Journal of Research in Reading
UKLA
The United Kingdom Literacy Association

*Journal of Research in Reading, ISSN 0141-0423*
*Volume 46, Issue 1, 2023, pp 1–21*
*DOI:10.1111/1467-9817.12411*





# The role of vocabulary components in second language learners' early reading comprehension

Martina Röthlisberger

Bern University of Teacher Education, Bern, Switzerland

Christoph Zangger

University of Bern, Bern, Switzerland and Ludwig-Maximilians-University Munich, Munich, Germany

Britta Juska-Bacher

Bern University of Teacher Education, Bern, Switzerland

**Background:** In countries with German as an official language, children with German as a second language perform overall worse in school than their German native speaking peers. This particularly affects written language skills, which require advanced language knowledge. The reasons are manifold, but one is prominent, namely poor vocabulary knowledge. Vocabulary, however, consists not only of the number of known words but also of a complex and hitherto under-researched lexico-semantic framework, which is referred to as vocabulary depth.
**Method:** In the present study, a sample of 373 children (322 German native speakers and 51 with German as a second language) was examined longitudinally in Grade 2 and 3 for their reading comprehension and vocabulary breadth and depth. Vocabulary depth was defined as relational and semantic word knowledge. Latent change score models on vocabulary breadth and depth development including reading precursor skills, non-verbal intelligence and SES were conducted.
**Results:** The children with German as a second language scored lower than the native German speakers in reading and vocabulary tests, with the difference in vocabulary depth being more pronounced than in vocabulary breadth. Importantly, the differences did not tend to diminish over time, which could have been expected. The differences in reading comprehension can be accounted for mainly by the variations in vocabulary, especially in vocabulary depth.
**Conclusions:** These findings emphasise the importance of including not only vocabulary breadth but also vocabulary depth in the context of reading research on children who do not speak the majority language.

**Keywords:** second language reading, vocabulary breadth, vocabulary depth

## Implications for Practice

- L2 children lag behind L1 children in reading comprehension.
- A large part of this variance is usually explained by vocabulary.
- Vocabulary, however, is not a homogeneous construct but includes at least two dimensions, vocabulary breadth and depth.
- The present study adopts a theoretically sound division of the construct of vocabulary into vocabulary breadth, relational and semantic word knowledge and relates these components to reading comprehension skills in L1 and L2 children of Grades 2 and 3.
- Structural equation models uncover a large influence of vocabulary and especially of relational word knowledge, substantiating the prominent role of the semantic network in L2 second graders' reading.
- Latent change score models show that the same constellation emerges 1 year later in Grade 3.
- Special teaching for bilingual children should not focus exclusively on the sheer number of words learned.
- The relations of the words in the mental lexicon and their semantic scope should be enhanced.
- These findings should also be taken into account in the teaching materials for L2 children.

Children who acquire German as a second language (GL2) face important disadvantages at school due to poorer language skills, such as deferral from school enrolment and lower performance in language-related subjects. As adolescents, they often attend the lowest level of the second stage of basic education (Kempert et al., 2016). In 2018, 23% to 36% of the 15-year-olds had a migration background in German-speaking countries. PISA 2018 revealed that adolescents from families speaking a non-majority language showed statistically significant poorer reading competences than their native peers (GL1; Austria: Höller et al., 2019; Germany: Weis et al., 2019; Switzerland: Petrucci et al., 2019). Poor reading skills limit participation in social and political life (Blossfeld et al., 2015) and pose an increased risk for depression and low self-esteem in adulthood (Eloranta et al., 2019).

On the other hand, GL1 children also show varying reading acquisition trajectories, which are shaped by general cognitive abilities and by specific precursor skills such as phonological awareness and decoding ability (e.g., Nation, 2008), with vocabulary often not adequately addressed (e.g., Ouellette, 2006). The present study emerges from a longitudinal project which investigated the relation between different vocabulary and reading skills in GL1 elementary school children. This project was expanded by a group of GL2 children, based on the understanding that the relationship between vocabulary and reading is of special interest in bilingual children (Proctor et al., 2012).

## Reading Acquisition

Early reading involves matching phonemes to graphemes and combine them into words, a process which is referred to as decoding (e.g., de Jong & van der Leij, 1999). Decoding is part of the precursor skills of reading, along with phonological awareness and rapid automatised naming. There is some evidence that letter-sound knowledge (Chiappe et al., 2002; Muter & Diethelm, 2001) and decoding (Jongejan et al., 2007; Verhoeven, 2000) have a comparable influence on early reading acquisition in L1 and L2 children. The ability to match phonemes to graphemes is apparently not affected by reading acquisition in a second language.

Both phonological awareness—the ability to perceive and manipulate phonemes as well as larger speech parts (Castles & Coltheart, 2004 for a review)—and rapid automatised naming (RAN; Denckla & Rudel, 1974) of visually presented items such as objects, colours, digits, or letters as quickly as possible have a high predictive value on reading acquisition (de Jong & van der Leij, 1999). In contrast, the evidence for the influence of phonological awareness on bilingual children's reading acquisition is contradictory. Studies from English-speaking countries show a transfer of phonological awareness from first to second language, resulting in a comparable influence in mono- and bilingual children (Jongejan et al., 2007; Muter & Diethelm, 2001). Studies from German areas however identified phonological awareness as less influential in L2 than L1 children (Duzy et al., 2013; Limbird et al., 2014).

Regarding RAN, it may be assumed that L2 children perform worse because access to lexical representations in the second language is slower. Various studies, though, have shown that they perform comparably or even better than their GL1 peers (e.g., Jongejan et al., 2007), probably due to the positive consequences of bilingualism (Bialystok, 2007). Moreover, the predictive force of RAN on reading efficiency in the highly transparent Italian orthography was stronger in L2 than in L1 children (Bellocchi et al., 2017).

However, literacy means not only decoding but also reading comprehension. According to the simple view of reading (Gough & Tunmer, 1986), word decoding and linguistic comprehension, defined as understanding spoken language, are the foundation of reading comprehension (Hjetland et al., 2020). In a more sophisticated framework, Perfetti and Stafura (2014) distinguish three constructs that form reading comprehension, that is, general cognitive resources, processes, and knowledge. Processes subsume decoding, meaning retrieval, integration of words into sentences or texts, that is, in a situation model, and meaning inference. Knowledge finally refers to linguistic, orthographic, and general knowledge. Alongside with grammar knowledge, the most important component of linguistic knowledge is vocabulary. If the meaning of a word is unknown, even a sound lexical process will not retrieve it.

In bilingual children's reading acquisition, it is challenging to discern the effects of bilingualism and the cross-linguistic influence of cognitive skills on phonological, orthographic, and metalinguistic knowledge as Bialystok (2007) points out in her seminal work. It is therefore crucial to control for cognitive and reading precursor skills to disentangle the effects of vocabulary, which has been shown to play a critical role in children's second language reading acquisition (August et al., 2005 for a review).

## Vocabulary

At school onset at about 7 years of age, GL1 children in Switzerland have a productive vocabulary of about 3000 to 6000 and a receptive vocabulary of about 10,000 to 14,000 words (Moser et al., 2005). Vocabulary development is shaped by the environment, especially the home literacy environment, which in turn is related to the socio-economic status (van Steensel, 2006).

The conception of vocabulary as vocabulary breadth (i.e., the number of known words) is widely used but not uncontroversial in its exclusivity. Anderson and Freebody (1981) already postulated an additional dimension, vocabulary depth, which considers the quality of the individual word entries in the mental lexicon as well as their interconnections. However, the definition of vocabulary depth varies (e.g., Cain & Oakhill, 2014; Schmitt, 2014). According to Read (2004), the first definition is the scope of semantic representations of words, that is, the precision of meaning (henceforth semantic word knowledge; Juska-Bacher & Jakob, 2014). The second definition is broader and includes also grammatical, phonological, and orthographic knowledge. Finally, the third definition addresses the network knowledge, the knowledge of the paradigmatic, syntagmatic, and formal links between words (henceforth relational word knowledge).

Vocabulary depth is thus characterised as a complex construct that cannot be adequately captured by tests surveying only single aspects (Cain & Oakhill, 2014; Yanagisawa & Webb, 2020). Moreover, the individual aspects are not distinct but overlap (Schmitt, 2014). To date, there is no consensus on how to operationalise them correctly. Semantic word knowledge is often assessed by word definition tasks and relational word knowledge by hypernym and antonym knowledge.

Vocabulary knowledge has an undeniable impact already on early reading. Children who know many words and have solid mental representations of words decode them more easily, and rich and sophisticated representations enhance reading comprehension (Nation, 2008; Ouellette, 2006).

In children learning to read in a second language this impact is even greater than in GL1 children. Reading comprehension relies on good vocabulary knowledge (Oakhill & Cain, 2012) and less on decoding (Melby-Lervåg & Lervåg, 2014). Additionally, gaps in vocabulary can negatively affect phonological representations, which are crucial for reading acquisition in L2 (Chiappe et al., 2002). Lervåg and Aukrust (2010) found that a pooled vocabulary breadth and depth measure in Grade 2 was exclusively predicting reading comprehension 12 and 18 months later. Verhoeven (2000) reported that in Grade 1 and 2 the differences in vocabulary between Dutch-speaking L1 and L2 children were marked and increased over the period from mid-first to late second grade. Moreover, the influence of vocabulary on reading was much stronger for the L2 children, affecting their reading comprehension negatively. Another longitudinal study from authors of the same group revealed similar results in Grade 3 and 4 with large differences in reading comprehension and vocabulary between the L1 and L2 group in all tests and on all timepoints, with relational word knowledge showing a large and over time increasing difference in favour of the L1 children. Moreover, reading comprehension in the L2 group was modulated directly and via oral text comprehension by vocabulary, but not in the L1 group (Droop & Verhoeven, 2003).

These results indicate that vocabulary depth should be considered in comparative studies with L1 and L2 children. Hutchinson et al. (2003) showed for children with English as L2 that Grade 2 vocabulary depth, measured as relational and semantic word knowledge, was the only predictor for reading comprehension in Grade 4. For German, detailed analyses of the impact of vocabulary are rare; only Limbird et al. (2014) reported a significant relationship of relational and semantic word knowledge on reading comprehension in Grade 2 and a significant effect of Grade 2 vocabulary on Grade 3 reading comprehension in Turkish-German L2 children, but not in the GL1 group.

However, there is a lack of longitudinal studies investigating the development of the individual vocabulary components and their relationship with reading comprehension using adequate statistical models—which was the aim of the presented study.

## This Study

The present study aims to uncover differences in the relationship between vocabulary and reading comprehension in second and third graders with German as L1 or L2, both simultaneously per grade level and longitudinally, and controlling for reading-specific precursor skills, socioeconomic status, and non-verbal cognitive performance by applying advanced statistical models. Particularly the roles of vocabulary breadth and depth are to be disentangled to investigate potential differences between children with German as L1 or L2 regarding their roles in reading comprehension. Vocabulary depth was defined broadly following Read's (2004) proposal as relational and semantic word knowledge. The three major questions were thus:

- (1) Is vocabulary the main predictor of reading comprehension differences in children with German as L1 (GL1) or L2 (GL2)?
- (2a) Do vocabulary breadth, relational and semantic word knowledge have a differential influence on reading comprehension?
- (2b) Are these possible influences different for GL1 compared with GL2 children?
- (3) How does this influence develop over the period from Grade 2 to Grade 3?

# Methods

## Participants and Design

The present work arose in the context of the longitudinal study 'The development of vocabulary and reading. An investigation in primary school' (https://p3.snf.ch/Project-173245) that examined the relationship between reading skills, different vocabulary dimensions, and precursor reading skills from first to third grade in 403 GL1 children. Given the high percentage of GL2 children in Swiss primary schools and the lack of information regarding the relationship between vocabulary and reading in bilingual children, an additional subsample of 54 children with German as a second language was included starting in Grade 2. All bilingual children came from classes that were already part of the study and were schooled under the same conditions as the GL1 children (Röthlisberger et al., 2021).

For the present study, a sub-sample of 376 children was available for the assessments in Grade 2 (t1) and Grade 3 (t2). The attrition rate of 7% is due to children changing schools or absence during the data collection. Three GL1 children had to be excluded due to incomplete data. The final sample consisted of 373 children, the remaining 322 GL1 speakers (177 girls, 145 boys, mean age at t1: 8 years, 0.5 months, *SD* = 4.6 months; mean age at t2: 9 years, 7.4 months, *SD* = 5.1 months) and 51 GL2 speakers (29 girls, 22 boys, mean age at t1: 8 years, 8.2 months, *SD* = 6.4 months; mean age at t2: 9 years, 9.1 months, *SD* = 6.8 months). The proportion of girls and boys in GL1 did not differ significantly from GL2, $\chi^2(1, N = 373) = 0.8$, $p = .880$. The larger standard deviation in age at t2 was caused by COVID-19, as the data collection had to be spread over a longer period. A description of the corresponding COVID measures can be found below.

The children were distributed among 40 school classes from agglomerations and urban or rural areas. The socio-economic status was coded as ISEI-08 score of occupation as defined by the International Standard Classification of Occupations ISCO (Ganzeboom et al., 1992), ranging between 10 and 89. It was significantly lower for GL2, $M = 38.55$, $SD = 16.21$, compared with GL1, $M = 57.39$, $SD = 14.33$, $\chi^2(56, N = 373) = 173.26$, $p < .001$. In all GL2 children, the native language, the time spent in Switzerland, language support lessons, as well as the most frequently used language in the family and during free time (i.e., outside of school and family) were surveyed using a standardised questionnaire from 'Sprachgewandt', a test tool for GL2 school children (Bayer et al., 2013). Of the 51 children, 44 were born in German-speaking countries, the others had been living in German-speaking countries for several years at the time of the survey ($M = 4.4$ years, $SD = 1$, range: 3 years—6 years). The children had various first languages ($N = 11$ Slavic, $N = 10$ Albanian, $N = 9$ Romance, $N = 7$ Iranian, $N = 7$ Ethio-Semitic, $N = 3$ Turkish, $N = 4$ diverse). All of them had been socialised in another language than German but spoke German in their free time. Some of the children spoke their native language and German at home ($N = 28$). All others spoke exclusively another language at home. A multivariate ANOVA showed that the children speaking German and another language did not differ from those speaking only their native language at home in terms of reading comprehension either at t1, $F(1, 49) = 1.18$, $p = .283$, $\eta^2 = 0.02$, or at t2, $F(1, 49) = 0.88$, $p = .354$, $\eta^2 = 0.02$.

## Materials

### *Reading*

Reading comprehension was tested with a standardised German reading test 'Ein Leseverständnistest für Erst- bis Siebtklässler—Version II' (ELFE-II; Lenhard et al., 2018). ELFE consists of three parts, a word, a sentence, and a text comprehension task. The word reading part contains 75 pictures accompanied by four written words each. The task is to select the appropriate word from graphemically and phonologically similar distractors. In the sentence reading part, 36 sentences with a cloze are presented. A suitable word for the cloze is to be selected from five alternatives. In the text comprehension task, 26 short texts followed by a question and four answers are presented. The task is to select the correct answer among four alternatives. The time limits for the word and the sentence task are 3 minutes each, for the text comprehension 7 minutes. The internal consistency between the three tasks was high, Cronbach's alpha was 0.92 and 0.90 for Grade 2 and 3 respectively in the GL1 group, 0.87 and 0.88 for the GL2 group.

### *Decoding*

Decoding fluency was determined with the pseudowords reading list (form A) of the 'Salzburger Lese- und Rechtschreibtest' (SLRT-II; Moll & Landerl, 2014). It contains 156 pseudowords ordered in ascending difficulty. The time limit is 1 minute.

### *Vocabulary Breadth*

Vocabulary breadth was measured by the German version of the 'Peabody Picture Vocabulary Test' (PPVT-4; Lenhard et al., 2015). In the PPVT, the target corresponding to an auditorily presented word must be selected from four pictures. The test was halved (only the odd items were used) due to time constraints. Accordingly, the ceiling was reached with four or more errors in a set.

### *Relational Word Knowledge*

Relational word knowledge was determined using the expressive subtest in its abbreviated version (items 16 to 55) of the 'Wortschatz- und Wortfindungstest' (WWT 6–10; Glück, 2011). The items for the target group of 7 to 9 years were presented as specified in the test manual. In the WWT, knowledge of semantic relations is determined with the help of pictures. In addition to the active naming of objects and actions ('What is he doing?' at the picture of a yawning lion), antonyms and hypernyms ('What are all these?' at the picture of different vegetables) are to be named explicitly and without retrieval assistance.

### *Semantic Word Knowledge*

To measure the semantic word knowledge, the 15 odd items of the definition task from the German version of the 'Wechsler Intelligence Test' (HAWIK-IV; Petermann & Petermann, 2007) were applied. In this task, the children are given a spoken word, the meaning of which they are supposed to describe. Because the scoring system of HAWIK is designed to measure verbal intelligence, customised evaluation criteria with 0 to 3 points were developed for a semantic analysis (Juska-Bacher & Röthlisberger, 2021). Responses were scored by two independent raters. With Cohen's kappa $\kappa = 0.79$, $SD = 0.1$, range 0.6–1 for Grade 2, and 0.81, range 0.7–0.92 for Grade 3, inter-rater reliability was satisfactory. In case of discrepancies in the scores, they were independently assessed by a third rater.

### *Phonological Awareness*

As a phonological awareness measure, three subtests from the German phonological awareness test 'Basiskompetenzen für Lese-Rechtschreibleistungen' (BAKO1–4; Stock et al., 2017) were administered.

**Vowel replacement.** In this test, the children are presented verbally with words and non-words, in which all /a/sounds are to be replaced by /i/. The standard version from BAKO with eight words and four pseudowords was applied.

**Word remainder.** The task is to delete the first phoneme of a word or non-word and to pronounce the remaining part. Four of the seven items are non-words.

**Phoneme blending.** The standardised task from BAKO with 11 items (five non-words) was applied. The children were asked to exchange the first and the second phoneme of the item and to pronounce the result.

***Rapid Automatised Naming.*** Rapid automatised naming (RAN) was measured with the standardised German test 'Test zur Erfassung der phonologischen Bewusstheit und der Benennungsgeschwindigkeit' (TEPHOBE; Mayer, 2016). The subtests colours, letters and numbers were used.

***Cognitive abilities.*** The subtest matrices from the German version of the Culture Fair Intelligence Test (CFT 20-R; Weiß, 2006) was applied in Grade 3. The matrices were selected because they have the highest correlation with the total value of the CFT.

## Procedure

The data were collected by the project collaborators and trained Masters-level students. Reading and cognitive measures were assessed in classes, with a duration of about half an hour. Vocabulary tests, phonological awareness, RAN, and decoding were applied in two individual sessions with a duration of about 20 minutes each, with the same test order within the sessions. Data were collected in spring 2019 for t1. In Grade 3, because of COVID, the testing period had to be extended from spring to the summer 2020. After having tested 37 GL1 children onsite, the individual sessions had to be conducted virtually. For this, one person from the project team was on site to provide the technical equipment and to apply the classroom part. To test the impact of the virtual assessments, a multivariate ANOVA with the results of the individual tests (vocabulary breadth, relational and semantic word knowledge, phonological awareness, and RAN) as dependent variables, and assessment method (virtual or onsite) as a between-subject variable was conducted on the data of the L1 children in Grade 3. The results showed that the assessment method had an impact on the results of the phonological awareness, $F(1, 320) = 7.19$, $p = .008$, $\eta^2 = 0.02$, with the children assessed onsite scoring higher ($M = 23.51$, $SD = 0.88$) than those assessed virtually ($M = 20.98$, $SD = 0.32$), especially for vowel replacement, $F(1, 320) = 4.46$, $p = .036$, and the phoneme blending, $F(1, 320) = 4.99$, $p = .026$, but less for word remainder, $F(1, 320) = 4.06$, $p = .045$. For all other tests, the assessment method did not play a significant role, all $F$s $< 1.7$, all $p$s $> .21$.

## Data Analysis

Because this study focuses vocabulary development, it was essential to use an approach accounting for the longitudinal structure of the data to identify the variation of the vocabulary development for the two groups GL1 and GL2. Latent change score models (LCSMs) conceptualise the difference between two subsequent measures within an autoregressive structure (Ghisletta & McArdle, 2012). The latent factor depicting the change between the subsequent measures can be subjected further to time-variant and invariant covariates. Moreover, it is essential to ensure measurement invariance of the latent constructs in LCSMs (Kievit et al., 2018). In the present study, this was the case for the vocabulary factor, comprising relational and semantic word knowledge and vocabulary breadth.

All analyses were conducted in R using lavaan (Rosseel, 2012). To account for potential violations of the assumed multivariate normality and for robust inference, maximum likelihood estimation with Huber–White standard errors and scaled test statistics was used (Rosseel, 2012).

## Results

### Descriptive Statistics

Table 1 contains means, standard deviations, and the results of unpaired *t*-test between the two groups of the raw scores through both grades. Data for non-verbal cognitive skills (CFT) were collected only in Grade 3. Standard deviations were generally higher for GL1 due to the larger sample size.

### Correlations

Due to the non-normal data distribution, non-parametric correlations of the standardised data were calculated. Table 2 contains the correlations between the single variables at t1, Table 3 at t2. The results of the GL1 group are below the diagonal, those of the GL2 group above. For the reading variables, only the T-value was considered because the following structural equation models rely on it as well.

The correlations clearly show a stronger relationship at both timepoints between decoding, phonological awareness, and RAN for the GL1 than for GL2 group. Contrariwise, the vocabulary variables are more closely related to GL2 children's reading performance.

### Structural Equation Models

To test measurement invariance in the vocabulary construct, a CFA (Figure 1) with equality constraints on the factor loadings across time was compared with a model with freely estimated loadings. Freeing these three parameters was associated with a $\chi^2$ difference of 4.49, $p$ = .210. Therefore, measurement invariance was assumed for the following modelling. The two latent factors fit the data well with maximum values of 1.00 for both CFI and TLI, a robust RMSEA with a 90% upper confidence interval at 0.045 and a SRMR of 0.027.

Subsequently, a latent change score model was estimated for vocabulary development between t1 and t2 (Figure 2). With a CFI and TLI of 0.97 and 0.95 respectively, a RMSEA of 0.08, and a SRMR of 0.04, the model fits the data well (Hu & Bentler, 1999). The latent change score, $\delta\eta$, reflects the vocabulary development between t1 and t2. Although the standardised intercept of 1.48 indicates a considerable average increase in 1 year for all children, GL2 affiliation had already a negative impact on vocabulary and on reading comprehension at t1. Moreover, a higher ISEI and CFT scores as well as a better reading score at t1 were positively associated with vocabulary. No differences between boys and girls were found, and no other covariate did significantly influence the development of vocabulary. With a standardised value of −0.357, the autoregressive effect of vocabulary at t1 on the development marginally missed significance at the 0.05 level. Thus, although those with better vocabulary skills at t1 experienced smaller gains, the disadvantages for the GL2 children persisted or increased in relation with their SES.

**Table 1.** Descriptive statistics and *t*-tests of differences between GL1 and GL2 groups in Grade 2 and 3

| | Grade 2 | | | | Grade 3 | | | |
|---|---|---|---|---|---|---|---|---|
| Group | All | GL1 | GL2 | Lang. Group | All | GL1 | GL2 | Lang. Group |
| Measure | *M* (*SD*) | *M* (*SD*) | *M* (*SD*) | *t* (Cohen's *d*) | *M* (*SD*) | *M* (*SD*) | *M* (*SD*) | *t* (Cohen's *d*) |
| Word reading | 32.33 (9.72) | 32.90 (9.78) | 28.73 (8.60) | 2.88** (0.46) | 46.50 (10.65) | 47.17 (10.55) | 42.24 (10.37) | 3.11** (0.47) |
| Sentence reading | 12.20 (5.87) | 12.66 (5.95) | 9.31 (4.32) | 4.84*** (0.64) | 19.65 (6.27) | 20.30 (6.14) | 15.59 (5.57) | 5.15*** (0.80) |
| Text reading | 7.78 (4.62) | 8.21 (4.62) | 5.06 (3.54) | 5.66*** (0.77) | 13.45 (5.04) | 13.90 (5.07) | 10.57 (3.85) | 5.48*** (0.74) |
| T-value reading | 47.53 (10.29) | 48.65 (10.15) | 40.45 (8.15) | 5.49*** (0.89) | 50.52 (9.92) | 51.39 (9.88) | 44.98 (8.28) | 4.39*** (0.70) |
| Vocabulary breadth | 150.02 (21.34) | 154.05 (19.21) | 124.57 (16.00) | 10.41*** (1.67) | 164.56 (24.62) | 169.42 (21.70) | 133.92 (19.40) | 11.01*** (1.72) |
| Relational word knowledge | 21.83 (7.91) | 23.86 (6.23) | 9.02 (4.76) | 19.76*** (2.35) | 25.05 (7.54) | 27.01 (5.75) | 12.63 (5.39) | 16.75*** (2.58) |
| Semantic word knowledge | 18.69 (6.52) | 19.96 (5.81) | 10.71 (4.88) | 10.78*** (1.72) | 21.72 (5.94) | 22.96 (5.16) | 13.88 (4.33) | 11.91*** (1.91) |
| Decoding | 30.10 (9.24) | 30.14 (9.40) | 29.88 (8.23) | 0.19 (0.03) | 35.82 (10.62) | 35.73 (10.73) | 36.37 (10.01) | −0.40 (0.06) |
| RAN colours | 0.91 (0.21) | 0.91 (0.21) | 0.94 (0.23) | −0.95 (0.14) | 1.01 (0.25) | 1.00 (0.25) | 1.06 (0.26) | −1.69 (0.24) |
| RAN letters | 1.63 (0.21) | 1.62 (0.37) | 1.68 (0.38) | −1.04 (0.16) | 1.86 (0.38) | 1.86 (0.38) | 1.86 (0.36) | 0.01 (0.00) |
| RAN numbers | 1.57 (0.36) | 1.57 (0.36) | 1.60 (0.34) | −0.58 (0.08) | 1.72 (0.39) | 1.70 (0.39) | 1.90 (0.41) | −3.41** (0.5) |
| Vowel replacement | 6.50 (3.33) | 6.77 (3.24) | 4.76 (3.36) | 4.08*** (0.61) | 7.48 (2.67) | 7.67 (2.59) | 6.24 (2.84) | 3.64*** (0.53) |
| Word remainder | 5.38 (1.69) | 5.41 (1.73) | 5.20 (1.44) | 0.83 (0.13) | 5.97 (1.36) | 5.94 (1.40) | 6.12 (1.07) | −0.85 (0.14) |
| Phoneme blending | 7.08 (2.92) | 7.24 (2.89) | 6.08 (2.92) | 2.66** (0.40) | 7.55 (2.86) | 7.66 (2.81) | 6.90 (3.11) | 1.75 (0.26) |
| CFT | | | | | 8.49 (2.51) | 8.69 (2.46) | 7.20 (2.49) | 4.03*** (0.61) |

*Note:* GL1: German as First Language, $N = 322$; GL2: German as Second Language, $N = 51$. Mean number of correct items with standard deviations in parentheses; *t*-values with Cohen's *d* in parentheses; RAN values: correct items per second.
*$p < .05$.
**$p < .01$.
***$p < .001$.

**Table 2.** Non-parametric correlations between the single measurements in Grade 2 (t1)

| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | T-value reading | | 0.50$^{***}$ | 0.55$^{***}$ | 0.34$^{*}$ | 0.59$^{***}$ | 0.20 | 0.42$^{**}$ | 0.25 | 0.33$^{*}$ | 0.19 | 0.41$^{**}$ |
| 2 | Vocabulary breadth | 0.26$^{***}$ | | 0.54$^{***}$ | 0.57$^{***}$ | 0.27 | 0.20 | 0.27 | 0.24 | 0.21 | 0.04 | 0.32$^{*}$ |
| 3 | Relational word knowledge | 0.34$^{***}$ | 0.55$^{***}$ | | 0.59$^{***}$ | 0.01 | 0.21 | 0.08 | −0.01 | 0.00 | −0.18 | 0.22 |
| 4 | Semantic word knowledge | 0.28$^{***}$ | 0.39$^{***}$ | 0.44$^{***}$ | | 0.09 | 0.26 | 0.27 | 0.22 | 0.17 | −0.05 | 0.25 |
| 5 | Decoding | 0.72$^{***}$ | 0.11 | 0.17$^{**}$ | 0.18$^{**}$ | | 0.18 | 0.55$^{***}$ | 0.44$^{**}$ | 0.49$^{***}$ | 0.45$^{***}$ | 0.42$^{**}$ |
| 6 | RAN colours | 0.19$^{***}$ | 0.04 | 0.03 | 0.16$^{**}$ | 0.20$^{***}$ | | 0.33$^{*}$ | 0.57$^{***}$ | 0.08 | 0.27 | 0.45$^{***}$ |
| 7 | RAN letters | 0.45$^{***}$ | 0.07 | 0.09 | 0.16$^{**}$ | 0.47$^{***}$ | 0.38$^{***}$ | | 0.55$^{***}$ | 0.12 | 0.36$^{**}$ | 0.39$^{**}$ |
| 8 | RAN numbers | 0.32$^{***}$ | −0.06 | −0.02 | 0.20$^{***}$ | 0.33$^{***}$ | 0.58$^{***}$ | 0.56$^{***}$ | | 0.00 | 0.29$^{*}$ | 0.37$^{**}$ |
| 9 | Vowel replacement | 0.35$^{***}$ | 0.14$^{*}$ | 0.33$^{***}$ | 0.20$^{***}$ | 0.26$^{***}$ | 0.09 | 0.18$^{**}$ | 0.16$^{**}$ | | 0.44$^{**}$ | 0.25 |
| 10 | Word remainder | 0.31$^{***}$ | 0.07 | 0.19$^{***}$ | 0.10 | 0.25$^{***}$ | 0.07 | 0.09 | 0.08 | 0.45$^{***}$ | | 0.51$^{***}$ |
| 11 | Phoneme blending | 0.43$^{***}$ | 0.06 | 0.22$^{***}$ | 0.11 | 0.39$^{***}$ | 0.19$^{***}$ | 0.20$^{***}$ | 0.16$^{**}$ | 0.50$^{***}$ | 0.51$^{***}$ | |

*Note:* GL1: $N$ = 322; GL2: $N$ = 51. Correlation coefficients are Spearman rank coefficients.
$^{*}p < .05$.
$^{**}p < .01$.
$^{***}p < .001$.

**Table 3.** Non-parametric correlations between the single measurements in Grade 3 (t2)

| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | T-value reading | | 0.45*** | 0.61*** | 0.53*** | 0.35* | 0.16 | −0.02 | 0.09 | 0.25 | 0.21 | 0.43** |
| 2 | Vocabulary breadth | 0.24*** | | 0.46*** | 0.50*** | 0.09 | 0.17 | −0.05 | 0.11 | 0.03 | 0.21 | 0.16 |
| 3 | Relational word knowledge | 0.36*** | 0.57*** | | 0.66*** | 0.02 | −0.04 | 0.02 | −0.13 | 0.04 | 0.15 | 0.15 |
| 4 | Semantic word knowledge | 0.25*** | 0.39*** | 0.52*** | | 0.19 | 0.00 | 0.10 | −0.06 | −0.04 | 0.11 | 0.12 |
| 5 | Decoding | 0.69*** | 0.16** | 0.16** | 0.14* | | 0.26 | 0.28* | 0.22 | 0.29* | 0.33* | 0.46*** |
| 6 | RAN colours | 0.15** | 0.01 | −0.02 | 0.10 | 0.21*** | | 0.45*** | 0.59*** | 0.10 | 0.06 | 0.16 |
| 7 | RAN letters | 0.39*** | 0.06 | 0.11* | 0.18** | 0.49*** | 0.40*** | | 0.58*** | 0.09 | 0.11 | 0.11 |
| 8 | RAN numbers | 0.32*** | 0.00 | −0.01 | 0.11* | 0.39*** | 0.42*** | 0.61*** | | 0.22 | 0.19 | 0.00 |
| 9 | Vowel replacement | 0.44*** | 0.20*** | 0.27*** | 0.14* | 0.28*** | 0.08 | 0.22*** | 0.16** | | 0.46*** | 0.31* |
| 10 | Word remainder | 0.23*** | 0.13* | 0.1 | 0.09 | 0.17** | 0.13* | 0.13* | −0.01 | 0.37*** | | 0.27 |
| 11 | Phoneme blending | 0.42*** | 0.22*** | 0.24*** | 0.16** | 0.35*** | 0.15** | 0.24*** | 0.11* | 0.50*** | 0.41*** | |

*Note:* GL1: $N = 322$; GL2: $N = 51$. Correlation coefficients are Spearman rank coefficients.
*$p < .05$.
**$p < .01$.
***$p < .001$.

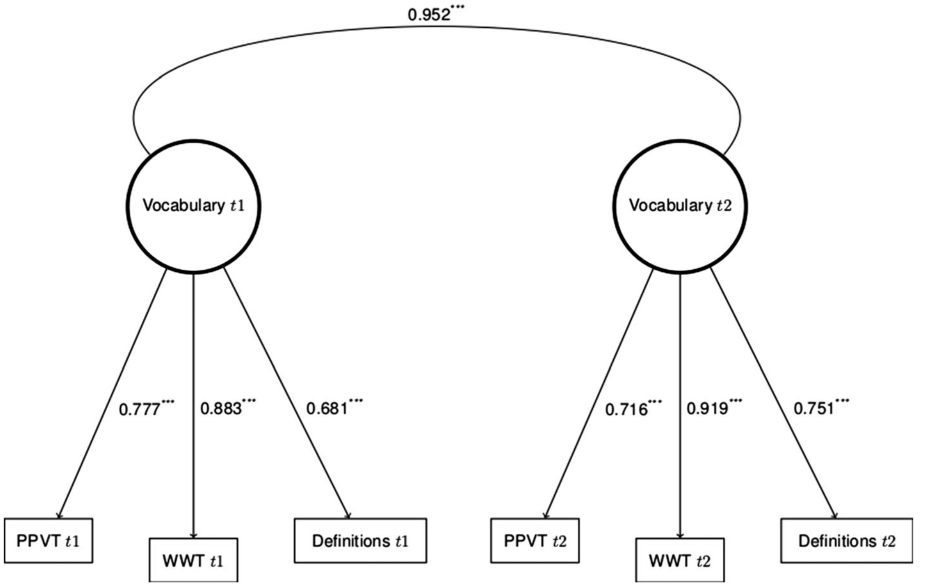


**Figure 1.** Vocabulary factors for both timepoints. ***$p < .001$

To account for the manifest vocabulary components (vocabulary breadth, relational and semantic word knowledge), the model in Figure 2 was re-estimated. Because the single models are just identified, the corresponding fit indices could not be calculated. Given the consistent pattern of effects observed in Figure 2, it can be assumed that the models fit the data well.

Table 4 shows that the covariates have similar influences as in the overall model. Moreover, the effects observed in Figure 2 seem to be driven by the association of the relational word knowledge with the covariates, substantiated by the high standardised loadings of the vocabulary factors on relational word knowledge. Although the GL2 children had lower scores on all vocabulary indicators at t1, the ISEI and the CFT score were positively associated with relational word knowledge, but not with other vocabulary components. Reading comprehension at t1, measured as T-values, was positively associated with all vocabulary indicators.

The effects in terms of differences between t1 and t2 were more pronounced than in the overall model. Higher scores in manifest vocabulary indicators at t1 were associated with smaller increases between t1 and t2. On the other hand, children of the GL2 group witnessed smaller increases in all vocabulary indicators. Generally, those with a higher ISEI increased their advantage in relational and semantic word knowledge. Those with a higher reading score at t1 also experienced larger gains in all vocabulary measures. Finally, there was a small positive effect of the CFT score on the development of vocabulary breadth, and no gender differences.

## Discussion

In the present study, potential differences in the effect of vocabulary breadth and depth on reading comprehension of GL1 and GL2 children in Grade 2 and 3 as well as their

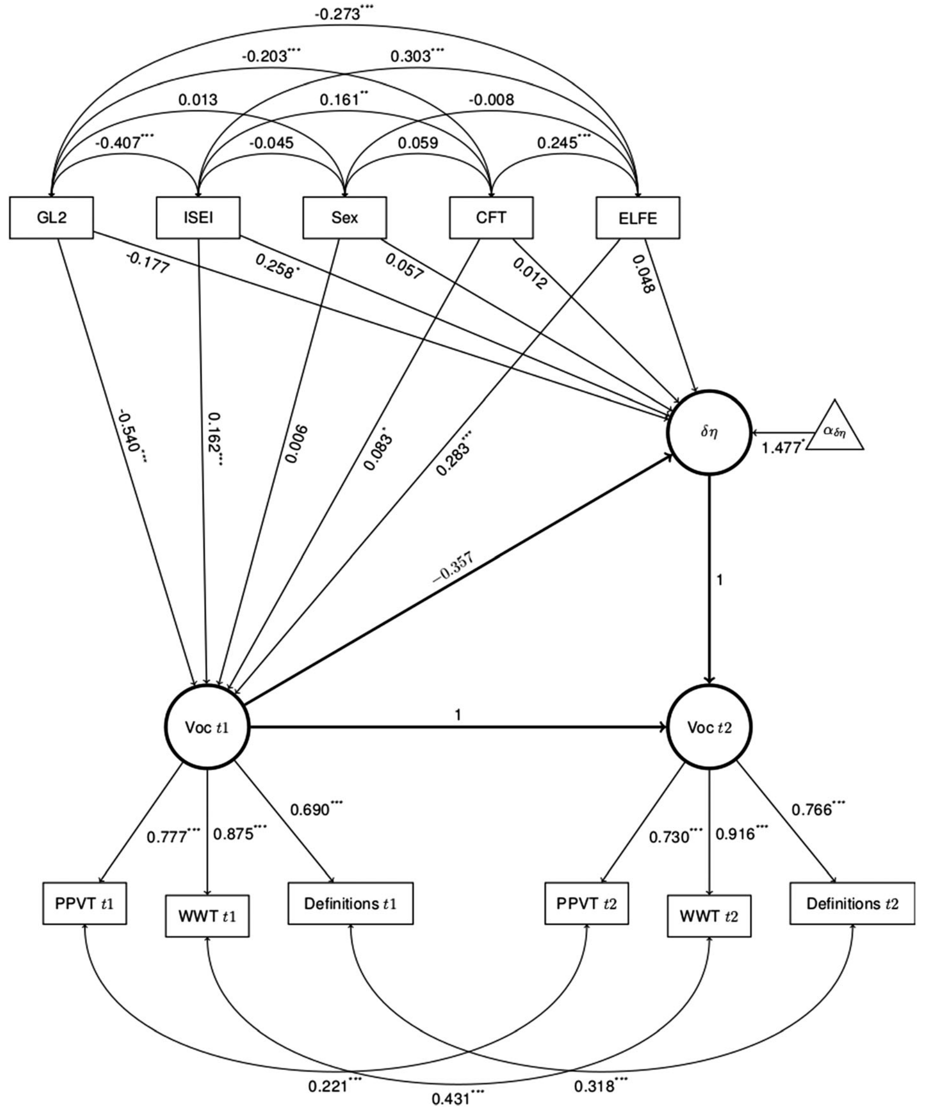


**Figure 2.** Latent change score model vocabulary. Circles represent latent constructs, rectangles are manifest variables, and the triangle represents the estimated intercept (average change) of the latent change score δη. All coefficients are standardised. $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$

development were explored. Moreover, the role of the hypothesised components of vocabulary depth, that is, relational, and semantic word knowledge, in reading comprehension was addressed. All analyses were conducted including the reading precursor skills and controlling for the SES and non-verbal cognitive abilities.

Large differences between the two groups were found in reading comprehension. At both timepoints, children of the GL1 group outperformed those of GL2 in word, sentence, and text reading with T-values differing at $p < .001$ and medium to high effect sizes. In the model, the GL2 affiliation was negatively associated with reading comprehension directly and indirectly through vocabulary.

**Table 4.** Coefficients latent change score models with manifest vocabulary indicators

| | Vocabulary knowledge | | | | | |
|---|---|---|---|---|---|---|
| | Vocabulary breadth | | Relational word knowledge | | Semantic word knowledge | |
| Variables | *β* | *Z* | *β* | *Z* | *β* | *Z* |
| Intercept (t2) | | | | | | |
| L2 | −0.365*** | −8.490 | −0.499*** | −14.933 | −0.370*** | −8.728 |
| ISEI | 0.088 | 1.899 | 0.167*** | 4.128 | 0.093 | 1.846 |
| Sex | −0.090* | −2.035 | 0.050 | 1.394 | 0.022 | −0.522 |
| CFT | 0.070 | 1.686 | 0.072* | 2.162 | 0.044 | 0.944 |
| ELFE | 0.216*** | 4.297 | 0.236*** | 6.578 | 0.230*** | 5.047 |
| Change score (δη) | | | | | | |
| t1 | −0.512*** | −8.923 | −0.670*** | −10.308 | −0.734*** | −12.616 |
| L2 | −0.244*** | −4.612 | −0.286*** | −4.970 | −0.247*** | −5.694 |
| ISEI | 0.101 | 1.871 | 0.182*** | 3.452 | 0.134** | 2.808 |
| Sex | −0.038 | −0.832 | 0.069 | 1.490 | −0.068 | −1.686 |
| CFT | 0.127* | 2.532 | 0.007 | 0.150 | 0.035 | 0.861 |
| ELFE | 0.119* | 2.353 | 0.129* | 2.503 | 0.137** | 3.134 |

*Note:* $N$ = 375; Maximum Likelihood estimation with robust (Huber–White) standard errors.
*$p < .05$.
**$p < .01$.
***$p < .001$.

A similar picture emerged for the vocabulary variables, where GL1 outperformed GL2 children in all measures at both timepoints, with the greatest difference in relational word knowledge, followed by semantic word knowledge. The high between-group differences with $p < .001$ and the very large effect sizes (all $d$s > 1) illustrate further the gap in vocabulary knowledge of the GL2 group.

In terms of phonological awareness, the essentially comparable performance between the groups is consistent with existing findings (Duzy et al., 2014; Jongejan et al., 2007; Limbird et al., 2014; Muter & Diethelm, 2001), except vowel replacement and phoneme blending at t1. Especially in vowel replacement, the difference might be due to the complex nature and instruction of the task, a presumption supported by the COVID induced mode effect of the online vs. onsite assessments, which was higher for these two tasks in GL1 children as well. Performance on the RAN also differed little between groups, mirroring the evidence of Chiappe et al. (2002), and Jongejan et al. (2007). Thus, GL2 children appeared by no means weaker in reading precursor skills than GL1 children.

The relation between phonological awareness and reading comprehension was slightly weaker for GL2, in line with the results of Duzy et al. (2014) and Limbird et al. (2014) but not with Jongejan et al. (2007) and Muter and Diethelm (2001). However, there is a caveat about the mixed results of the cited studies. Some of them included children coming from educated, bilingual environments, and being taught to read in both languages (Melby-Lervåg & Lervåg, 2014). The transfer of phonological awareness from L1 to L2 might be restricted to those children and not apply to children who are taught to read exclusively in their second language. Finally, the equivalent decoding results of the two groups fit in with those from previous studies (Droop & Verhoeven, 2003; Lervåg & Aukrust, 2010; Limbird et al., 2014; Verhoeven, 2000).

These results foreshadow what the correlations and structural equation models provide evidence for: the GL2 children's disadvantage in reading comprehension is associated with differences in vocabulary. Therefore, Research Question 1, 'Is vocabulary the main predictor of reading comprehension differences in children with German as L1 or L2?', must be answered positively, confirming the prominent role of vocabulary in the reading comprehension of children learning to read in L2 (Chiappe et al., 2002; Droop & Verhoeven, 2003; Lervåg & Aukrust, 2010; Verhoeven, 2000).

The second research question concerned the partitioning of the vocabulary into breadth and depth according to Anderson and Freebody (1981), Read (2004), and Juska-Bacher and Jakob (2014). The definition of vocabulary as a multidimensional construct is supported by the confirmatory factor analysis yielding an excellent model fit and proves to be legitimate (Cain & Oakhill, 2014; Schmitt, 2014).

In terms of the individual vocabulary components, the correlations for both groups and timepoints showed different patterns. Relational word knowledge exhibited the strongest association with reading comprehension in both groups. Research Question 2a, whether vocabulary breadth, relational and semantic word knowledge have a differential influence on reading comprehension in all children, can thus also be confirmed.

These correlations were even stronger in GL2 children, especially for vocabulary breadth. The low correlation of semantic word knowledge at t1 reflects a floor effect. The GL2 children had difficulties to define the 15 items and produced frequently invalid answers. Moreover, the high intercept values of the manifest vocabulary variables for GL2 in Table 4 show that Research Question 2b 'Are these possible influences different for GL1 compared to GL2 children?' must be answered positively. Belonging to group GL2 is associated with a negative effect on all vocabulary variables. Again, relational word knowledge stands out with higher values, a finding already reported by Limbird et al. (2014). Apparently, there are great differences between GL1 and GL2 in the ability to classify a word in the structure of the mental lexicon. Because the intercept coefficients of the vocabulary variables have a positive association with reading, an indirect negative effect of GL2 is also present in reading through the vocabulary.

A major concern was to delineate the development of vocabulary and its influence on reading comprehension. It could be hypothesised that the GL2 children are gradually catching up with their delay, which would translate into greater development and a higher $\delta\eta$-score in the model. However, the development of the vocabulary as a construct differed only marginally between the groups, a finding consistent with Hutchinson et al. (2003), Limbird et al. (2014), and Proctor et al. (2012). Moreover, the vocabulary components showed the same development pattern, no single component exhibited an accelerated progress in the GL2 group. Similar to Limbird et al. (2014), the level of the vocabulary components at t1 was the main determinant of the level at t2, and it was significantly lower in the GL2 group. Thus, the children were not able to catch up with the GL1 children. The answer to Research Question 3 'How does this influence develop […] from Grade 2 to Grade 3?' is therefore that the influence of vocabulary on reading has not developed positively in children of GL2. GL2 is strongly associated with a negative change score.

### Limitations and Future Research Directions

Apparently, the sheer quantity of lexical entries in the mental lexicon is not sufficient to induce adequate reading comprehension. It is rather the connections of these entries and the insight into their meaning that deepens the difference between GL1 and GL2 children. These aspects should therefore be given special attention in the teaching and support of GL2 children (Hutchinson et al. 2003).

Caution is required when interpreting vocabulary breadth measurement and semantic word knowledge. Halving the PPVT may have led to slightly less precise results. However, because the even-odd split-half reliability of the German PPVT is very high, $\alpha = 0.92$ and $\alpha = 0.97$ for Grade 2 and 3, respectively (Lenhard et al., 2015), this effect should be negligible. In contrast, the adapted scoring of the verbal intelligence subtest from the Wechsler Intelligence Test HAWIK-IV (Petermann & Petermann, 2007) has some weaknesses. The criteria catalogue and the rating by two raters or three in case of discrepancies are not free of arbitrariness. Given the impact of semantic word knowledge in reading comprehension when measured with a standardised test (Proctor et al., 2012), it is desirable that such a test is developed for German, including the definition of semantic knowledge as a continuum from not recognising a word to extensive contextual and decontextualised word knowledge (Cain & Oakhill, 2014; Schmitt, 2014; Yanagisawa & Webb, 2020). Deeper insights into L2 children's semantic word knowledge would also link to findings from studies of the influence of the home literacy environment in bilingual children (van Steensel, 2006). The large impact of socioeconomic status on all vocabulary aspects indicates that this is an important area of research. Systematic fostering of semantic word knowledge in regular and special classes for L2 children might counteract the negative influence of the SES and should be promoted.

Further limitations might have been induced by COVID-19. The phonological awareness results may have been biased and should be interpreted with caution. Moreover, the pandemic-related lockdown (in Switzerland limited to 1.5 months) may have slowed particularly the progresses of the L2 children. The present results thus would have to be confirmed in a regular school period. Moreover, it would be important to conduct longitudinal studies beyond Grade 3 to investigate whether increasing reading proficiency changes the relationship between reading and vocabulary in L2 children.

A last point to be discussed critically is the difference in the CFT to the benefit of the GL1 children, analogous to Duzy et al. (2014), but see Limbird et al. (2014), and which also appears in the standardisation data of the first part of the CFT, but not in the second (Weiß, 2006). Because here only a subtest from the first part was applied, the result reflects these findings. In a future study, this issue should be addressed by using subtests from both parts of the CFT.

As a conclusion, the present study shows that latent change score models can clarify the role of vocabulary depth in second language learner's early reading comprehension. Research on the relationships between literacy, vocabulary, and individual factors faced by children from non-majority language backgrounds is an essential contribution to providing these same children with better prospects for the future.

## Acknowledgements

We thank the Swiss National Science Foundation for funding the present study (100019_173245). This article has been published with the financial support of the University of Teacher Education, Bern. [Correction added on 16 November 2022, after first online publication: The funding details have been added in the preceding sentence, in this version.] Further, we thank Dr. Ladina Brugger for the valuable input and the joint work in the project. Finally, we want to thank participating teachers, the school administrators and especially the children.

## Conflict of Interest

The authors declare no conflicts of interest with regard to the funding source for this study. This study has been approved by the ethics committee of the University.

## Data Availability Statement

Data available on request due to privacy/ethical restrictions.

## References

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77–117). International Reading Association.

August, D., Carlo, M., Dressler, C., & Snow, C. (2005). The critical role of vocabulary development for English language learners. *Learning Disabilities Research & Practice*, 20(1), 50–57. https://doi.org/10.1111/j.1540-5826.2005.00120.x

Bayer, N., Berweger, S., & Moser, U. (2013). *Sprachgewandt. Kindergarten und 1*. Klasse. Lehrmittelverlag Zürich.

Bellocchi, S., Tobia, V., & Bonifacci, P. (2017). Predictors of reading and comprehension abilities in bilingual and monolingual children: A longitudinal study on a transparent language. *Reading and Writing*, 30(6), 1311–1334. https://doi.org/10.1007/s11145-017-9725-5

Bialystok, E. (2007). Acquisition of literacy in bilingual children: A framework for research. *Language Learning*, 57, 45–77. https://doi.org/10.1111/j.1467-9922.2007.00412.x

Blossfeld, H. P., Skopek, J., Triventi, M., & Buchholz, S. (Eds.) (2015). *Gender, education and employment: An international comparison of school-to-work transitions*. Edward Elgar Publishing. https://doi.org/10.4337/9781784715038

Cain, K., & Oakhill, J. (2014). Reading comprehension and vocabulary: Is vocabulary more important for some aspects of comprehension? *L'Année Psychologique*, 114(4), 647–662. https://doi.org/10.3917/anpsy.144.0647

Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition*, 91(1), 77–111. https://doi.org/10.1016/S0010-0277(03)00164-1

Chiappe, P., Siegel, L. S., & Wade-Woolley, L. (2002). Linguistic diversity and the development of reading skills: A longitudinal study. *Scientific Studies of Reading*, 6(4), 369–400. https://doi.org/10.1207/S1532799XSSR0604_04

de Jong, P. F., & van der Leij, A. (1999). Specific contributions of phonological abilities to early reading acquisition: Results from a Dutch latent variable longitudinal study. *Journal of Educational Psychology*, 91(3), 450–476. https://doi.org/10.1037/0022-0663.91.3.450

Denckla, M. B., & Rudel, R. (1974). Rapid 'automatized' naming of pictured objects, colors, letters and numbers by normal children. *Cortex*, 10(2), 186–202. https://doi.org/10.1016/S0010-9452(74)80009-2

Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first-and second-language learners. *Reading Research Quarterly*, 38(1), 78–103. https://doi.org/10.1598/RRQ.38.1.4

Duzy, D., Ehm, J.-H., Souvignier, E., Schneider, W., & Gold, A. (2013). Prädiktoren der Lesekompetenz bei Kindern mit Deutsch als Zweitsprache. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 45(4), 173–190. https://doi.org/10.1026/0049-8637/a000093

Duzy, D., Souvignier, E., Ehm, J. H., & Gold, A. (2014). Early decoding speed and later reading competencies in children with German as a second language. *Child Indicators Research*, 7(4), 787–804. https://doi.org/10.1007/s12187-014-9242-x

Eloranta, A. K., Närhi, V., Ahonen, T., & Aro, T. (2019). Does childhood reading disability or its continuance into adulthood underlie problems in adult-age psychosocial well-being? A follow-up study. *Scientific Studies of Reading*, 23(4), 273–286. https://doi.org/10.1080/10888438.2018.1561698

Ganzeboom, H. G. B., De Graaf, P. M., & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research*, 21, 1–56. https://doi.org/10.1016/0049-089X(92)90017-B

Ghisletta, P., & McArdle, J. J. (2012). Latent curve models and latent change score models estimated in R. *Structural Equation Modeling: A Multidisciplinary Journal*, 19(4), 651–682. https://doi.org/10.1080/10705511.2012.713275

Glück, C. W. (2011). Wortschatz- und Wortfindungstest für 6- bis 10-Jährige *(WWT 6–10)*. Elsevier.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *RASE: Remedial & Special Education*, 7, 6–10. https://doi.org/10.1177/074193258600700104

Hjetland, H. N., Brinchmann, E. I., Scherer, R., Hulme, C., & Melby-Lervåg, M. (2020). Preschool pathways to reading comprehension: A systematic meta-analytic review. *Educational Research Review*, 30, 100323. https://doi.org/10.1016/j.edurev.2020.100323

Höller, I., Lindemann, R., Wallner-Paschon, C., & Schaubmair, V. (2019). Kompetenzentwicklung im Kontext individueller und familiärer Faktoren. In B. Suchan, I. Höller, & C. Wallner-Paschon (Eds.), *PISA 2018. Grundkompetenzen am Ende der Pflichtschulzeit im internationalen Vergleich* (pp. 75–80). Leykam.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. https://doi.org/10.1080/10705519909540118

Hutchinson, J. M., Whiteley, H. E., Smith, C. D., & Connors, L. (2003). The developmental progression of comprehension-related skills in children learning EAL. *Journal of Research in Reading*, 26(1), 19–32. https://doi.org/10.1111/1467-9817.261003

Jongejan, W., Verhoeven, L., & Siegel, L. S. (2007). Predictors of reading and spelling abilities in first-and second-language learners. *Journal of Educational Psychology*, 99(4), 835–851. https://doi.org/10.1037/0022-0663.99.4.835

Juska-Bacher, B., & Jakob, S. (2014). Wortschatzumfang und Wortschatzqualität und ihre Bedeutung im fortgesetzten Spracherwerb. *Zeitschrift für Angewandte Linguistik*, 61(1), 49–75. https://doi.org/10.1515/zfal-2014-0016

Juska-Bacher, B., & Röthlisberger, M. (2021). Das Konstrukt Wortschatz: Dimension (en) Umfang und Tiefe? Empirische Ergebnisse aus der Unterstufe. *Bulletin Suisse de Linguistique Appliquée*, 113, 49–68. https://www.vals-asla.ch/fileadmin/user_upload/Journal/VALS_113_BAT_Definitif_SiteVALS.pdf

Kempert, S., Edele, A., Rauch, D., Wolf, K. M., Paetsch, J., Darsow, A., Maluch, J., & Stanat, P. (2016). Die Rolle der Sprache für zuwanderungsbezogene Ungleichheiten im Bildungserfolg. In C. Diel, C. Hunkler, & C. Kristen (Eds.), *Ethnische Ungleichheiten im Bildungsverlauf* (pp. 157–241). Springer VS. https://doi.org/10.1007/978-3-658-04322-3_5

Kievit, R. A., Brandmaier, A. M., Ziegler, G., van Harmelen, A.-L., de Mooij, S. M. M., Moutoussis, M., Goodyer, I. M., Bullmore, E., Jones, P. B., Fonagy, P., NSPN Consortium, Lindenberger, U., & Dolan, R. J. (2018). Developmental cognitive neuroscience using latent change score models: A tutorial and applications. *Developmental Cognitive Neuroscience*, 33, 99–117. https://doi.org/10.1016/j.dcn.2017.11.007

Lenhard, A., Lenhard, W., Segerer, R., & Suggate, S. (2015). *Peabody picture vocabulary test* (4th ed.). Pearson.

Lenhard, W., Lenhard, A., & Schneider, W. (2018). *ELFE II. Ein Leseverständnistest für Erst- bis Siebtklässler – Version II: Manual* (2nd ed.). Hogrefe Schultests.

Lervåg, A., & Aukrust, V. G. (2010). Vocabulary knowledge is a critical determinant of the difference in reading comprehension growth between first and second language learners. *Journal of Child Psychology and Psychiatry*, 51(5), 612–620. https://doi.org/10.1111/j.1469-7610.2009.02185.x

Limbird, C. K., Maluch, J. T., Rjosk, C., Stanat, P., & Merkens, H. (2014). Differential growth patterns in emerging reading skills of Turkish-German bilingual and German monolingual primary school students. *Reading and Writing*, 27(5), 945–968. https://doi.org/10.1007/s11145-013-9477-9

Mayer, A. (2016). *Test zur Erfassung der phonologischen Bewusstheit und der Benennungsgeschwindigkeit (TEPHOBE)* (3rd ed.). Ernst Reinhardt. https://doi.org/10.1007/978-3-658-15617-6

Melby-Lervåg, M., & Lervåg, A. (2014). Reading comprehension and its underlying components in second-language learners: A meta-analysis of studies comparing first-and second-language learners. *Psychological Bulletin*, 140(2), 409–433. https://doi.org/10.1037/a0033890

Moll, K., & Landerl, K. (2014). *SLRT-II. Lese- und Rechtschreibtest: Weiterentwicklung des Salzburger Lese- und Rechtschreibtests (SLRT)* (2nd ed.). Huber.

Moser, U., Stamm, M., & Hollenweger, J. (2005). Für die Schule bereit? Lesen, Wortschatz, und soziale Kompetenzen bei Schuleintritt. Sauerländer.

Muter, V., & Diethelm, K. (2001). The contribution of phonological skills and letter knowledge to early reading development in a multilingual population. *Language Learning*, 51(2), 187–219. https://doi.org/10.1111/1467-9922.00153

Nation, K. (2008). Learning to read words. *Quarterly Journal of Experimental Psychology*, 61(8), 1121–1133. https://doi.org/10.1080/17470210802034603

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

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology*, 98(3), 554–566. https://doi.org/10.1037/0022-0663.98.3.554

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

Petermann, F., & Petermann, U. (2007). *HAWIK IV Hamburg-Wechsler Intelligenztest für Kinder-IV Manual. Übersetzung und Adaptation der WISC-IV von David Wechsler*. Hogrefe.

Petrucci, F., Nidegger, C., & Roos, E. (2019). Lesekompetenz in der Schweiz. In A. B. Erzinger, M. Verner, N. König, F. Petrucci, C. Nidegger, E. Roos, et al. (Eds.), *PISA 2018. Schülerinnen und Schüler der Schweiz im internationalen Vergleich* (pp. 11–26). Konsortium PISA.ch.

Proctor, C. P., Silverman, R. D., Harring, J. R., & Montecillo, C. (2012). The role of vocabulary depth in predicting reading comprehension among English monolingual and Spanish–English bilingual children in elementary school. *Reading and Writing*, 25(7), 1635–1664. https://doi.org/10.1007/s11145-011-9336-5

Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language* (pp. 209–227). John Benjamins. https://doi.org/10.1075/lllt.10.15rea

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Röthlisberger, M., Schneider, H., & Juska-Bacher, B. (2021). Lesen von Kindern mit Deutsch als Erst- und Zweitsprache – Wortschatz als limitierender Faktor. *Zeitschrift für Grundschulforschung*, 14, 359–374. https://doi.org/10.1007/s42278-021-00115-w

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning*, 64(4), 913–951. https://doi.org/10.1111/lang.12077

Stock, C., Marx, P., & Schneider, W. (2017). *BAKO 1–4. Basiskompetenzen für Lese-Rechtschreibleistungen: ein Test zur Erfassung der phonologischen Bewusstheit vom ersten bis vierten Grundschuljahr: Manual* (2nd ed.). Hogrefe.

van Steensel, R. (2006). Relations between socio-cultural factors, the home literacy environment and children's literacy development in the first years of primary education. *Journal of Research in Reading*, 29(4), 367–382. https://doi.org/10.1111/j.1467-9817.2006.00301.x

Verhoeven, L. (2000). Components in early second language reading and spelling. *Scientific Studies of Reading*, 4(4), 313–330. https://doi.org/10.1207/S1532799XSSR0404_4

Weis, M., Müller, K., Mang, J., Heine, J. H., Mahler, N., & Reiss, K. (2019). Soziale Herkunft, Zuwanderungshintergrund und Lesekompetenz. In K. Reiss, E. Klieme, & O. Köller (Eds.), *PISA 2018. Grundbildung im internationalen Vergleich* (pp. 129–162). Waxmann.

Weiß, R. H. (2006). *CFT 20-R: Grundintelligenztest Skala 2-Revision*. Hogrefe.

Yanagisawa, A., & Webb, S. (2020). Measuring depth of vocabulary knowledge. In S. Webb (Ed.), *The Routledge handbook of vocabulary research* (pp. 371–386). Routledge.

**Martina Röthlisberger** is a scientific collaborator at the University of Teacher Education in Berne, Switzerland and at the Universiry Clinic for Chold and Adolescent Psychiatry and Psychotherapy at the University of Zurich. Her research interests cover Reading acquisition and development in L1 and L2 children, vocabulary, dyslexia, and reading trainings.

**Christoph Zangger** is a postdoc at the Institute of Sociology at the University of Berne and at LMU Munich. His scientific interests cover quantitative methods, social stratification and inequality, and sociology of education.

**Britta Juska-Bacher** is a professor of linguistics and language didactics at the University of Teacher Education in Berne. Her research interests cover reading, vocabulary, (historical) phraseology, and reading primers.

*Received 25 November 2021; revised version received 7 September 2022.*

**Address for correspondence:** Martina Röthlisberger, Institute of Research, Development and Evaluation, Bern University of Teacher Education, 3012 Bern, Switzerland. Email: martina.roethlisberger@phbern.ch