The Associations between Child and Item Characteristics, Use of Vocabulary Scaffolds, and Reading Comprehension in a Digital Environment: Insights from a Big Data Approach

Laura Diprossimo[1], Anastasia Ushakova[2], Jennifer Zoski[3], Harrison Gamble[3], Robin Irey[4] and Kate Cain[1]

[1]Department of Psychology, Lancaster University

[2]Centre for Health Informatics, Computing, and Statistics, Lancaster University

[3]Amplify Education, Inc

[4]Department of Education, Stanford University

Accepted for publication in Contemporary Educational Psychology 06/02/2023

Author Note

Declaration of interest: Laura Diprossimo, Anastasia Ushakova, Robin Irey and Kate Cain have no competing interests. However, Jennifer Zoski and Harrison Gamble are employed by Amplify Education, Inc, the company which developed the software and features under

investigation. Anastasia Ushakova was supported by philanthropic funding gifted to Kate Cain from a private donor. As a group, we took all the reasonable steps to mitigate potential bias in the analysis, reporting and interpretation. Before accessing the dataset, the academic authors Laura Diprossimo, Anastasia Ushakova and Kate Cain conceptualized and preregistered the research questions and analytical plan on the Open Science Framework (Diprossimo, L., Cain, K., & Ushakova, A. (2021, June 9). Vocabulary Scaffolding Features and Young Readers' Comprehension of Digital Text: Insights from a Big Observational Dataset. https://doi.org/10.17605/OSF.IO/62C4Q).

Correspondence concerning this article should be addressed to Laura Diprossimo, Department of Psychology, Lancaster University, Lancaster, LA1 4YF, UK. l.diprossimo@lancaster.ac.uk

Abstract

Scaffolding features that provide multimodal support for the pronunciation and meaning of words are increasingly common in digital reading environments. These vocabulary scaffolds are intended to aid the accurate pronunciation and understanding of individual words in context, thus supporting both vocabulary development and comprehension of text. However, the evidence on their efficacy remains inconclusive. The present study adds to the evidence base by examining: 1) whether child characteristics predict the use of vocabulary scaffolds; 2) whether the use of vocabulary scaffolds is associated with reading comprehension performance; and 3) whether the association between the use of scaffolds and reading comprehension is modulated by child and/or item characteristics. A large cohort (N ~ 120,000) of 5- to 8-year-old children in the United States interacted with a gamified digital reading environment with embedded vocabulary scaffolds, thereby generating a large observational dataset of user log files. Confirmatory analyses with Generalized Linear Mixed Models (GLMMs) indicated that children with lower literacy skills, beginning readers, girls, and bilingual students were more likely to use the scaffold. Overall, the use of scaffolds was associated with better reading comprehension performance. The association between the use of scaffolds and reading comprehension was modulated by both child and item characteristics. We conclude that vocabulary scaffolds may be promising tools to facilitate reading comprehension and reduce performance differences amongst diverse learners in digital reading environments. Educational implications and recommendations for future research are discussed.

Keywords: vocabulary, reading comprehension, scaffolding features, multimedia learning, big data

The Associations between Child and Item Characteristics, Use of Vocabulary Scaffolds, and Reading Comprehension in a Digital Environment: Insights from a Big Data Approach

Digital texts afford novel opportunities to scaffold children's reading comprehension (Dalton et al., 2011; Gonzalez, 2014; Proctor et al., 2007; Zou et al., 2021). They provide a critical opportunity for learning through in-built scaffolding features that offer multimodal support (i.e., auditory and visual) for the pronunciation and meaning of words. These features, also referred to in the literature as hypertext glosses (Yun, 2011), dictionaries (Furenes et al., 2021), annotations (Akbulut, 2007) and vocabulary supports (Gonzalez, 2014; Proctor et al., 2007), are intended to support the accurate pronunciation and understanding of individual words in context, potentially promoting both vocabulary development and comprehension of the text. To date, evidence supporting the benefits of digital scaffolds remains elusive, and a thorough evaluation of these features is lacking. First, little is known about which child characteristics predict the use of such scaffolds when they are available to the reader on an as-needed basis. Second, the association between use of scaffolds and reading comprehension remains unclear, with mixed evidence arising potentially from diverse participants and item characteristics across studies (Abraham, 2008; Furenes et al., 2021).

Our study was designed to address these gaps in our knowledge to provide a detailed and critical evaluation of the associations between child and item characteristics, use of vocabulary scaffolds, and reading comprehension in a digital environment. We exploited a large observational dataset of children's interactions with a real-world digital reading product that was supplementary to the established classroom curriculum. Our analyses identified critical factors that were associated with the use of scaffolds, and which modulated the associations between the use of scaffolds and reading comprehension.

Scaffolding Principle

A scaffold was originally defined as a process that enables a child or novice to solve a problem, carry out a task, or achieve a goal that is beyond their unassisted efforts (Wood et al., 1976). Contemporary definitions of scaffolds are, however, more nuanced and include several tools and aids. Scaffolds have been investigated across a wide range of learning contexts, including instruction in narrative skills (Pesco & Gagné, 2017), problem-based learning for science, technology, engineering and mathematics education (N. J. Kim et al., 2018), and simulation-based learning in medical or teacher training (Chernikova et al., 2020). Scaffolds have been found to benefit learning in different age groups, including children (Pesco & Gagné, 2017) and adults (Chernikova et al., 2020).

With respect to reading comprehension, instructional scaffolding represents a flexible and adaptable model (Clark & Graves, 2005). A scaffolded reading experience may involve a range of activities that can take place before, during, or after reading. These activities include pre-teaching vocabulary, relating the content to children's lives, asking questions and discussing the text (Clark & Graves, 2005). These scaffolds have all proved beneficial for reading comprehension (Blything et al., 2020; Degener & Berne, 2017; Elleman et al., 2009; McKeown et al., 2009; Zucker et al., 2010). In relation to the specific focus of this study, a meta-analysis evaluating the effect of vocabulary instruction on passage-level comprehension found a positive impact for classroom-based instruction focused on increasing word knowledge (Elleman et al., 2009). This provides strong evidence for the potential benefit of scaffolds for vocabulary items to support understanding of text.

Beyond traditional instructional scaffolding, technology-based scaffolding is becoming increasingly common. Digital environments, and consequently digital texts, afford novel

opportunities to scaffold children's reading comprehension (Dalton et al., 2011; Gonzalez, 2014; Proctor et al., 2007; Zou et al., 2021). Indeed, the medium lends itself to support independent reading with built-in scaffolds. For instance, empirical work has shown that text-to-speech read aloud functionality scaffolds text comprehension for grade 3 and 4 students with disabilities (Gonzalez, 2014). Moreover, an intervention study found that the use of scaffolds that provide comprehension strategies embedded in digital texts positively relates to reading comprehension gains in 4th graders (Proctor et al., 2007). Importantly, digital environments often incorporate gamification and adaptivity principles, which are both powerful tools for promoting motivation and learning (Cohen et al., 1982; Lämsä et al., 2018; Manzano-León et al., 2021). Despite the promise of digital supplements in education, there is consensus that specific features need to be evaluated carefully in relation to learning outcomes to disentangle which features works best and for whom (Lämsä et al., 2018; Manzano-León et al., 2021; Mayer, 2020). Our focus here is on the evaluation of scaffolding features which were designed to promote understanding of single words in context and embedded in a gamified digital reading environment. The scaffolds included written definitions, and auditory and visual support for pronunciation and meaning of target words, which could be used on an as-needed basis. A comprehensive evaluation of these features is crucial given the strong and reciprocal relationship between vocabulary knowledge and reading comprehension (Cain & Oakhill, 2018).

Vocabulary Knowledge and Reading Comprehension

Vocabulary knowledge is a unique predictor of word reading in kindergarteners (N. J. Kim et al., 2014), and reading comprehension in later grades (Oakhill & Cain, 2012; Quinn et al., 2015). As noted above, a meta-analysis exploring the impact of vocabulary instruction on passage-level comprehension for school-age children found a positive impact for this type of

scaffold, and indicated that the impact was three times greater for students with reading difficulties (Elleman et al., 2009). Thus, vocabulary instruction is viewed as a crucial ingredient of interventions targeting reading fluency and reading comprehension (Wolf & Katzir-Cohen, 2009). Importantly, vocabulary development and reading comprehension, and its core skills such as inference, are interdependent (Cain & Oakhill, 2018; Language and Reading Research Consortium (LARRC) et al., 2019): vocabulary knowledge is both foundational to reading comprehension, and also boosted by reading practice, presumably because good text comprehension skill enables readers to derive the meanings of unfamiliar words from context (Cain et al., 2003). Written text indeed provides a rich source of exposure to a range of less frequent and less familiar words in context (Cunningham & Stanovich, 1998; Montag et al., 2015).

Vocabulary Scaffolding Features

Given both the strong relationship between vocabulary and reading comprehension (Cain & Oakhill, 2018), and written text itself as a rich source of new word learning (Cunningham & Stanovich, 1998; Montag et al., 2015), there is a need to understand the role of vocabulary scaffolds in digital texts. These scaffolds have proven beneficial for vocabulary learning for struggling readers and second language learners (Liu & Leveridge, 2017; Proctor et al., 2007; Yun, 2011), as well as for young children in the context of shared reading (Furenes et al., 2021). Indeed, inferring the meanings of novel words from linguistic contexts may be particularly challenging for children with poor reading comprehension (Cain et al., 2003) and for second language learners (Nassaji, 2003).

The positive effect of vocabulary scaffolding features that combine pictorial support and definitions on word learning aligns with the wealth of research documenting the benefits of

studying vocabulary with verbal explanations and images as compared to studying with only verbal explanations (Akbulut, 2007; Andrä et al., 2020; Hald et al., 2016; D. Kim & Gilman, 2008; Rowe et al., 2013; Tonzar et al., 2009). This so-called picture superiority effect in vocabulary acquisition has been explained by the dual coding theory (Paivio & Csapo, 1973), which posits that image and verbal memory codes are independent and have an additive effect on recall. Similarly, the cognitive theory of multimedia learning explains the picture superiority effect as a multimedia effect, in which students learn more deeply from words and pictures than from words alone (Mayer, 2017; Mayer, 2003; Mayer & Moreno, 2003; Mayer & Sims, 1994).

Evidence documenting the impact of vocabulary scaffolds on reading comprehension is mixed. Whilst some meta-analytic evidence suggests that vocabulary scaffolds have an overall medium size effect on second language reading comprehension in university students (Abraham, 2008), other meta-analytic research shows that dictionaries embedded in digital storybooks aid word learning, but have a negligible or negative impact on story comprehension in children up to 8 years of age (Furenes et al., 2021). However, these results refer to a small subsample (n = 5) of the studies included in this meta-analysis. Indeed, another study of children aged 4 to 6 (Korat & Shamir, 2012), not included in the synthesis by Furenes and colleagues (2021), found a significant and positive association between story comprehension and gains in knowledge and decoding of words that received direct computer support (i.e., support was given a priori to all children for target words) but did not find gains in knowledge and decoding of words that did not receive the computer supports. This finding suggests a positive association between the provision of vocabulary scaffolds and young children's story comprehension in digital environments. Studies reporting a positive association between use of vocabulary scaffolding features and reading comprehension are in line with both dual coding theory (Paivio & Csapo, 1973) and the

cognitive theory of multimedia learning (Mayer, 2017; Mayer, 2003; Mayer & Moreno, 2003; Mayer & Sims, 1994), and also with research involving older children that has demonstrated the benefits on reading comprehension of training children to use mental imagery (Francey & Cain, 2015; Joffe et al., 2007). The inconsistent findings across studies potentially arise from diverse child and item characteristics. To date, there is a dearth of studies that have systematically investigated the associations between use of vocabulary scaffolds and young readers' comprehension of short digital text during independent reading, while accounting for both child and item characteristics. This is an innovative feature of our study.

Child Characteristics

Vocabulary scaffolds can be made available to the learner on an as-needed basis. This approach assumes that readers are aware of their vocabulary and comprehension support needs as they progress through the text. Such metacognitive awareness is likely to vary as a function of reader characteristics, such as ability level and age (Baker & Cerro, 2000; Kirby & Moore, 1987), gender (Sadeghi & Khezrlou, 2012; Tseng et al., 2006; Wu, 2014), and degree of bilingualism (Abu Rabia, 2019). To the best of our knowledge, research on individual differences in the selection of vocabulary scaffolds embedded in digital texts is lacking. However, reading skills predict 11-year-old students' selection of cohesive, semantically related hyperlinks (Salmerón & García, 2011), suggesting that literacy skills may influence how young readers go about navigating and searching for information in digital texts. With respect to gender, it has been shown that female students between 11 to 25 years employ more self-regulated strategies in vocabulary learning when reading texts compared to male students (Sadeghi & Khezrlou, 2012). Additionally, girls aged 5 to 7 years have been observed to outperform boys on a task designed to assess novel word learning, but only when learning

phonologically or semantically familiar information (Kaushanskaya et al., 2013), suggesting that girls use different strategies compared to boys when learning novel words. Girls also outperform boys in knowledge of metacognitive strategies and navigation skills when engaged in online reading (Wu, 2014). Regarding language status, bilingualism is also associated with increased use of metacognitive reading strategies (Abu Rabia, 2019).

Child characteristics may also modulate the associations between use of vocabulary scaffolds and reading comprehension. This association may depend on both the child's background knowledge and their use of strategies. For example, meta-analytic evidence documenting the process of acquiring novel words in a foreign language suggests that the impact of scaffolds is greater for beginner learners than advanced learners (Yun, 2011). To date, little is known about which child characteristics predict the use of scaffolding features and modulate the associations between use of vocabulary scaffolds and reading comprehension. This information is essential to understand who can benefit most from this feature, and who may need additional support.

Item Characteristics

Another dimension to consider is that word class, as well as the inherent psycholinguistic properties of the scaffolded words, may pose different challenges to learners, and thus modulate the associations between use of vocabulary scaffolds and reading comprehension. Studies on early language acquisition suggest that novel nouns are learnt more easily than novel verbs (Childers & Tomasello, 2002; Gentner, 1982; Goldin-Meadow et al., 1976) and the same advantage for nouns is apparent in second language acquisition (Ellis & Beaton, 1993). Nevertheless, the concept of noun-bias has been challenged: Fourth graders demonstrate a learning advantage for non-nouns (i.e., verbs, adjective and adverbs) compared to nouns during

story reading (Schwanenflugel et al., 1997). These mixed findings emphasize the need to take word class into account when examining the associations between use of vocabulary scaffolds and reading comprehension. Psycholinguistic properties of the words may also influence how easily they are learned. For instance, concreteness influences children's ability to learn words (Hadley et al., 2016). Moreover, the frequency of a word in child-directed speech and imageability ratings are both important predictors of lexical development (Hansen, 2017).

Present Study

The research evidence reviewed suggests that vocabulary scaffolds embedded in digital texts can play an important role in word learning. However, the strength of the association between use of scaffolds and reading comprehension remains unclear, with mixed findings arising potentially from diverse participant and item characteristics. The current study was designed to address these gaps in our knowledge using secondary data analysis to answer the following critical research questions:

1) Do child characteristics predict the use of vocabulary scaffolds?

2) Is the use of vocabulary scaffolds associated with reading comprehension?

3) Is the association between use of scaffolds and reading comprehension conditional on child and/or item characteristics?

We examined the performance of a large cohort of young readers in the United States in a digital reading environment. During a gamified comprehension task, students acted as a newspaper editor and judged whether picture-text pairs matched in meaning or not. The task was designed to promote interest level, which has been shown to enhance comprehension monitoring in children with poor reading comprehension (De Sousa & Oakhill, 1996). Throughout the game, vocabulary scaffolds for target words were signalled by underlining. The student could select the

word to get audio and visual support for its pronunciation and meaning. Our analysis examined the use of the scaffolds and the associations between use of scaffolds and reading comprehension as evidenced by performance on the sense-matching task, across a range of child and item characteristics.

Research Question 1: Do child characteristics predict the use of vocabulary scaffolds?

Our first research question investigated child characteristics as predictors of scaffold use. We examined four child characteristics: Literacy skill, grade (as a proxy for chronological age), gender, and language status (monolingual vs bilingual students). Whilst previous research on hyperlink selection strategies shows that reading ability influences how children search information and navigate digital texts (Salmerón & García, 2011), the reading environment and type of scaffold in that study differed substantially from our own, so we did not make directional predictions about the influence of reading skills and age on scaffold use. With respect to gender, however, we predicted that girls would be more likely to use scaffolds, because previous research has shown that girls are more likely to use self-regulated strategies in vocabulary learning (Sadeghi & Khezrlou, 2012), and have better knowledge of metacognitive strategies and navigation skills in online reading (Wu, 2014). Regarding language status, we hypothesized that bilingual students would make greater use of scaffolds, since bilingualism is associated with an increased use of metacognitive reading strategies (Abu Rabia, 2019).

Research Question 2: Is the use of vocabulary scaffolds associated with reading comprehension?

Our second research question investigated the associations between use of vocabulary scaffolds and reading comprehension. We predicted that scaffold use would be associated with greater item accuracy (i.e., reading comprehension), given the strong relationship between

vocabulary and reading comprehension (Cain et al., 2003) and previous meta-analytic evidence (Abraham, 2008). However, we note the findings of a recent meta-analysis indicating that an embedded dictionary aided word learning had a negligible or negative impact on story comprehension (Furenes et al., 2021).

Research Question 3: Is the association between use of scaffolds and reading comprehension conditional on child and/or item characteristics?

Our third research question addressed whether the associations between use of vocabulary scaffolds and reading comprehension was modulated by child and/or item characteristics. Our child characteristics of interest were early literacy skills, grade, gender, and language status. Yun's (2011) meta-analysis showed that the impact of similar scaffolds was greater in beginner as opposed to more advanced adult learners, during the acquisition of novel words in a foreign language. If those findings generalize to children, and to the domain of reading comprehension, the associations between scaffold use and reading comprehension should be positive and greater for children with lower literacy skills and beginner readers. Due to a lack of previous studies in this area examining the influence of gender and language status, we did not make specific predictions about how gender and language status might modulate the associations between use of vocabulary scaffolds and reading comprehension.

Lastly, we hypothesize that the associations between use of vocabulary scaffolds and reading comprehension would differ by word type, because they pose different challenges to learners (Ellis & Beaton, 1993; Gentner, 1982; Goldin-Meadow et al., 1976; Schwanenflugel et al., 1997). Specifically, we predicted that the associations between use of vocabulary scaffolds and reading comprehension would be positive and stronger for nouns than for other word types, due to research demonstrating that nouns are easier to learn (Childers & Tomasello, 2002), and

also because nouns are typically more concrete (Goldin-Meadow et al., 1976), making them easier to depict and define with the type of scaffold under investigation.

Methods

To address our research questions, we exploited an existing large observational dataset. The dataset was supplied through a project-specific data sharing agreement with *Amplify Education,* a private company dedicated to the development and commercialization of digital educational tools. The dataset consists of behavioural/usage data (i.e., user log files) relative to the *Reveal Words* vocabulary scaffolding feature embedded in the *Read All About It* reading game. The data procured was for students who played the game between 01/09/2020 and 28/04/21, as part of a broader digital reading supplement, Amplify Reading. This study has received Ethical approval from the Faculty of Science and Technology, Lancaster University (reference number: FST-2022-1098-RECR-5). Our analysis plan was preregistered and is available on the Open Science Framework (OSF): Diprossimo, L., Cain, K., & Ushakova, A. (2021, June 9). Vocabulary Scaffolding Features and Young Readers' Comprehension of Digital Text: Insights from a Big Observational Dataset. https://doi.org/10.17605/OSF.IO/62C4Q.

Participants

Children across the U.S., primarily based in Southern California and Texas, played the *Read All About It* game as part of a broader digital reading supplement. Observations with missing data on key variables of interest (i.e., literacy skills, grade level[1], gender[2], and language status) were removed. We then randomly sampled within the large observational dataset to

[1] *Grade level* was our proxy for chronological age, as precise information on chronological age was not available in the anonymised dataset.

[2] Here we use the term *gender* in line with the previous literature and to acknowledge the fact that literacy is embedded in a socio-cultural context. Strictly speaking however, we classed our participants on reported biological sex.

reduce it to half of all cases per each game item/use of scaffold combination. This was done to save computational time while preserving the dataset's inherent characteristics. With large enough samples, there is a certain point after which increasing the sample size will not affect the results but may bring additional computational challenges. After subsampling, our sample was still much larger (N ~120,000) than those commonly reported in the literature. English monolingual and bilingual students[3] aged between 5 and 8 years were included. The data set used for this analysis included only student data from classroom use of *Amplify Reading*. No personally identifiable information (for either children, schools, or school districts) was included in the dataset that was analysed.

Early literacy skills were measured for each child with the Dynamic Indicators of Basic Early Literacy Skills (DIBELS), 8th edition, (University of Oregon, Center on Teaching and Learning, 2018). DIBELS is a set of short, standardized assessments that can be used for universal screening, benchmark assessment, and progress monitoring. DIBELS was administered by school teachers to evaluate letter naming fluency, phonemic segmentation, nonsense word fluency, word reading fluency, oral reading fluency, and reading comprehension (using a maze – or cloze – task). The scores of the different DIBELS subtests enabled the classification of students into four composite performance levels: *well below benchmark*, *below benchmark*, *at benchmark*, and *above benchmark*. Our analysis employed composite performance levels as an overall measure of literacy skills. The composition of our sample is summarized in Table 1.

[3] Here we use the term *bilingual* in an inclusive way, encompassing a variety of types of bilingualism. Our dataset does not permit distinction between, for example, simultaneous or sequential bilingualism. Examination of these more fine-grained distinctions were, therefore, beyond the scope of this study.

Table 1

Descriptives for the Key Variables of Interest in the Study Sample: Literacy Skills, Grade, Gender, and Language Status. Information on Reported Race is also Provided

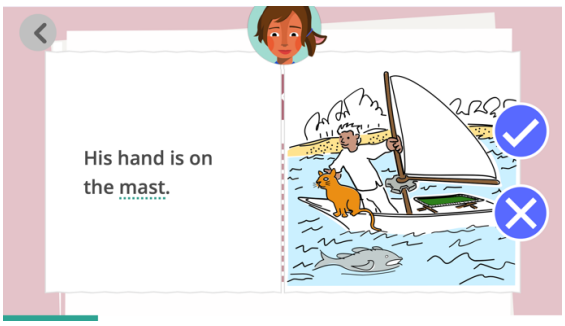
		Count	Proportion (%)
Literacy skills (DIBELS)	Above Benchmark	26956	23.1
	At Benchmark	29024	24.87
	Below Benchmark	18163	15.56
	Well Below Benchmark	42570	36.47
Grade	K	21369	18.31
	1	32673	27.99
	2	41775	35.79
	3	20896	17.9
Gender	Male student	59019	50.57
	Female student	57694	49.43
Language status	English monolingual	90636	77.66
	Bilingual student	26077	22.34
Race	Black or African American	18471	15.83
	White	22813	19.55
	Hispanic or Latino	61884	53.02
	Asian	5256	4.5
	Multiracial/other	2387	2.05
	Native Hawaiian, Pacific Islander, American Indian, Alaskan Native	2386	2.04
	Not available	3516	3.01

Learning Materials

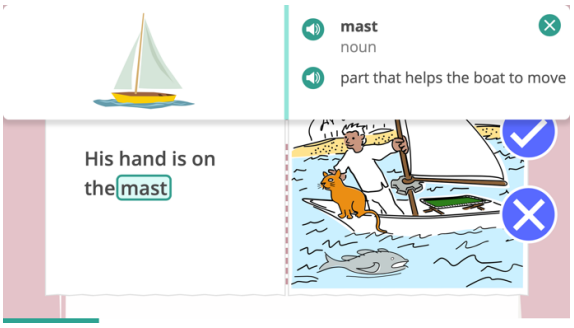
Read All About It is a reading game designed to foster automaticity in reading. In a gamified comprehension task, students are presented with text-picture pairs that assess understanding of a target item. They are encouraged to act as newspaper editors and judge whether the text matches the picture by selecting (i.e., mouse clicking or tapping) either the correct (i.e., √) or incorrect (i.e., X) option (Figure 1a). The decodable text includes the sound-spelling correspondences, word features (e.g., prefixes/suffixes), and phonics rules that they have learned and practised in other games. The game is introduced after students have been taught 10 individual letter sounds. Content builds as students learn new letter-sound correspondences, letter combinations, and phonics rules through multisyllabic decoding. Content difficulty increases as the child progresses to more advanced levels. Each child is initially assigned to a game level based on their reading ability, and provisionally removed from a game level if they repeatedly fail; this is done to minimize their frustration.

Figure 1

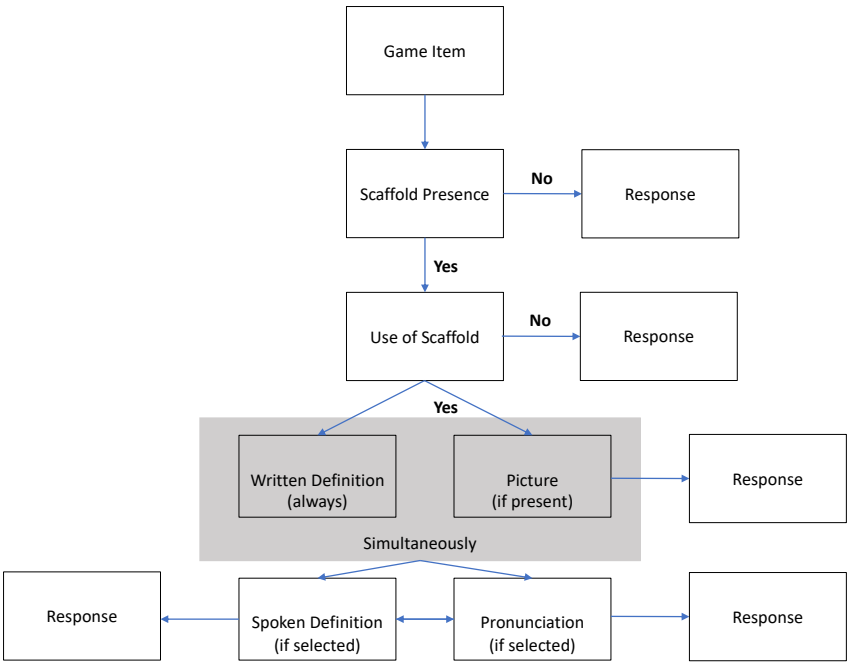
Example of Game Item, Scaffolding Feature and Child's Decision Tree



a)



b)



c)

Note. (a) Gamified reading comprehension item, (b), *Reveal Word* scaffolding feature for the word "mast", (c) child's decision tree while navigating an item in the game.

Throughout the game, the *Reveal Words* scaffold (Figure 1b) appears with the target word underlined. The student may select this word to get immediate support for its meaning via written definition and pictorial support, which appears simultaneously. We note however that pictorial support was not provided for a small proportion of scaffolded words[4]. Auditory support for pronunciation and definition were available by further clicking on the speaker symbols (see Figure 1c for a sketch). *Reveal Words* scaffolds were available for a range of word types (i.e., nouns, adjective, adverbs, verbs, and prepositions)[5]. The game comprises a total of 54 game levels, 505 game items, and 581 *Reveal Words* scaffolds. The type and number of *Reveal Words* scaffolds varied across game items. The words that received *Reveal Words* support were selected based on the Biemiller ratings (Biemiller, 2010), playtesting feedback, and the English Vocabulary Profile (www.englishprofile.org). The full list of scaffolded words is available in the Supplementary Materials.

Procedures

Children played the *Read All About It* game via a tablet or a PC in school settings. The data collection was a continuous process, as data were collected iteratively, every time a child logged into the game. Log files provided information about usage, such as *Reveal Words* feature selection, and accuracy for the comprehension task. Each log file was associated with a unique identification code for each child. Early literacy skills, grade, gender, and language status were also encoded in the log files. A codebook with full list of information available in the log files can be consulted in the Supplementary Materials.

[4] Overall, pictorial support was available for 71% of the scaffolds. We return to this point in our exploratory analysis.

[5] The scaffold was available for 311 nouns, 59 adjectives, 20 adverbs, 188 verbs, and 2 prepositions. Pictorial support was provided for 77% of nouns, 63% of adjective, 25% of adverbs, 69% verbs and 100% of prepositions.

Analytical Approach

Statistical Models

Research Question 1: Do child characteristics predict the use of vocabulary scaffolds? To investigate which child characteristics influenced the probability of using at least one scaffold per game item, we fitted a Generalized Linear Mixed Model (GLMM; Baayen 2008) with binomial error structure and logit link function (McCullagh & Nelder, 1989). Our binary outcome was the use of the scaffold. We included early literacy skills as measured by DIBELS, grade level (as a proxy for age), gender, and language status, all as fixed effects; we further included game item and level as random effects to account for potential differences arising from individual game items and levels.

Research Question 2: Is the use of vocabulary scaffolds associated with reading comprehension? Similar to the first analysis, we fitted a Generalized Linear Mixed Model (GLMM; Baayen 2008) with binomial error structure and logit link function (McCullagh & Nelder, 1989) to investigate what influenced the probability of being correct in each game item. Here our binary outcome was the accuracy in each game item. We included the use of at least one scaffold per game item as a test predictor, and literacy skills as measured by DIBELS, grade, gender, and language status, as control variables. As before, we controlled for the random effect of each game item and level to account for potential differences arising from individual game items and levels.

Research Question 3: Is the association between use of scaffolds and reading comprehension conditional on child and/or item characteristics? We further expanded our previous model by including the interactions between the use of the scaffold and each of the child characteristics of interest (i.e., early literacy skills, grade, gender, and language status) as

well as the number of scaffolded nouns, adjectives, adverbs, verbs, and prepositions as test predictors. As before, we controlled for the random effect of each game item and level to account for potential differences arising from individual game items and levels.

Model Selection

Model selection was guided by the results of the incremental likelihood ratio test (Matuschek et al., 2017), which indicates whether a specific term significantly improves the model fit. The selection process started at the baseline model (a model lacking our test predictor), adding a single predictor at a time until we reached the full model, as indicated in our pre-registration (https://doi.org/10.17605/OSF.IO/62C4Q).

Inference Criteria

To evaluate the significance of the contribution of our predictors. whilst avoiding multiple testing (Forstmeier & Schielzeth, 2011), we compared our full models with null models lacking our test predictors but being otherwise identical using the likelihood ratio test. The significance of the beta coefficients was indicated by $p < .05$. Null hypothesis significance testing was complemented by examination of odds ratios and relative confidence intervals (see tables in Supplementary Materials). Odds ratio provided an index of the strength of the relationships between our predictors and outcome, thus enabling meaningful comparison between predictors. Marginal effects were plotted to provide straightforward visualization of predicted probabilities for the results of the more complex models (Lüdecke, 2018). Significant interaction terms accompanied by improvement in the model fit, as indicated by a significant likelihood ratio test, indicated the presence of conditional effects. Marginal R^2 indicated the variance explained by the fixed effects, while conditional R^2 indicated the variance explained by both the fixed and random effects.

Model Implementation

The models were implemented in R (version 4.0.4. 2021-02-15) with the function glmer of the R package lme4 (version 1.1- 26) (Bates et al., 2015). Marginal effects (i.e., predicted probabilities) were computed using the function ggpredict of the R package ggeffects (Lüdecke, 2018). Predicted probabilities plots are provided for visualisation purposes and were used to guide the interpretation of interaction terms. Best practice guidelines for the reporting of mixed models (Meteyard & Davies, 2020) were followed (i.e., we report the version of the software and packages used, data preparation steps, model selection and model output including coefficients, standard errors, confidence intervals and p-values, random effects, and R^2 as an index of model fit).

Exploratory Analysis

Exploring the Contribution of Pictorial Support, Pronunciation Support, Spoken and Written Definition and their Combinations.

We ran an exploratory analysis to examine the unique contributions of the different components of the scaffold (i.e., picture, pronunciation, written and spoken definitions) and their combinations (i.e., picture-pronunciation, picture-definition, pronunciation-definition, and all those together) on reading comprehension performance. To enable this, we subset our dataset and focused on instances where a single *Reveal Words* scaffold was used in each game item. This approach enabled us to quantify the proportion of correct responses across different components of the scaffold that were available and used, while fixing the number of scaffolds constant at 1 to ensure comparability across game items.

Do Age of Acquisition and Concreteness of Scaffolded Words Influence the Associations between Use of Scaffolds and Reading Comprehension?

We modelled an observational dataset, which was not designed for research purposes. Therefore, we checked whether key psycholinguistic variables such as age of acquisition and concreteness varied systematically across word types in our dataset. We further explored whether age of acquisition and concreteness influenced the effectiveness of the scaffold. We analysed the same subset of data as that used in the previous exploratory analysis. By working with instances where only a single *Reveal Words* scaffold was used in each game item, we were able to examine the relation between the psycholinguistic properties of a specific scaffolded word and reading comprehension performance (indicated by game item accuracy). The age of acquisition and concreteness ratings used in this analysis were derived from open sources databases (Brysbaert et al., 2014; Kuperman et al., 2012).

Results

Confirmatory analysis

Research Question 1: Do Child Characteristics Predict the Use of Vocabulary Scaffolds?

A clear impact of early literacy skills, grade, gender, and language status on the probability of using the scaffolding feature was evident. This was indicated by the significant likelihood ratio test comparing the full model including our test predictors and the null model lacking our predictors but being otherwise identical ($\chi 2 = 21877.00$, $df = 8$, $p < .001$). We discuss the output of the model (Table 2) first, followed by the model-generated predicted probability plots (Figure 2).

Table 2

Results of the Model Estimating the Effects of Literacy Skills, Grade, Gender and Language Status on the Probability of Using the Reveal Word Scaffold

	Use of the Scaffold			
Predictors	*Estimate*	*SE*	*CI*	*p*
Intercept	-0.76	0.06	-0.88 – -0.64	**<0.001**
DIBELS [Well Below Benchmark]	0.24	0.01	0.23 – 0.25	**<0.001**
DIBELS [Below Benchmark]	0.22	0.01	0.21 – 0.24	**<0.001**
DIBELS [Above Benchmark]	-0.31	0.01	-0.32 – -0.30	**<0.001**
Grade [1]	-0.56	0.01	-0.57 – -0.54	**<0.001**
Grade [2]	-1.21	0.01	-1.23 – -1.19	**<0.001**
Grade [3]	-1.44	0.01	-1.46 – -1.42	**<0.001**
Gender [Male]	-0.20	0.00	-0.21 – -0.19	**<0.001**
Language Status [Bilingual]	0.24	0.01	0.23 – 0.25	**<0.001**
Random Effects				
σ^2	3.29			
$\tau_{00\ q\_id}$	0.12			
$\tau_{00\ level\_name}$	0.37			
ICC	0.13			
N_{q_id}	505			
N_{level_name}	54			
Observations	1716787			

Marginal R^2 / Conditional R^2	0.065 / 0.186

Note. Literacy skills [DIBELS] are dummy coded with 'at benchmark' set as the reference category. Grade is dummy coded with kindergarten set as the reference category. Language status is dummy coded with English monolingual set as the reference category. Gender is dummy coded with girls set as the reference. Model fit is indicated by R^2.

With respect to literacy skills, children classified as being *well below benchmark* and *below benchmark* were more likely to use the scaffolding feature compared to children classified as being *at benchmark*. This was indicated by the positive sign of the respective coefficients (ß = 0.24, $p < .001$; ß = 0.22, $p < .001$). In contrast, children classified as being *above benchmark* were less likely to use the scaffolding feature compared to those *at benchmark*. This was indicated by the negative sign of the coefficient (ß = -0.31, $p < .001$). These results suggest that literacy skills predict the use of the scaffold. Students who need more support seek more support, as shown by the increased probability of using the scaffold in lower ability levels. The opposite was true for children in higher ability levels.

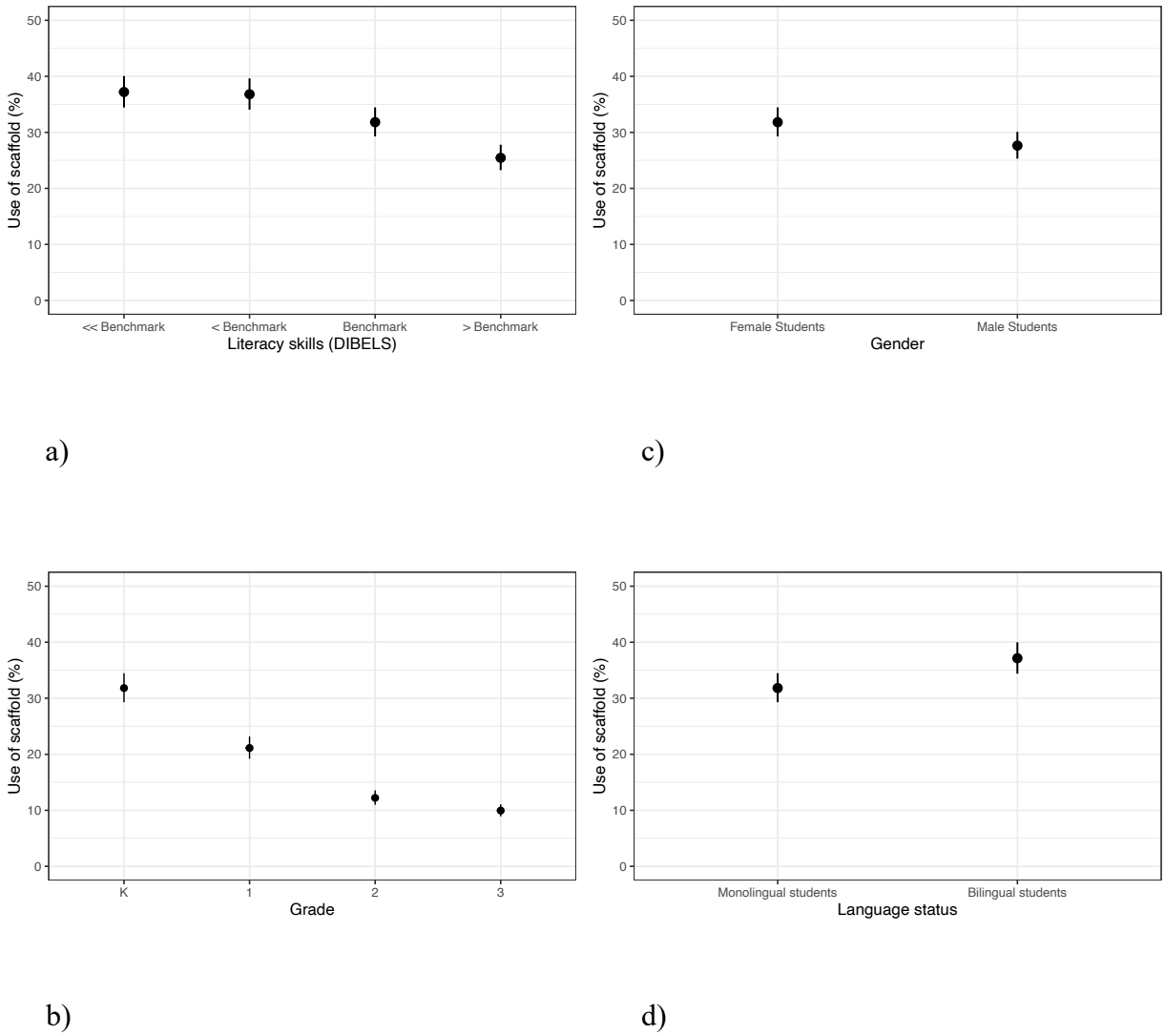
Grade served as a proxy for chronological age. A decreased probability of using the scaffold was associated with each of the higher grades, as compared to kindergarten. This was indicated by the negative sign of the coefficients for grades 1 (ß = -0.56, $p < .001$), 2 (ß = -1.21, $p < .001$), and 3 (ß = -1.44, $p < .001$). This pattern of results suggests a progressive *decrease* in the probability of using the scaffolding feature as students move up through the grades (i.e., increase in age). Beginner readers were more likely to make use of the scaffold, and this probability decreased in more experienced readers.

We also found an influence of gender and language status. Boys were less likely to use the scaffold compared to girls, as indicated by the negative sign of the coefficient (ß = - 0.20, $p <$ .001). Furthermore, students reporting as bilingual were more likely to use the scaffold compared to those reporting as English monolingual, as shown by the positive coefficient (ß = 0.23, $p <$ .001).

The predicted probability of using the scaffold expressed in percentage on the y axis, with child characteristics on the x axis is reported in Figure 2. Each panel displays one of the four child characteristics of interest. Error bars represent the confidence intervals. Figure 2a shows the extent to which the predicted probability of using the scaffold varied across literacy skills. Children *well below benchmark* had the highest predicted probability of using the scaffold (~ 35%) and this probability was less (~ 25%) for children *above benchmark*. Figure 2b illustrates how the predicted probability of using the scaffold varied across grades. Children in kindergarten had the highest predicted probability of use (~ 30%), while children in grade 3 had the lowest (~ 10%). Differences in the predicted probability of using the scaffold by gender are reported in Figure 2c. The predicted probability of using the scaffold for girls was above 30%, while for boys it was below 30%. Differences in the predicted probability of using the scaffold between monolingual and bilingual students are illustrated in Figure 2d. The predicted probability of using the scaffold for monolingual students was below 35%, while for bilingual students it was above 35%.

Figure 2

Predicted Probability of Using the Scaffold across Child Characteristics



Note. Predicted probability of using the scaffold by: (a) early literacy skills, (b) grade, (c) gender, and (d) language status, computed holding non-focal terms constant at their reference level. Error bars represent confidence intervals.

The strength of the relationships between child characteristics and use of scaffold as indicated by odds ratios is reported in the Supplementary Materials. Overall, the associations were modest, with odds ranging from 0.24 to 1.27. Children classified as being well below benchmark according to their literacy skills were 1.27 times more likely to use the scaffold as compared to children at benchmark.

Research Question 2: Is the Use of Vocabulary Scaffolds Associated with Reading Comprehension?

We examined whether use of the scaffold was associated with reading comprehension performance. Overall, the use of the scaffolding feature was associated with an increased probability of getting the associated item correct in the reading comprehension (text-picture matching) task. This was indicated by the significant likelihood ratio test comparing our full model with a null model lacking our test predictor but being otherwise identical ($\chi 2 = 2226.80$, $df = 1$, $p < 0.001$). The results of the full model are reported in Table 3.

Table 3

Results of the Model Estimating the Effects of the Scaffold, Literacy Skills, Grade, Gender and Language Status on Task Accuracy

	Accuracy in the Task			
Predictors	*Estimate*	*SE*	*CI*	*p*
Intercept	1.34	0.04	1.26 – 1.42	**<0.001**
Use of Scaffold [yes]	0.26	0.01	0.25 – 0.27	**<0.001**

DIBELS [Well Below Benchmark]	-0.39	0.01	-0.40 – -0.38	**<0.001**
DIBELS [Below Benchmark]	-0.10	0.01	-0.11 – -0.09	**<0.001**
DIBELS [Above Benchmark]	0.23	0.01	0.22 – 0.24	**<0.001**
Grade [1]	-0.01	0.01	-0.03 – 0.00	0.050
Grade [2]	-0.03	0.01	-0.05 – -0.02	**<0.001**
Grade [3]	0.16	0.01	0.14 – 0.18	**<0.001**
Gender [Male]	-0.09	0.00	-0.10 – -0.08	**<0.001**
Language Status [Bilingual]	-0.21	0.00	-0.22 – -0.21	**<0.001**

Random Effects

σ^2	3.29
$\tau_{00\ q\_id}$	0.70
$\tau_{00\ level\_name}$	0.04
ICC	0.18
N_{q_id}	505
N_{level_name}	54
Observations	1716787
Marginal R^2 / Conditional R^2	0.020 / 0.200

Note. Literacy skills [DIBELS] are dummy coded with at benchmark set as the reference category. Grade is dummy coded with kindergarten set as the reference category. Language status is dummy coded with English monolingual set as the reference category. Gender is dummy coded with girls set as the reference. Model fit is indicated by R^2.

Use of the scaffolding feature was associated with an increased probability of getting the item correct compared to not using the scaffold. This is indicated by the positive sign of the coefficient (ß = 0.26, $p < .001$). Importantly, the effect of the scaffold was significant after controlling for child literacy skills, grade, gender, and language status. Of note, each of the four control predictors was significant and indicated that children with lower literacy skills, boys, and bilingual students were less likely to get the item correct in the reading comprehension task. In addition, children in grade 2 were less likely to get the item correct compared to kindergarteners while children in grade 3 were more likely to get the item correct compared to kindergarteners. Odds ratios are reported in Supplementary Materials to illustrate the strength of the association between use of scaffold and reading comprehension. This association was modest ($OR = 1.30$). We consider the practical significance of these findings in the discussion.

Research Question 3: Is the Association between Use of Scaffolds and Reading Comprehension Conditional on Child and/or Item Characteristics?

In these analyses, we further investigated the interactions between the use of the scaffold and child characteristics and the influence of item characteristics (word type) on the probability of getting the game item correct. The comparison of the full and null model ($\chi 2 = 1350.00$, df = 13, $p < .001$) revealed significant interactions between use of scaffold and child characteristics and a significant effect of word type. The results of the full model are reported in Table 4.

Table 4

Results of the Model Estimating the Effects of the Scaffolded Word Type, and Interaction Between Scaffold Use and Literacy Skills, Grade, Gender, and Language Status on the Accuracy in the Task

	Accuracy in the Task			
Predictors	*Estimate*	*SE*	*CI*	*p*
Intercept	1.30	0.04	1.22 – 1.38	**<0.001**
Use of Scaffold [yes]	0.45	0.02	0.41 – 0.49	**<0.001**
DIBELS [Well Below Benchmark]	-0.41	0.01	-0.42 – -0.40	**<0.001**
DIBELS [Below Benchmark]	-0.12	0.01	-0.13 – -0.10	**<0.001**
DIBELS [Above Benchmark]	0.25	0.01	0.24 – 0.26	**<0.001**
Grade [1]	0.03	0.01	0.02 – 0.05	**<0.001**
Grade [2]	0.02	0.01	0.01 – 0.04	**0.009**
Grade [3]	0.22	0.01	0.20 – 0.24	**<0.001**
Gender [Male]	-0.09	0.00	-0.10 – -0.09	**<0.001**
Language Status [Bilingual]	-0.22	0.00	-0.23 – -0.21	**<0.001**
Use of Scaffold [yes]:DIBELS [Well Below Benchmark]	0.20	0.01	0.17 – 0.23	**<0.001**
Use of Scaffold [yes]:DIBELS [Below Benchmark]	0.11	0.02	0.08 – 0.14	**<0.001**

Use of Scaffold [yes]:DIBELS [Above Benchmark]	-0.15	0.01	-0.18 – -0.13	**<0.001**
Use of Scaffold [yes]:Grade [1]	-0.23	0.02	-0.26 – -0.20	**<0.001**
Use of Scaffold [yes]:Grade [2]	-0.34	0.02	-0.38 – -0.31	**<0.001**
Use of Scaffold [yes]:Grade [3]	-0.40	0.02	-0.44 – -0.36	**<0.001**
Use of Scaffold [yes]:Gender [Male]	0.03	0.01	0.01 – 0.05	**0.015**
Use of Scaffold [yes]:Language Status [Bilingual]	0.03	0.01	0.00 – 0.05	**0.023**
Scaffolded Nouns	0.03	0.01	0.01 – 0.05	**0.001**
Scaffolded Verbs	-0.02	0.01	-0.04 – 0.00	0.057
Scaffolded Adjectives	-0.35	0.02	-0.40 – -0.30	**<0.001**
Scaffolded Adverbs	-0.18	0.05	-0.28 – -0.07	**0.001**
Scaffolded Prepositions	-0.02	0.06	-0.14 – 0.10	0.712

Random Effects

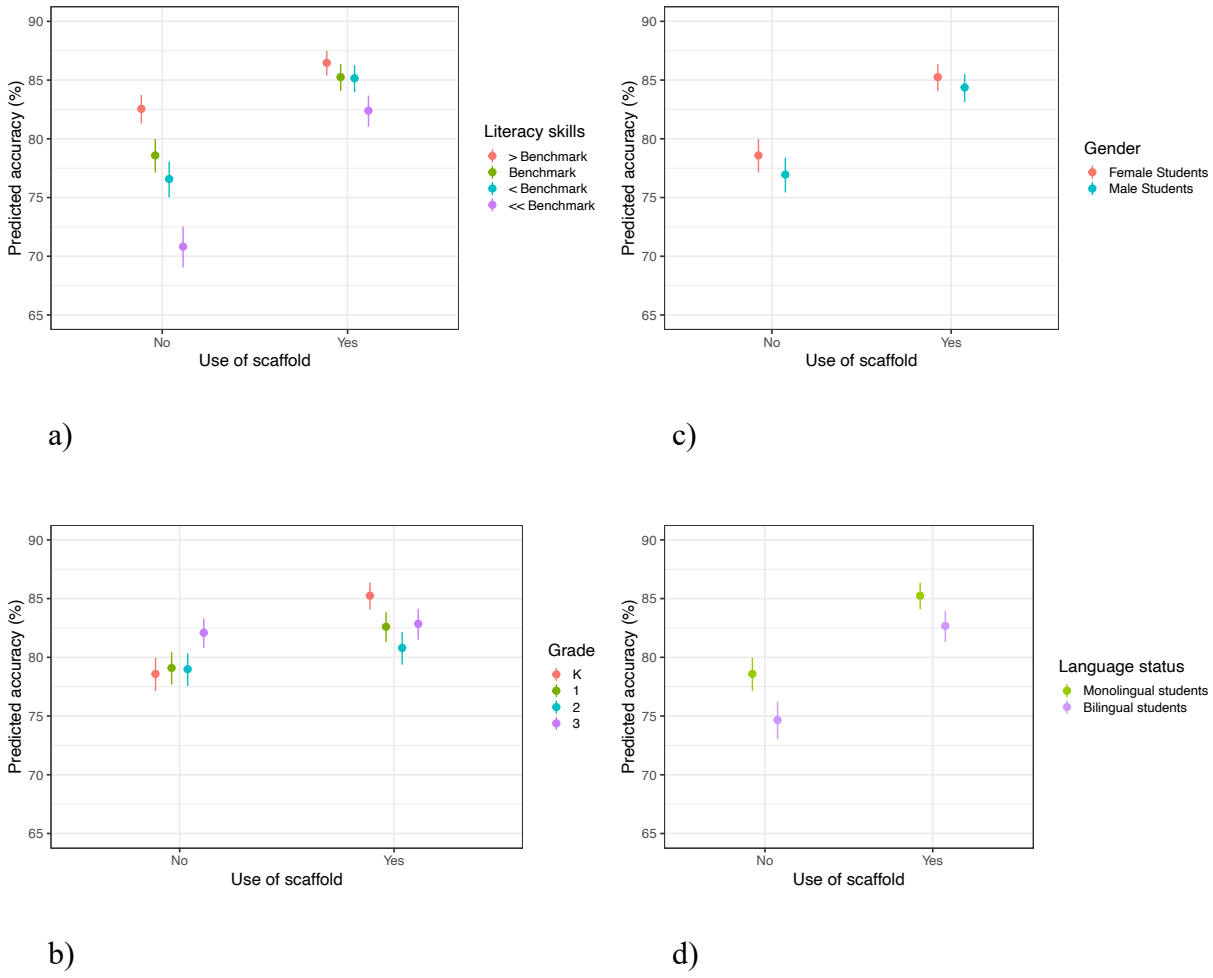
σ^2	3.29
$\tau_{00\ q\_id}$	0.69
$\tau_{00\ level\_name}$	0.05
ICC	0.18
$N_{\ q\_id}$	505

N_{level_name}	54
Observations	1716787
Marginal R^2 / Conditional R^2	0.021 / 0.201

Note. Literacy skills [DIBELS] are dummy coded with at benchmark set as the reference category. Grade is dummy coded with kindergarten set as the reference category. Language status is dummy coded with English monolingual set as the reference category. Gender is dummy coded with girls set as the reference. Model fit is indicated by R^2.

Figure 3

Predicted Probability of Accuracy in the Task by Use of Scaffold across Child Characteristics



Note. Predicted probability of accuracy (y axis) plotted against the use of scaffold (x axis) across child characteristics (colour coded) for: (a) literacy skills, (b) grade, (c) gender, and (d) language status. Predicted probability are computed holding non-focal terms constant at their reference level.

To guide interpretation of the interaction between scaffold use and child characteristics, we provide visualization of predicted probabilities (Figure 3). The significant interactions between scaffold use and literacy skills (ß = 0.20, $p < .001$; ß = 0.11, $p < .001$; ß = -0.15, $p < .001$) suggest that the association between use of scaffolds and reading comprehension was modulated by literacy skills. As illustrated in Figure 3a, use of the scaffold was associated with higher performance for all children, but to a different extent depending on their literacy skills. Specifically, this association was greater for readers at lower ability levels. In other words, the use of the scaffold tended to minimise performance differences between ability levels. The significant interactions between scaffold use and grade (ß = -0.23, $p < .001$; ß = -0.34, $p < .001$; ß = -0.40, $p < .001$) suggest that the association between use of scaffolds and reading comprehension was modulated by grade. Specifically, the predicted probability plot (Figure 3b) illustrates that, although the use of the scaffold was associated with higher performance for all children, this association was greater for younger children such as those in kindergarten.

There was a significant interaction between scaffold use and gender (ß = 0.03, $p = .01$). The predicted probability plot (Figure 3c) illustrates that the differences in performance due to gender were slightly reduced when the scaffold was used. Finally, there was a significant interaction between scaffold use and language status (ß = 0.03, $p = .02$). The predicted probability plot (Figure 3d) shows that performance differences associated with language status were reduced when the scaffold was used.

The number of scaffolds used for nouns in each game item had a positive impact on the probability of getting the game item correct, as compared to not using a scaffold for nouns, as indicated by the positive sign of the coefficient (ß = 0.03, $p < .001$). That is, for nouns, use of the scaffold was associated with better performance. In contrast, the number of scaffolds used for

adjectives and adverbs in each game item had a negative impact on accuracy, as compared to not using a scaffold for adjectives and adverbs respectively. This is indicated by the negative sign of the respective coefficients (ß = -0.35, $p < .001$; ß = -0.18, $p = .001$). There was no clear effect of the number of scaffolds used for verbs and prepositions on item accuracy (ß = -0.02, $p = .06$; ß = -0.02, $p = .71$). Odds ratios are reported in the Supplementary Materials to illustrate the strength of the associations under investigations. Similar to previous analyses, the associations were modest.

Exploratory Analysis

Exploring the Contribution of Pictorial Support, Pronunciation Support, Spoken and Written Definition and their Combinations

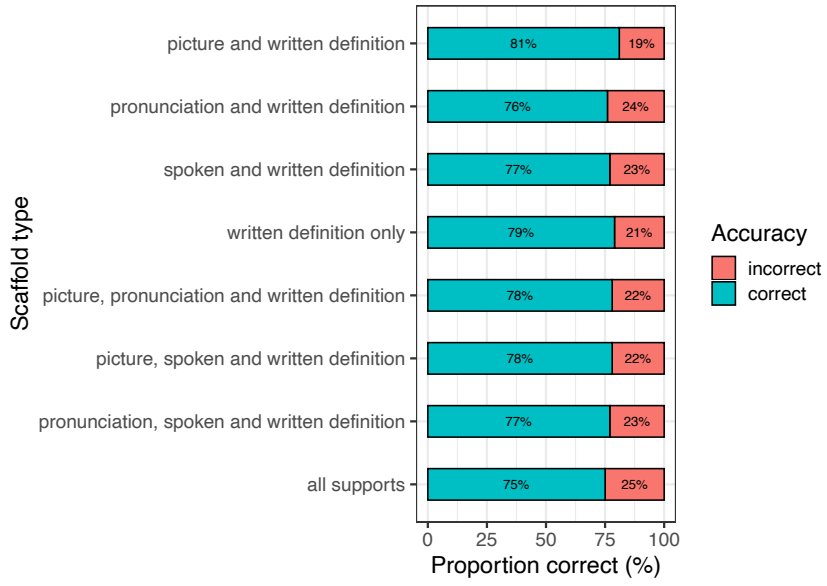
We conducted an exploratory analysis with the aim to disentangle the associations between use of different components of the scaffold (i.e., picture, pronunciation, written and spoken definitions) and their combinations (i.e., picture-pronunciation, picture-definition, pronunciation-definition, and all those together) and reading comprehension performance. Written definitions were available for all *Reveal Words* scaffolds, whilst pictures were available for most, but not all (see learning materials in the methods section). Of note, when children clicked on the scaffold, written definitions and pictures (if available) were immediately visible, whereas children had to click further on the speaker icon to obtain the pronunciation and spoken definition supports. To examine the contributions of these distinct scaffold components and their combinations, we subset our dataset and focused on instances where a single *Reveal Words* scaffold was used in each game item. This was because some items had scaffolds for more than a single word. This approach enabled us to quantify how the proportion of correct responses varied across the components of the scaffold that were available and used, while fixing the number of

scaffolds constant at 1 to ensure comparability across game items. Descriptive statistics revealed that the proportion of correct responses was broadly comparable across the different scaffold components and their combinations, with the proportion of correct responses ranging between 75 to 81% (Figure 4).

These descriptives also suggest that the effects of different components of a scaffold were not additive. For instance, the proportion of correct responses when all the supports were available and used (i.e., picture, pronunciation, spoken and written definition) was lower (75%) compared to when only the picture and written definition support were available and used (81%), or only the written definition was available (79%). Overall, pictorial support and written definition in combination were associated with the highest proportion of correct responses, followed by written definition alone (proportion correct 81% and 79%, respectively).

Figure 4

Proportion of Correct Responses across Scaffold Components and Combinations



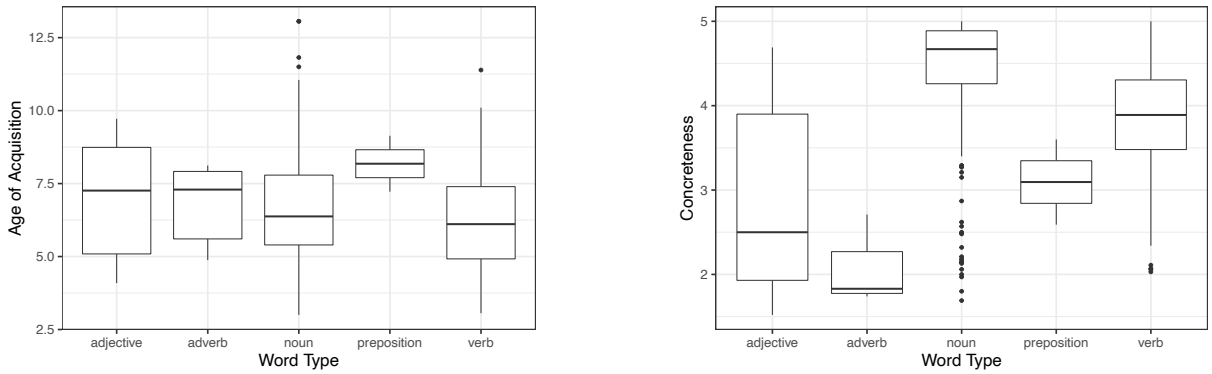
Note. This analysis was conducted on a subset dataset where only a single scaffold per game item was used to enable meaningful comparisons.

Do Age of Acquisition and Concreteness of Scaffolded Words Modulate the Associations between Use of Scaffolds and Reading Comprehension?

Because we modelled an observational dataset we checked whether key psycholinguistic variables of the words, such as age of acquisition and concreteness, varied systematically across word types in our dataset. To this end, we produced the boxplots illustrated in Figure 5. The horizontal line in the body of each boxplot represents the median value of age of acquisition for each category of word. The body of the boxplot represents the interquartile range. The vertical line in each boxplot signals the minimum and maximum value. Dots represent potential outliers. Figure 5 illustrates that age of acquisition was comparable across word type (Figure 5a), with median age of acquisition between 6 and 8 years of age (i.e., the age range of our participating children). In contrast, median concreteness differed across word types (Figure 5b) and may thus act as a confounding variable. We will return to this point in the discussion.

Figure 5

Age of Acquisition and Concreteness across Word Types



a) b)

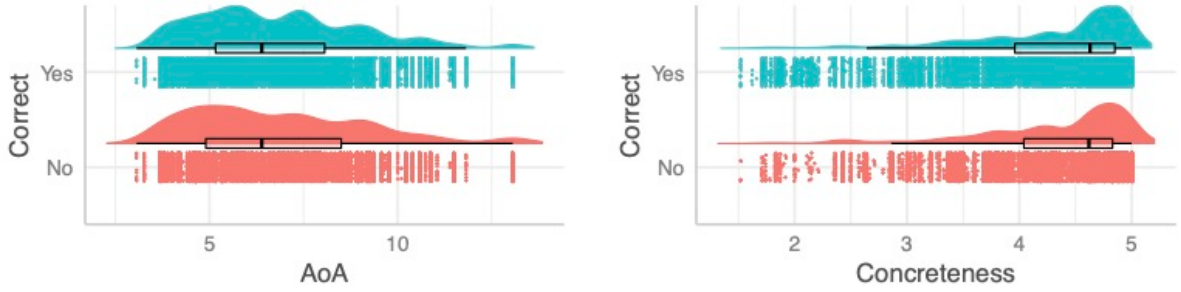
Note. Box plots of (a) age of acquisition, and (b) concreteness, across word type

To explore whether age of acquisition and concreteness of scaffolded words influence the associations between use of scaffolds and reading comprehension, we analysed the same subset of data used in the previous exploratory analysis. As before, we worked with instances where only a single *Reveal Words* scaffold was used in each game item. This enabled us to examine the relation between the psycholinguistic properties of a specific scaffolded word and reading comprehension performance as evidenced by game item accuracy. The results are illustrated with rainclouds plots in Figure 6. Correct and incorrect responses are shown on the y axis and age of acquisition (Figure 6a) and concreteness (Figure 6b) on the x axis. These visualizations combine boxplots (see previous paragraph for interpretation), probability density, and the jittered raw dataset (Allen et al., 2019). This comprehensive data overview suggests that, in a scenario where children use a single scaffold, the psycholinguistic properties of the scaffolded word do not influence accuracy. Indeed, the median and distribution of age of acquisition and concreteness of

Reveal Words scaffolds is comparable across correct and incorrect responses. This pattern suggests that the age of acquisition and concreteness of scaffolded words do not influence the associations between use of scaffolds and reading comprehension.

Figure 6

Age of Acquisition and Concreteness across Correct and Incorrect Responses



a) b)

Note. Raincloud plots showing (a) age of acquisition and (b) concreteness on the x axis, across correct and incorrect responses on the y axis, in a scenario where a single scaffold was used.

Discussion

Our findings advance our understanding of vocabulary scaffolding features embedded in digital reading environments in several ways. First, we found that child characteristics predicted the use of scaffolds: Children with lower literacy skills, beginner readers, girls, and students who reported as bilingual were more likely to use the scaffold. Second, the use of the scaffold was associated with better reading comprehension. This association was stronger for three of the four child characteristics associated with greater scaffold use: Children with lower literacy skills, beginner readers, and bilingual students. In addition, this association was stronger for boys than for girls. Third, the association between the use of scaffold and reading comprehension varied

across word types, with a positive association for nouns, a negative association for adjective and adverbs, and no evidence of an association for verbs and prepositions. We discuss each finding in relation to the previous literature, followed by recommendations for future research and educational implications.

Child Characteristics Predicted the Use of Scaffolds

Our first research question sought to establish which child characteristics predicted the use of vocabulary scaffolds that were available to learners on an as-needed basis. Children with lower literacy skills and beginner readers were more likely to use the scaffold. This suggests that struggling and beginner readers have sufficient metacognitive awareness to seek support while navigating a digital reading environment. In addition, girls and bilingual students were more likely to use the scaffold. These findings are in line with evidence for enhanced metacognitive skills in girls (Wu, 2014) and bilingual readers (Abu Rabia, 2019). On the one hand, these results are encouraging as they show that children with lower literacy skills, beginner readers, and bilingual readers, all groups that are associated with lower reading comprehension (Catts et al., 2006; LARRC, 2015; Lesaux & Kieffer, 2010), are more likely to use a scaffold. On the other hand, these results point to the need to further prompt the use of such scaffolds amongst male students. Our analyses showed that boys were less accurate compared to girls, and that performance differences due to gender were reduced when the scaffold was used. This has important and positive implications for educational practice, due to the common finding of lower reading attainment in boys than in girls (Duncan et al., 2016; Wu, 2014). In addition, we note that it may be beneficial to prompt the use of the scaffolding feature across all students, as the overall use of the scaffold was lower than desirable, and scaffold use was associated greater reading comprehension. Two potential ways to promote the use of the scaffold are the addition of

extended on-boarding instructions reviewing how to use the feature, and/or initial adult modelling. These strategies have the potential to promote children's independent use of the scaffold. App developers should also consider providing these types of scaffolds based on a student's performance, in an adaptive manner (Sampayo-Vargas et al., 2013).

The Use of Scaffolds was Associated with Better Reading Comprehension

Critically, we found that the use of the vocabulary scaffolding feature was associated with better reading comprehension of short digital texts. These results extend previous meta-analytic evidence reporting a positive effect of similar scaffolds on the reading comprehension of university students when reading in a foreign language (Abraham, 2008). The benefit of vocabulary scaffolds appears to generalize to younger students, such as those included in our study (i.e., 5 to 8 years), and to first language reading comprehension[6]. Our findings also add to the literature demonstrating that other types of scaffold, such as graphic organizers, can benefit reading comprehension (Elbro & Buch-Iversen, 2013). However, our results contrast with meta-analytic evidence showing a null or negative effect on children's story comprehension when dictionary supports are embedded in storybooks (Furenes et al., 2021). These inconsistencies are likely to arise from the longer texts used in the storybooks examined in other research, and possibly from the co-reading context that was the focus of the meta-analysis. Future research should determine any specific influence of the context (educational, recreational, individual, co-reading), as well as length of text, to inform specific recommendations for the inclusion of this support.

The Association between Use of Scaffolds and Reading Comprehension Was Modulated by Child and Item Characteristics

[6] Our sample included both English monolingual and bilingual students, both reading in English, thus providing evidence for first language reading comprehension.

Our study provides the first evidence that the association between use of scaffolds and reading comprehension was modulated by both child and item characteristics. Specifically, our analysis suggested that the associations between the use of scaffolds and reading comprehension was greater for children with lower literacy skills and beginner readers. This extends previous research showing that the beneficial effects of diagrams as scaffolds for learners' cognitive and metacognitive processes is stronger for participants with low verbal ability (Cuevas et al., 2002). Similarly, our findings align with a meta-analysis suggesting that the impact of similar scaffolds is greater in beginner learners when acquiring novel words in a foreign language (Yun, 2011) and extends those findings to the domain of reading comprehension.

Additionally, we found that performance differences due to gender and language status were reduced when the scaffold was used. Because boys and bilingual students are typically found to have poorer reading comprehension skills (Duncan et al., 2016; Wu, 2014), our findings are especially encouraging. In short, the use of scaffolds is associated with higher performance especially in typically lower achieving student subgroups. This has important educational implications and speaks in favour of embedding scaffolding features in digital texts, especially for children with or at risk of reading comprehension difficulties.

Regarding item characteristics, the use of scaffolds was associated with higher performance for nouns and lower performance for adjective and adverbs. No differences emerged for verbs and prepositions. The advantage observed for scaffolded nouns can be explained in two ways. One possibility is that nouns are simply easier to learn (see Ellis & Beaton, 1993, but also Schwanenflugel, Stahl, & McFalls, 1997, for an alternative account). Another is that because nouns are more concrete than the other word types in in our dataset, they

were easier to depict and define with in-text scaffolds, thus making scaffolding feature more effective for this word type.

Practical significance

The strength of the associations reported in this study was modest, that is odds ratios were lower than 1.68 (H. Chen, Cohen & S. Chen, 2010). However, in terms of practical significance these associations may be meaningful. First, the literature on incidental vocabulary learning and reading development suggests that small effects, when cumulative, can have long term benefits over time (Anderson, Wilson, & Fielding, 1988; Fukkink & de Glopper, 1998). Second, the associations reported here were captured in highly naturalistic settings and are potentially very conservative estimates, given the noise inherent in this type of dataset. In addition, vocabulary scaffolds embedded in digital supplements represent a relatively low-cost approach to support, as compared to more resource intensive approaches, such as tutoring or classroom-based instructions. In this context, our identification of positive associations suggests the potential utility of including scaffolds in digital supplements to support independent reading and, through that, both comprehension and vocabulary growth. Taken together, these observations lead us to recommend the inclusion of embedded scaffolds in digital reading materials to provide timely support.

Exploring the Contribution of Different Scaffold Components and their Combinations

Our exploratory analysis further suggests that a written definition alone, and especially in conjunction with pictorial support, is associated with a higher proportion of correct responses compared to other combinations of scaffold components, such as the combination of picture, pronunciation, written and spoken definitions. These findings are in line with, and extend, previous evidence indicating that children's narrative skills benefit more from the combination of

verbal and non-verbal scaffolds, than from just verbal scaffolds alone (Pesco & Gagné, 2017), as well as evidence documenting the benefits of imagery training to children's reading comprehension (Francey & Cain, 2015; Joffe et al., 2007). Our descriptive analysis also suggests that the use of multiple scaffold components at a time is not necessarily associated with higher performance. This can be interpreted in several ways. One possibility is that performance when using the scaffold was already high, leaving little room for improvement when additional spoken supports were used. Another possibility is that interacting with several components of the scaffold may increase cognitive load and distract students from the main task (Mayer & Moreno, 2003). This may explain why performance was less accurate when children used several scaffold components for the same item. An alternative account of these exploratory results is that children clicked on more components of the scaffold when they faced a particularly challenging item, suggesting that the benefits of scaffolds may vary by item difficulty. Future experimental work is needed to validate this pattern of results under more controlled conditions.

Exploring the Influence of Key Psycholinguistic Variables

Our exploratory analysis further investigated whether key psycholinguistic variables such as age of acquisition and concreteness modulated the association between use of scaffolds and reading comprehension. Our data suggest that they did not. Again, future experimental work is needed to test the reproducibility of this set of results under carefully controlled conditions. Information about any influence of the psycholinguistic properties of words in relation to both the use of scaffolds and their associations with reading performance is crucial to inform both practice and theory.

Limitations and Future Research

This study comes with limitations and suggestions for future research, additional to those already discussed. First, causal inference is beyond the scope of this analysis, due to the observational nature of the dataset and lack of experimental manipulation. It is however plausible that the nature of the associations described in this work are causal; that is, the use of the scaffold resulted in enhanced item accuracy. This possibility should be taken up by the research community and tested in future experimental work, in a cumulative science framework. Engagement and motivation may also play an important in reading performance and future work should take those factors into account when examining the role of scaffolding features. A strength of our approach was the interrogation of a substantial dataset of children's interactions with a digital reading environment providing the statistical power to examine several child characteristics. Concerns have been raised in relation to high-powered studies. We note here that those studies are problematic in conjunction with null-hypothesis significance testing only when the sample is increased arbitrarily to achieve significance and/or p-values are the only criterion used to evaluate results. This does not apply to our study as we have taken the strength of the relationships into account and have exploited an existing big dataset. A large sample in this context provides more stable estimates, increasing the chance that these patterns of results will be observed in future studies (addressing current and critical concerns about the reproducibility of scientific findings). Of course, our findings are limited to the specific reading game under investigation; future research is needed to clarify whether these associations generalize to a different sample, context, type of text, and reading comprehension measure. Finally, the differential impact of the scaffolds across word types is an important area for future research to inform how to best scaffold the meaning of adjectives and adverbs, the impact of scaffolds for

verbs and prepositions, and the role of other key psycholinguistic variables. This information is critical to develop appropriate scaffolds for a range of vocabulary.

Conclusions

Vocabulary scaffolding features may be promising tools to promote reading comprehension in general, and to reduce performance differences amongst diverse readers in digital environments. This work can be viewed as a proof of principle for the feasibility of collaborations between academics and industrial partners in the field of language and literacy research in the era of large app-generated data. Specifically, we have shown that such collaborations can be undertaken in accordance with academic standards and that they can contribute meaningfully to inform a broader audience of app developers, educators, and policy makers.

Acknowledgments

We sincerely thank all the participating schools, teachers, parents, and children for contributing data. We also thank the Literacy Acceleration through Scalable Research (LASeR) consortium and all the collaborators across the industrial and academic network. This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Actions grant agreement No 857897.

CRediT authorship contribution statement

L. Diprossimo: Conceptualization, Formal analysis, Data curation, Writing original draft, Visualization, Project administration; **A. Ushakova:** Conceptualization, Formal analysis, Writing original draft, Project administration; **J. Zoski:** Resources, Writing - review & editing; **H. Gamble:** Resources, Data curation, Writing - review & editing; **R. Irey:** Writing - review & editing; **K. Cain:** Conceptualization, Writing original draft, Supervision, Funding acquisition.

References

Abraham, L. B. (2008). Computer-mediated glosses in second language reading comprehension and vocabulary learning: A meta-analysis. *Computer Assisted Language Learning*, *21*(3), 199–226. https://doi.org/10.1080/09588220802090246

Abu Rabia, S. (2019). The effect of degrees of bilingualism on metacognitive linguistic skills. *International Journal of Bilingualism*, *23*(5), 1064–1086. https://doi.org/10.1177/1367006918781060

Akbulut, Y. (2007). Effects of multimedia annotations on incidental vocabulary learning and reading comprehension of advanced learners of english as a foreign language. *Instructional Science*, *35*(6), 499–517. https://doi.org/10.1007/s11251-007-9016-7

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. A. (2019). Raincloud plots: a multi-platform tool for robust data visualization. Wellcome open research, 4, 63. https://doi.org/10.12688/wellcomeopenres.15191.1

Anderson, R. C., Wilson, P. T., & Fielding, L. G. (1988). Growth in reading and how children spend their time outside of school. *Reading research quarterly*, 285-303.

Andrä, C., Mathias, B., Schwager, A., Macedonia, M., & von Kriegstein, K. (2020). Learning foreign language vocabulary with gestures and pictures enhances vocabulary memory for several months post-learning in eight-year-old school children. *Educational Psychology Review*, *32*(3), 815–850. https://doi.org/10.1007/s10648-020-09527-z

Baker, L., & Cerro, L. C. (2000). *Assessing Metacognition in Children and Adults*.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. Journal of Statistical Software, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Blything, L. P., Hardie, A., & Cain, K. (2020). Question asking during reading comprehension instruction: a corpus study of how question type influences the linguistic complexity of primary school students' responses. *Reading Research Quarterly*, *55*(3), 443–472. https://doi.org/10.1002/rrq.279

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand English word lemmas. *Behavior Research Methods*, *46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Cain, K., & Oakhill, J. V. (2018). Vocabulary development and reading comprehension: a reciprocal relationship. *Oxford University Press.*, 1. http://oxford.ly/wordgap

Cain, K., Oakhill, J. V., & Elbro, C. (2003). The ability to learn new word meanings from context by school-age children with and without language comprehension difficulties. *Journal of Child Language*, *30*(3), 681–694. https://doi.org/10.1017/S0305000903005713

Catts, H. W., Adlof, S. M., & Weismer, S. E. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language, and Hearing Research*, *49*(2), 278–293. https://doi.org/10.1044/1092-4388(2006/023)

Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics—simulation and Computation®*, *39*(4), 860-864.

Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T., & Fischer, F. (2020). Simulation-based learning in higher education: a meta-analysis. *Review of Educational Research*, *90*(4), 499–541. https://doi.org/10.3102/0034654320933544

Childers, J. B., & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology*,

38(6), 967–978. https://doi.org/10.1037/0012-1649.38.6.967

Clark, K. F., & Graves, M. F. (2005). Scaffolding students' comprehension of text. *The Reading Teacher*, *58*(6), 570–580. https://doi.org/10.1598/rt.58.6.6

Cohen, P. A., Kulik, J. A., & Kulik, C.L. C. (1982). Educational outcomes of tutoring: a meta-analysis of findings. *American Educational Research Journal*, *19*(2), 237–248. https://doi.org/10.3102/00028312019002237

Cuevas, H. M., Fiore, S. M., & Oser, R. L. (2002). Scaffolding cognitive and metacognitive processes in low verbal ability learners: Use of diagrams in computer-based training environments. *Instructional Science*, *30*(6), 433–464. https://doi.org/10.1023/A:1020516301541

Cunningham, A. E., & Stanovich, K. E. (1998). What reading does for the mind. *Journal of Direct Instruction*, *1*(2), 137–149.

Dalton, B., Proctor, C. P., Uccelli, P., Mo, E., & Snow, C. E. (2011). Designing for diversity: The role of reading strategies and interactive vocabulary in a digital reading environment for fifth-grade monolingual English and Bilingual students. *Journal of Literacy Research*, *43*(1), 68–100. https://doi.org/10.1177/1086296X10397872

De Sousa, I., & Oakhill, J. (1996). Do levels of interest have an effect on children's comprehension monitoring performance? *British Journal of Educational Psychology*, *66*(4), 471–482. https://doi.org/10.1111/j.2044-8279.1996.tb01213.x

Degener, S., & Berne, J. (2017). Complex questions promote complex thinking. *Reading Teacher*, *70*(5), 595–599. https://doi.org/10.1002/trtr.1535

Diprossimo, L., Cain, K., & Ushakova, A. (2021, June 9). Vocabulary Scaffolding Features and Young Readers' Comprehension of Digital Text: Insights from a Big Observational Dataset.

https://doi.org/10.17605/OSF.IO/62C4Q

Duncan, L. G., Mcgeown, S. P., Griffiths, Y. M., Stothard, S. E., & Dobai, A. (2016). Adolescent reading skill and engagement with digital and traditional literacies as predictors of reading comprehension. *British Journal of Psychology*, *107*(2), 209–238. https://doi.org/10.1111/bjop.12134

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. https://doi.org/10.1080/10888438.2013.774005

Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). The impact of vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Research on Educational Effectiveness*, *2*(1), 1–44. https://doi.org/10.1080/19345740802539200

Ellis, N. C., & Beaton, A. (1993). Psycholinguistic determinants of foreign language vocabulary learning. *Language Learning*, *43*(4), 559–617. https://doi.org/10.1111/j.1467-1770.1993.tb00627.x

Forstmeier, W., & Schielzeth, H. (2011). Cryptic multiple hypotheses testing in linear models: Overestimated effect sizes and the winner's curse. *Behavioral Ecology and Sociobiology*, *65*(1), 47–55. https://doi.org/10.1007/s00265-010-1038-5

Francey, G., & Cain, K. (2015). Effect of imagery training on children's comprehension of pronouns. *Journal of Educational Research*, *108*(1), 1–9. https://doi.org/10.1080/00220671.2013.824869

Fukkink, R. G., & de Glopper, K. (1998). Effects of instruction in deriving word meaning from context: A meta-analysis. Review of educational research, 68(4), 450-469.

Furenes, M. I., Kucirkova, N., & Bus, A. G. (2021). A comparison of children's reading on paper versus screen: a meta-analysis. In *Review of Educational Research: Vol. XX* (Issue X). https://doi.org/10.3102/0034654321998074

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. A. Kuczaj (Ed.), *Language development: Language, thought, and culture (Vol. 2, pp. 301–334)*. Hillsdale, NJ: Erlbaum.

Goldin-Meadow, S., Seligman, M. E., & Gelman, R. (1976). Language in the two-year old. *Cognition*, *4*, 189–202.

Gonzalez, M. (2014). The effect of embedded text-to-speech and vocabulary ebook scaffolds on the comprehension of students with reading disabilities. *International Journal of Special Education*, *29*(3), 111–125.

Hadley, E. B., Dickinson, D. K., Hirsh-Pasek, K., Golinkoff, R. M., & Nesbitt, K. T. (2016). Examining the acquisition of vocabulary knowledge depth among preschool students. *Reading Research Quarterly*, *51*(2), 181–198. https://doi.org/10.1002/rrq.130

Hald, L. A., de Nooijer, J., van Gog, T., & Bekkering, H. (2016). Optimizing word learning via links to perceptual and motoric experience. *Educational Psychology Review*, *28*(3), 495–522. https://doi.org/10.1007/s10648-015-9334-2

Hansen, P. (2017). What makes a word easy to acquire? The effects of word class, frequency, imageability and phonological neighbourhood density on lexical development. *First Language*, *37*(2), 205–225. https://doi.org/10.1177/0142723716679956

Joffe, V. L., Cain, K., & Marić, N. (2007). Comprehension problems in children with specific language impairment: Does mental imagery training help? *International Journal of Language and Communication Disorders*, *42*(6), 648–664.

https://doi.org/10.1080/13682820601084402

Kaushanskaya, M., Gross, M., & Buac, M. (2013). Gender differences in child word learning. *Learning and Individual Differences*, *27*, 82–89. https://doi.org/10.1016/j.lindif.2013.07.002

Kim, D., & Gilman, D. A. (2008). Effects of text, audio, and graphic aids in multimedia instruction for vocabulary learning. *Journal of Educational Technology & Society*, *11*(3), 114–126. https://www.jstor.org/stable/10.2307/jeductechsoci.11.3.114

Kim, N. J., Belland, B. R., & Walker, A. E. (2018). Effectiveness of computer-based scaffolding in the context of problem-based learning for stem education: bayesian meta-analysis. *Educational Psychology Review*, *30*(2), 397–429. https://doi.org/10.1007/s10648-017-9419-1

Kim, Y.-S., Al Otaiba, S., Puranik, C., Folsom, J. S., & Gruelich, L. (2014). The contributions of vocabulary and letter writing automaticity to word reading and spelling for kindergartners. *Reading and Writing*, *27*(2), 237–253. https://doi.org/10.1007/s11145-013-9440-9

Kirby, J. R., & Moore, P. J. (1987). Metacognitive awareness about reading and its relation to reading ability. *Journal of Psychoeducational Assessment*, *2*, 119–137.

Korat, O., & Shamir, A. (2012). Direct and indirect teaching: Using e-books for supporting vocabulary, word reading, and story comprehension for young children. *Journal of Educational Computing Research*, *46*(2), 135–152. https://doi.org/10.2190/EC.46.2.b

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30 thousand English words. *Behavior Research Methods*, *44*(4), 978–990. https://doi.org/10.3758/s13428-012-0210-4

Lämsä, J., Hämäläinen, R., Aro, M., Koskimaa, R., & Äyrämö, S. M. (2018). Games for

enhancing basic reading and maths skills: A systematic review of educational game design in supporting learning by people with learning disabilities. *British Journal of Educational Technology*, *49*(4), 596–607. https://doi.org/10.1111/bjet.12639

Language and Reading Research Consortium. (2015). Learning to read: Should we keep things simple? *Reading Research Quarterly*, *50*(2), 151–169. https://doi.org/10.1002/rrq.99

Language and Reading Research Consortium, Currie, N. K., & Muijselaar, M. M. L. (2019). Inference making in young children: The concurrent and longitudinal contributions of verbal working memory and vocabulary. *Journal of Educational Psychology*, *111*(8), 1416–1431. https://doi.org/10.1037/edu0000342

Lesaux, N. K., & Kieffer, M. J. (2010). Exploring sources of reading comprehension difficulties among language minority learners and their classmates in early adolescence. In *American Educational Research Journal* (Vol. 47, Issue 3). https://doi.org/10.3102/0002831209355469

Liu, Y. T., & Leveridge, A. N. (2017). Enhancing L2 vocabulary acquisition through implicit reading support cues in e-books. *British Journal of Educational Technology*, *48*(1), 43–56. https://doi.org/10.1111/bjet.12329

Lüdecke, D. (2018). ggeffects: tidy data frames of marginal effects from regression models. *Journal of Open Source Software*, *3*(26), 772. https://doi.org/10.21105/joss.00772

Manzano-León, A., Camacho-Lazarraga, P., Guerrero, M. A., Guerrero-Puerta, L., Aguilar-Parra, J. M., Trigueros, R., & Alias, A. (2021). Between level up and game over: A systematic literature review of gamification in education. *Sustainability*, *13*(4), 1–14. https://doi.org/10.3390/su13042247

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error

and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Mayer, R. E. (2017). Using multimedia for e-learning. *Journal of Computer Assisted Learning*, *33*(5), 403–423. https://doi.org/10.1111/jcal.12197

Mayer, R. E. (2003). The promise of multimedia learning: using the same instructional design methods across different media. *Learning and Instruction*, *13*(2), 125–139. https://doi.org/10.1016/s0959-4752(02)00016-6

Mayer, R. E. (2020). Where is the learning in mobile technologies for learning?. Contemporary Educational Psychology, 60, 101824. https://doi.org/10.1016/j.cedpsych.2019.101824

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *American Psychological Association, Inc*, *86*(3), 389–401. http://libproxy.library.unt.edu:2067/ehost/pdfviewer/pdfviewer?vid=1&sid=cacecf36-cd5b-4ffe-9996-9b0332df5fb7%40pdc-v-sessmgr05

Mckeown, M. G., Beck, I. L., & Blake, R. G. K. (2009). Rethinking reading comprehension: a comparison of instruction for strategies and content approaches. *Reading Research Quaterly*, *44*(3), 218–253. https://doi.org/10.1598/RRQ.44.3.1

Meteyard, L., & Davies, R. A. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, 112, 104092. https://doi.org/10.1016/j.jml.2020.104092

Montag, J. L., Jones, M. N., & Smith, L. B. (2015). The words children hear: Picture books and

the statistics for language learning. *Psychological Science*, *26*(9), 1489–1496. https://doi.org/10.1177/0956797615594361

Nassaji, H. (2003). L2 vocabulary learning from context: strategies, knowledge sources, and their relationship with success in l2 lexical inferencing. *Tesol Quarterly*, *37*(4), 645–670. https://doi.org/10.2307/3588216

Oakhill, J. V., & Cain, K. (2012). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading*, *16*(2), 91–121. https://doi.org/10.1080/10888438.2010.529219

Paivio, A., & Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding? *Cognitive Psychology*, *5*(2), 176–206. https://doi.org/10.1016/0010-0285(73)90032-7

Pesco, D., & Gagné, A. (2017). Scaffolding narrative skills: a meta-analysis of instruction in early childhood settings. *Early Education and Development*, *28*(7), 773–793. https://doi.org/10.1080/10409289.2015.1060800

Proctor, C. P., Dalton, B., & Grisham, D. L. (2007). Scaffolding English language learners and struggling readers in a universal literacy environment with embedded strategy instruction and vocabulary support. *Journal of Literacy Research*, *39*(1), 71–93. https://doi.org/10.1080/10862960709336758

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

Rowe, M. L., Silverman, R. D., & Mullan, B. E. (2013). The role of pictures and gestures as nonverbal aids in preschoolers' word learning in a novel language. *Contemporary Educational Psychology*, *38*(2), 109–117. https://doi.org/10.1016/j.cedpsych.2012.12.001

Sadeghi, K., & Khezrlou, S. (2012). Glossing mode in self-regulated vocabulary learning, and its relationship with gender, age, and field of study. *Journal of Asia TEFL*, *9*(3), 51–74.

Salmerón, L., & García, V. (2011). Reading skills and children's navigation strategies in hypertext. *Computers in Human Behavior*, *27*(3), 1143–1151. https://doi.org/10.1016/j.chb.2010.12.008

Sampayo-Vargas, S., Cope, C. J., He, Z., & Byrne, G. J. (2013). The effectiveness of adaptive difficulty adjustments on students' motivation and learning in an educational computer game. *Computers & Education*, *69*, 452-462.

Schwanenflugel, P. J., Stahl, S. A., & McFalls, E. L. (1997). Partial word knowledge and vocabulary growth during reading comprehension. *Journal of Literacy Research*, *29*(4), 531–553. https://doi.org/10.1080/10862969709547973

Tonzar, C., Lotto, L., & Job, R. (2009). L2 vocabulary acquisition in children: Effects of learning method and cognate status. *Language Learning*, *59*(3), 623–646. https://doi.org/10.1111/j.1467-9922.2009.00519.x

Tseng, W. T., Dörnyei, Z., & Schmitt, N. (2006). A new approach to assessing strategic learning: The case of self-regulation in vocabulary acquisition. *Applied Linguistics*, *27*(1), 78–102. https://doi.org/10.1093/applin/ami046

Wolf, M., & Katzir-Cohen, T. (2009). Reading fluency and its intervention. *Scientific Studies of Reading*, *5*(3), 211–239. https://doi.org/10.1207/S1532799XSSR0503

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, *17*(2), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

Wu, J. Y. (2014). Gender differences in online reading engagement, metacognitive strategies,

navigation skills and reading literacy. *Journal of Computer Assisted Learning*, *30*(3), 252–271. https://doi.org/10.1111/jcal.12054

Yun, J. (2011). The effects of hypertext glosses on L2 vocabulary acquisition: A meta-analysis. *Computer Assisted Language Learning*, *24*(1), 39–58. https://doi.org/10.1080/09588221.2010.523285

Zou, D., Huang, Y., & Xie, H. (2021). Digital game-based vocabulary learning: where are we and where are we going? *Computer Assisted Language Learning*, *34*(5–6), 751–777. https://doi.org/10.1080/09588221.2019.1640745

Zucker, T. A., Justice, L. M., Piasta, S. B., & Kaderavek, J. N. (2010). Preschool teachers' literal and inferential questions and children's responses during whole-class shared reading. *Early Childhood Research Quarterly*, *25*(1), 65–83. https://doi.org/10.1016/j.ecresq.2009.07.001