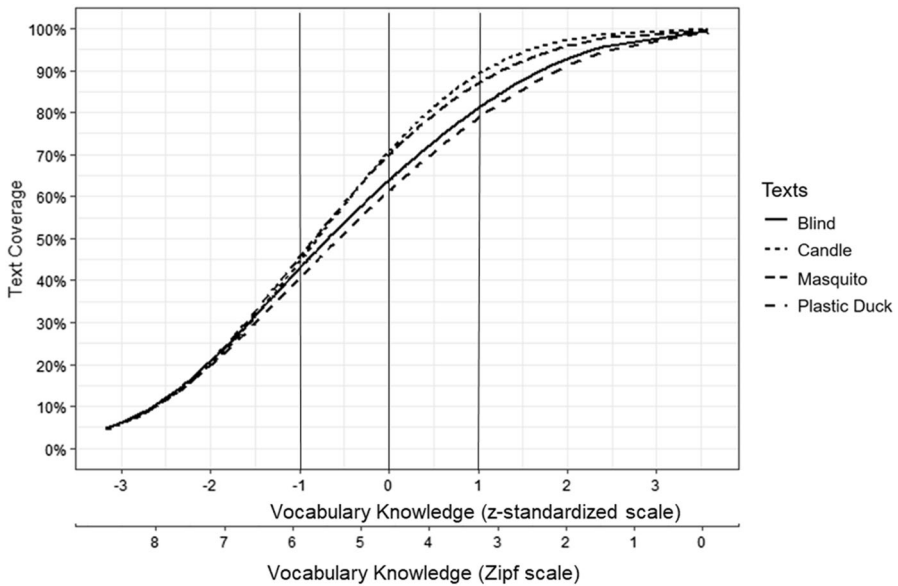
The text coverage estimation yielded average text coverage scores ranging from 65% for "Plastic duck" to 74% for "Mosquito". Figure 7 shows the text coverage relative to vocabulary knowledge for each text. The estimated text coverage increases with vocabulary knowledge up to mean + 2 *SD* and then nears 100% for all texts.

For instance, the "Plastic duck" text has the most infrequent words and "Candle" the fewest infrequent words. The differences in text coverage between texts are higher for students with high and mean vocabulary knowledge than for students with

**Table 3** Example sentence with probabilities of knowing each word relative to the WF of words and vocabulary knowledge

[English]	Only	the	females	, of the	mosquitoes	Suck	blood
[German]	Nur	die	Weibchen	, der	Stechmücken	Saugen	Blut
WF	6.4	7.7	3.7	7.7	2.8	4.3	4.9
Probability of knowing the word by vocabulary knowledge							
High <sup>1</sup>	98%	100%	65%	100%	36%	79%	89%
Mean	91%	98%	25%	98%	9%	41%	61%
Low	65%	92%	6%	92%	2%	11%	22%

Sentence from “Mosquito” text. <sup>1</sup>Ability estimates on Zipf scale:  $N(4.59, 1.30)$ . High = 3.3 ( $M-1 SD$ ), Mean = 4.59, Low = 5.9 ( $M+SD$ ). On the Zipf scale, low values correspond to high vocabulary knowledge because the value describes the probability of knowing infrequent words. The probability of knowing a word is 50% when the value of the vocabulary knowledge estimate equals the value of the word. The categorization (high, mean, low) was only used to derive illustrative examples and did not influence the actual estimation



**Fig. 7** Text coverage for each text in relation to vocabulary knowledge. *Note* Percent text coverage estimate (y-axis). Vocabulary knowledge on z-standardized scale and the Zipf scale (x-axis). The Zipf scale is inverse to a z-standardized scale. Low values correspond to high vocabulary knowledge because the scale refers to the word frequencies individuals are likely to know. Vertical lines indicate low vocabulary knowledge (i.e.,  $M-1 SD$ ), mean vocabulary knowledge, and high vocabulary knowledge (i.e.,  $M+1 SD$  above)

mean or low vocabulary knowledge. Text coverage for students with high (i.e.,  $M+1 SD$ ) vocabulary knowledge ranged from 78.27% for the Plastic Duck text to 86.60% for the Mosquito text, while text coverage for students with low vocabulary knowledge (i.e.,  $M-1 SD$ ) was around 42% for each text.

## Text coverage and reading comprehension

### RQ1: shape of the relationship between text coverage and text comprehension

The upper part of Table 4 summarizes the results comparing the baseline model (BLM) to a linear and a quadratic text coverage model. Both the linear and quadratic models had a substantively better fit than the BLM. This was indicated by a much lower *BIC*,  $\Delta_{\text{BLM-TCL}}(\text{BIC})=430$ ,  $\Delta_{\text{BLM-TCQ}}(\text{BIC})=488$ . However, the quadratic text coverage model fit significantly better,  $\chi^2=51.32$ ,  $p<0.001$ , and had the lowest *BIC*  $\Delta_{\text{TCQ-TCL}}(\text{BIC})=42$ . Although the variance explained by the quadratic term was rather small ( $R_0^2\Delta=0.026$ ), the fit indices suggest that the quadratic text coverage model fits better than the linear model.

The model parameters for the quadratic model are provided in Table 5. Both the linear,  $\beta_1=-1.26$ ,  $SE=0.63$ ,  $p=0.046$ , and the quadratic trend,  $\beta_2=3.90$ ,  $SE=0.54$ ,  $p<0.001$ , were significant. The signs of the predictors show that reading comprehension increased with text coverage, but that the effect leveled off in the low text coverage range.

### RQ2: the better predictor for text comprehension

We performed the same analysis including linear and quadratic terms with vocabulary knowledge to determine whether text coverage was a better predictor of text comprehension than vocabulary knowledge. Both vocabulary knowledge models were better than the baseline model. As expected, both vocabulary

**Table 4** Model comparisons between models with linear and quadratic terms explaining the probability of a correct answer in the text comprehension test

Model (i)	$np_i$	$\log(L_i)$	$BIC_i$	$\Delta_i BIC$	$w_i(\text{BIC})^*$	$R_0^2$	$\Delta_i R_0^2$	$\chi^2$	$p$
BLM Baseline	3	-9968	19,965	488	> .001	0			
<i>Text coverage (TC)</i>									
TCL Linear $\beta_1x$	4	-9740	19,519	42	> .001	.518	.026		
TCQ Quadratic $\beta_1x + \beta_2x^2$	5	-9714	19,478	0	> .999	.544	0	51.32	> .001
<i>Vocabulary knowledge (VK)</i>									
VKL Linear $\beta_1x$	4	-9746	19,531	54	< .001	.537	.007		
VKQ Quadratic $\beta_1x + \beta_2x^2$	5	-9744	19,538	60	< .001	.541	.003	3.34	.068

Sample:  $N=924$ , Items:  $I=20$ , Text:  $T=4$ , Observations = 17,932 ( $924 \times 20 = 18,480$ , difference due to 2.97% omitted and not reached responses)

$np_i$  = Number of estimated parameters for model $_i$ ;  $\log(L_i)$  = Natural logarithm of the maximum likelihood for model $_i$ ;  $BIC$  = Bayesian information criterion;  $\Delta_i BIC = [BIC_i - \min(BIC)]$ ;  $w_i(BIC)$  = Rounded Schwarz weights.  $R_0^2$  = Person variance explained by fixed effects obtained by (baseline model  $\sigma_0^2 - \sigma_0^2$  of model  $i$ )/baseline model  $\sigma_0^2$ .  $\chi^2$  and  $p$  were the test statistic and  $p$ -value of the likelihood ratio test comparing nested models

\* $w_i$  can be interpreted as the probability of each model being the best model in a BIC sense among the compared models (Wagenmakers & Farrell, 2004)



**Table 5** Fixed effects of quadratic text coverage and vocabulary knowledge model explaining the probability of correct responses in text comprehension items

	Text coverage			Vocabulary knowledge		
	logit	SE	p	logit	SE	p
$\beta_0$ Intercept <sup>i</sup>	-0.57	0.30	.054	0.33	0.26	.208
<i>Text coverage</i>						
$\beta_1x^c$	-1.26	0.63	.046			
$\beta_2x^2$	3.90	0.54	<.001			
<i>Vocabulary knowledge</i>						
$\beta_1x^z$				0.64	0.03	<.001
$\beta_2x^{2z}$				0.31	0.02	.067
<i>Random variances</i>						
$\sigma^2_{\theta}$	0.34			0.34		
$\sigma^2_{\alpha}$	1.12			1.34		

Participants:  $N_p=924$ , Items:  $N_I=20$ , Text:  $N_T=4$ , Observations = 17,932 ( $924 \times 20 = 18,480$ , difference due to 2.97% omitted responses)

<sup>i</sup>Intercept represents probability of a correct answer with a text coverage of 0% in TCQ and an average vocabulary score in VSL, <sup>c</sup>Text coverage ranges from 0 (0%) to 1 (100%), <sup>z</sup>z-score

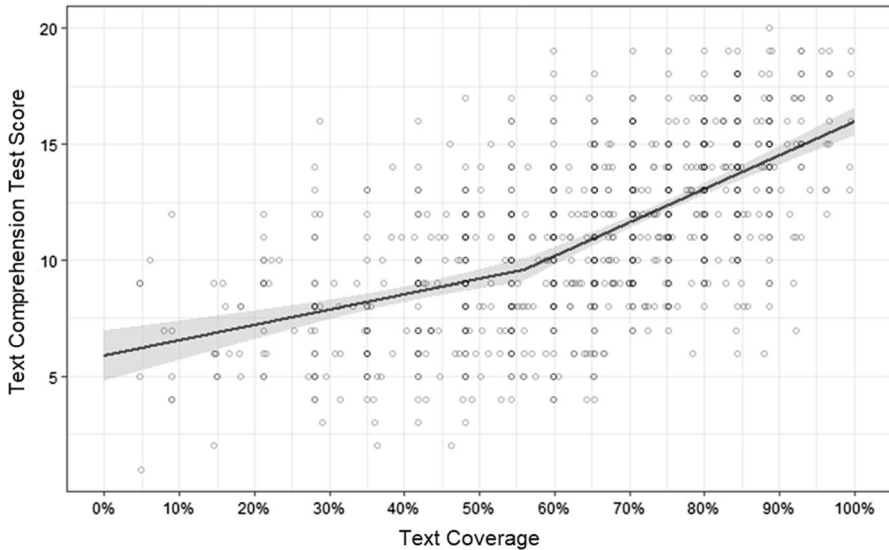
knowledge models explain a significant amount of variance in the responses and thus reading comprehension ability. In contrast to text coverage, the quadratic model for vocabulary knowledge was not substantively better than the linear model,  $\chi^2=3.34, p=0.068$ . This was most clearly indicated by the only marginally lower  $\Delta_{VKQ-VKL}(BIC)=-6$ , suggesting that the gains in explained variance and goodness of fit were due to overfitting. The linear vocabulary model showed a significant linear trend,  $\beta_1=0.63, SE=0.03, p<0.001$  effect, that was in line with previous findings in terms of size and direction (see Table 5).

In a direct comparison of the two text coverage and two vocabulary models, the quadratic text coverage model turned out to be the best model as indicated by the much lower *BIC*.<sup>1</sup> For better interpretability, we calculated the  $w_i$ , which represents the probability of each model being the best out of the five models (Wagenmakers & Farrell, 2004). The  $w_{TCQ}>0.999$  implies that there was an above 99.9% chance that the quadratic text coverage model was the best model among the five. In terms of explained variance, the differences between the quadratic text coverage model and the linear vocabulary knowledge model were small but significant for both outcomes concerning reading comprehension ability, ( $\Delta R^2_{\theta} = 0.011$ ).

**RQ3: amount of text coverage that defines the lexical threshold**

A broken-line regression was significantly better at explaining text comprehension scores than a linear regression,  $F(2, 920)=11.121, p<0.001$ . The broken-line

<sup>1</sup> The comparison between the text coverage and vocabulary knowledge models is only based on BIC because the models were not nested.



**Fig. 8** Relationship between text coverage and comprehension. *Note* Broken-line regression with a changepoint at 56% text coverage. x-axis: Average text coverage estimate for a student across texts. y-axis: Text comprehension test score with a maximum of 20. The grey area is the 95% confidence interval of the expected text comprehension test score. The dots represent the distribution of the test score

regression identified a changepoint at 55.99%, CI[51.59%; 60.39%], meaning that text coverage increases faster above this point than below it. Below the threshold, text comprehension increases at a rate of  $\beta_{1<0.56} = 6.59$ ,  $SE = 1.28$ ,  $p < 0.001$ , and above the threshold, it increases at a rate of  $\beta_{1>0.56} = 7.89$ ,  $SE = 1.68$ ,  $p < 0.001$ . The expected test score at the threshold was 9.56, slightly below the mean test score of  $M = 11.1$ . Thus, the threshold occurs at a mean reading comprehension level (Fig. 8).

## Discussion

In the present study, we investigated the relationship between (1) vocabulary knowledge and reading comprehension, (2) vocabulary knowledge and text coverage, (3) text coverage and text comprehension, as well as associated lexical thresholds.

In line with previous studies, our findings show a strong association between vocabulary knowledge and reading comprehension. As expected, the association between text coverage and comprehension was best described by a non-linear relationship (i.e., exponential or broken-line). Text comprehension increases with text coverage exponentially rather than linearly, and we were able to identify a threshold at 56% text coverage, above which text comprehension increases more rapidly. Overall, text coverage outperformed vocabulary knowledge as a predictor of text comprehension.

Our study provides strong evidence for the view that text coverage and text comprehension are non-linearly related. This is in line with the construction-integration model (Kintsch, 1988) that conceptualizes text comprehension as building an

associative network of links and nodes. A network loses connectivity exponentially when nodes are missing. Hence, there is a certain amount of text coverage that is necessary to activate relevant background knowledge, enable contextual inference and subsequently build comprehension.

Our estimation of the lexical threshold differed from previous studies. Previous studies reporting higher values used self-reports (Hsueh-Chao & Nation, 2000) or word translations (Laufer, 1989). The study with the most comparable research method found the most similar results, but only considered content words (59%; O'Reilly et al., 2019). Due to methodological differences, there is probably no 'one' lexical threshold, as the estimation crucially depends on how word knowledge and text comprehension are assessed. As a consequence, thresholds should only be used within the context in which they were defined.

Our findings also have implications for research and practice. In particular, the relationship between text coverage and text comprehension can be used to determine theoretically and statistically justified cut-off values to inform the selection of adequate learning material and for standard-setting procedures.

Selection of appropriate reading materials could benefit from the text range model and thresholds if text range estimation were implemented in a software tool or within a readability analysis. In most situations, readers benefit most from reading activities that are neither too difficult nor too easy (e.g., Wolfe et al., 1998). Reading activities that are too difficult may be more detrimental to motivation and reading engagement than activities that are too easy (Kahmann et al., 2022). The lexical threshold might be helpful for identifying too little text coverage. It is probably advisable to match readers to texts so that text coverage is above 56%.

A similar application is standard-setting. In the context of educational monitoring, it is often of great interest to determine what test score on a vocabulary test best represents "core vocabulary". Core vocabulary for reading is the vocabulary that allows students to understand a text on a basic level. Core vocabulary has been defined as the vocabulary that covers some (high) percentage of texts in corpora (e.g., Chujo & Utiyama, 2005). Thresholds for core vocabulary are usually defined based on expert ratings or norm values (Brown & Kappes, 2012). However, the text coverage function could be a model-based way to determine these thresholds. Wang et al. (2019) found a decoding threshold that is stable between Grades 5 and 10. They suggested that such thresholds are a function of skill rather than grade or age. According to our model, readers with a vocabulary corresponding to a text coverage above 56% start to gain comprehension. Whether this threshold is consistent across grades need to be investigated in further research.

## Limitations and outlook

One of the major limitations of the present study is that our results are based on a large group of participants but only a small number of texts. Our results should be replicated with a larger and more representative sample of authentic texts, different vocabulary tests and other participants in order to test the generalizability of the reported results.

The difference in explained variance between the non-linear and linear model of the relationship between text coverage and text comprehension was significant but very small. On the one hand, this small effect could still be relevant, because it could improve reader-text matching without requiring more test time and is based on already available information (vocabulary test and text word frequencies). On the other hand, several aspects of this study might have led to a particularly small effect. First, the vocabulary test used in the present study was not ideal for the purpose of estimating students' text coverage. Although it was a widely used, standardized instrument, items were not selected systematically based on their word frequency. Additionally, the test had relatively few items and a relatively low reliability. Future researchers may be advised to use tests that systematically manipulate word frequency and are more reliable (i.e., SET 5-10: subtest "lexicon"; Petermann, 2012). Second, the non-linear relationship between text coverage and text comprehension might have been more pronounced if we had investigated a broader range of texts, including texts with only frequent words or texts with very rare words, and a broader range of students, for instance, second to fifth graders.

There was relatively little precise information about the students' language background. About 35% of the students reported that they did not primarily speak German at home. This group of students could include recently arrived non-native speaking students (about 4% of German fourth graders at the time of the study) or bilingual students. However, we performed robustness checks and used language spoken at home as a mediator of the relationship between text coverage and text comprehension and did not find a significant difference in effects. Thus, the relationship we described is probably generalizable to students with diverse language backgrounds. However, future studies should take a more in-depth look at language background-specific effects.

There are also some theoretical problems that might need to be addressed in the future to further develop this technique. In particular, the present framework does not take into account the context in which words are encountered. Frequent words are usually useful in many contexts, whereas infrequent words are more context-specific. This issue is not addressed in our model. It is also not clear how different psychometric aspects such as measurement error, guessing and slipping influence the text coverage estimation. However, these aspects might primarily influence the absolute text coverage scores.

Despite these limitations, our study demonstrated that text coverage and the lexical threshold are useful concepts that are not yet well established in elementary school reading research and could help to align reading materials with readers and conduct standard-setting. Further research should further refine the method and test whether the thresholds are actually useful.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

## Declarations

**Conflict of interest** We have no known conflict of interest to disclose.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Adams, R. J., & Wu, M. L. (2007). The mixed-coefficients multinomial logit model. A generalized form of the Rasch model. In M. von Davier & C. H. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models: Extensions and applications* (pp. 55–76). Springer. [https://doi.org/10.1007/978-0-387-49839-3\\_4](https://doi.org/10.1007/978-0-387-49839-3_4)
- Artlet, C., Baumert, J., Julius-McElvany, N., & Peschar, J. (2003). Learners for life. Student approaches to learning. *Results from PISA, 2000*, 1–133. <https://doi.org/10.1787/9789264103917-en>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Brown, S., & Kappes, L. (2012). *Implementing the Common Core State Standards: A Primer*. 1–10. Doi: <https://doi.org/10.1177/002205741619600102>
- Brysbaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bølle, J., & Böhl, A. (2011). The word frequency effect. *Experimental Psychology*, 58(5), 412–424.
- Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, 27(1), 45–50. <https://doi.org/10.1177/0963721417727521>
- Burnham, K. P., & Anderson, D. R. (2002). A practical information-theoretic approach. *Model Selection and Multimodel Inference*. <https://doi.org/10.1007/b97636>
- Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology*, 96(4), 671–681. <https://doi.org/10.1037/0022-0663.96.4.671>
- Chujo, K., & Utiyama, M. (2005). Understanding the role of text length, sample size and vocabulary size in determining text coverage. *Reading in a Foreign Language*, 17(1), 1–22. <https://files.eric.ed.gov/fulltext/EJ689119.pdf>.
- De Boeck, P., & Wilson, M. (2004). *Explanatory item response models: A generalized linear and non-linear approach*. Springer Science & Business Media. <https://doi.org/10.1007/978-1-4757-3990-9>
- Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first- and second-language word recognition: A lexical entrenchment account. *Quarterly Journal of Experimental Psychology*, 66(5), 843–863. <https://doi.org/10.1080/17470218.2012.720994>
- Duke, N. K., & Cartwright, K. B. (2021). The science of reading progress: Communicating advances beyond the simple view of reading. *Reading Research Quarterly*, 56, 25–S44. <https://doi.org/10.1002/rrq.411>
- Ehri, L. C. (2014). Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning. *Scientific Studies of Reading*, 18(1), 5–21. <https://doi.org/10.1080/10888438.2013.819356>
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, 7, 6–10.
- Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Matrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, and Computers*, 36(2), 193–202. <https://doi.org/10.3758/BF03195564>
- Gustafsson, J. E. (1980). Testing and obtaining fit of data to the Rasch model. *British Journal of Mathematical and Statistical Psychology*, 33(2), 205–233. <https://doi.org/10.1111/j.2044-8317.1980.tb00609.x>

- Heller, K. A., & Perleth, C. (2000). Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision: KFT 4–12+ R. Beltz-Test.
- Hsueh-Chao, M. H., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, 13(1), 403–430. <https://doi.org/10.26686/wgtn.12560354>
- Karageorgos, P., Richter, T., Haffmans, M. B., Schindler, J., & Naumann, J. (2020). The role of word-recognition accuracy in the development of word-recognition speed and reading comprehension in primary school: A longitudinal examination. *Cognitive Development*, 56, 100949. <https://doi.org/10.1016/j.cogdev.2020.100949>
- Kahmann, R., Ozuer, Y., & Zedelius, C.M. et al. (2022). Mind wandering increases linearly with text difficulty. *Psychological Research*, 86, 284–293. <https://doi.org/10.1007/s00426-021-01483-9>.
- Kim, Y. S. G. (2020). Hierarchical and dynamic relations of language and cognitive skills to reading comprehension: Testing the direct and indirect effects model of reading (DIER). *Journal of Educational Psychology*, 112(4), 667–684. <https://doi.org/10.1037/edu0000407>
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95, 163–182. <https://doi.org/10.1037/0033-295X.95.2.163>
- Lauffer, B. (1989). What percentage of text-lexis is essential for comprehension. In *Special language: From Humans Thinking to Thinking Machines* (pp. 316–323). <https://www.jstor.org/stable/43267941>
- Lauffer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, 22(1), 15–30.
- Lehmann, R. H., Peek, R., Gänsfuß, R., & Husfeldt, V. (2002). *LAU 9: Aspekte der Lernausgangslage und der Lernentwicklung; Klassenstufe 9; Ergebnisse einer längsschnittlichen Untersuchung in Hamburg. Behörde für Bildung und Sport. [LAU 9: Aspects of learning background and learning development; grade 9; results of a longitudinal study in Hamburg. Authority for Education and Sports.]* (pp. 1–467). Waxmann.
- Li, M., & Kirby, J. R. (2015). The effects of vocabulary breadth and depth on english reading. *Applied Linguistics*, 36(5), 611–634. <https://doi.org/10.1093/applin/amu007>
- Muggeo, V. M. (2008). Segmented: an R package to fit regression models with broken-line relationships. *R News*, 8(1), 20–25.
- Nagy, W., & Townsend, D. (2012). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly*, 47(1), 91–108. <https://doi.org/10.1002/RRQ.011>
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133–142. <https://doi.org/10.1111/j.2041-210x.2012.00261.x>
- Nation, P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. *Vocabulary: Description, Acquisition and Pedagogy*, 14, 6–19.
- O'Reilly, T., Wang, Z., & Sabatini, J. (2019). How much knowledge is too little? When a lack of knowledge becomes a barrier to comprehension. *Psychological Science*, 30(9), 1344–1351. <https://doi.org/10.1177/0956797619862276>
- Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, 98(3), 554–566. <https://doi.org/10.1037/0022-0663.98.3.554>
- 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. A., & Hart, L. (2002). The lexical quality hypothesis. *Precursors of Functional Literacy*, 11, 67–86. <https://doi.org/10.1075/swll.11.14per>
- Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin & Review*, 21(5), 1112–1130. <https://doi.org/10.3758/s13423-014-0585-6>
- McElvany, N., Becker, M., & Lüdtke, O. (2009). Die Bedeutung familiärer Merkmale für Lesekompetenz, Wortschatz, Lesemotivation und Leseverhalten [The importance of family characteristics for reading literacy, vocabulary, reading motivation, and reading behavior]. *Zeitschrift Für Entwicklungspsychologie Und Pädagogische Psychologie [Journal of Developmental and Educational Psychology]*, 41(3), 121–131.
- Petermann, F. (2012). *Sprachstandserhebungstest für Kinder im Alter zwischen 5 und 10 Jahren [Language proficiency test for children between the ages of 5 and 10 years old]*: SET 5-10 (Vol. 3). Göttingen: Hogrefe

- Quinn, J. M., Wagner, R. K., Petscher, Y., & Lopez, D. (2015). Developmental relations between vocabulary knowledge and reading comprehension: A latent change score modeling study. *Child Development*, 86(1), 159–175. <https://doi.org/10.1111/cdev.12292>
- R Core Team (2021). Writing R extensions. R foundation for statistical computing.
- Richter, T., & Schnotz, W. (2018). Textverstehen [Text Comprehension]. In S. Buch, D. Rost & J. Sparfeldt (Hrsg.), *Handwörterbuch Pädagogische Psychologie [Hand dictionary educational psychology]* (4. Aufl., S. 826–837). Beltz. <https://doi.org/10.2378/peu2019.art16d>
- Richter, T., Isberner, M. B., Naumann, J., & Neeb, Y. (2013). Lexical quality and reading comprehension in elementary school children. *Scientific Studies of Reading*, 17(6), 415–434.
- Robitzsch, A., Kiefer, T., & Wu, M. (2021). *TAM: Test analysis modules*. R package version 3.7-16. <https://CRAN.R-project.org/package=TAM>.
- Robitzsch, A., Kiefer, T., & Wu, M. (2022). *TAM: Test Analysis Modules*. R package version 4.1-4. <https://CRAN.R-project.org/package=TAM>.
- 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. <https://doi.org/10.1111/j.1540-4781.2011.01146.x>.
- Stanat, P., Schipolowski, S., Rjosk, C., Weirich, S., & Haag, N. (2017). *IQB-Bildungstrend 2016. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Klasse. [IQB Education Trend 2016. Competencies in German and mathematics at the end of grade 4.]* Waxmann. <https://doi.org/10.25656/01:15477>
- Schneider, W. (2011). WLLP-R: Würzburger Leise Leseprobe-Revision: ein Gruppentest für die Grundschule. Hogrefe Verlag für Psychologie.
- Share, D. L., & Stanovich, K. E. (1995). Cognitive processes in early reading development: Accommodating individual differences into a model of acquisition. *Issues in Education: Contributions from Educational Psychology*, 1, 1–57.
- Trautwein, J., & Schroeder, S. (2018). WOR-TE: Ein Ja/Nein-Wortschatztest für Kinder verschiedener Altersgruppen [WOR-TE: A yes/no vocabulary test for children of different age groups]. *Diagnostica*, 65, 37–48. <https://doi.org/10.1026/0012-1924/a000212>.
- Van Heuven, W. J., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British english. *Quarterly Journal of Experimental Psychology*, 67(6), 1176–1190. <https://doi.org/10.1080/17470218.2013.850521>
- Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin and Review*, 11(1), 192–196. <https://doi.org/10.3758/bf03206482>
- 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.
- Wolfe, M. B., Schreiner, M. E., Rehder, B., Laham, D., Foltz, P. W., Kintsch, W., & Landauer, T. K. (1998). Learning from text: Matching readers and texts by latent semantic analysis. *Discourse Processes*, 25(2–3), 309–336. <https://doi.org/10.1080/01638539809545030>