Syllable-Based Reading Improvement:

Effects on Word Reading and Reading Comprehension in Grade 2

Bettina Müller, Tobias Richter & Panagiotis Karageorgos

University of Würzburg

Accepted for publication in the journal Learning and Instruction

Author Note

Data collection was partly funded by the German Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF, grant 01GJ1004 B). We would like to thank all the students and teachers taking part in this study, as well as all the student assistants collaborating within the project.
Abstract

This study examined the effects of a syllable-based reading intervention for German second graders who demonstrated difficulties in the recognition of written words. The intervention focused on fostering word reading via syllable segmentation. The materials consisted of the 500 most frequent syllables typically read by 6- to 8-year-old children. The aims were to practice phonological recoding, consolidate orthographic representations of syllables, and routinize the access to these representations. Compared to children randomly assigned to a wait-list group, poor readers in the treatment condition showed significant improvements in standardized measures of phonological recoding, direct word recognition, and text-based reading comprehension after the 24-session intervention. Poor readers in the treatment condition also showed greater improvements in development of word recognition compared to children with efficient word recognition skills. The results provide evidence that a syllable-based reading intervention is a promising approach to increase struggling readers’ word recognition skills, which in turn will improve their reading comprehension.

Keywords: primary school children, reading comprehension, syllable-based intervention, visual word recognition
Syllable-Based Reading Improvement: Effects on Word Reading and Reading Comprehension in Grade 2

1 Introduction

Poor readers in primary school usually have difficulties with word reading. They often fail to especially make the transition from letter-by-letter phonological recoding to recognizing words directly by lexical access. The transition requires learning regularities of letter strings and acquiring orthographical representations that eventually lead to the recognition of whole words (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). The segmentation of words into syllables seems to be an important yet often overlooked intermediate step between phonological recoding and using orthographic representations of word forms in the mental lexicon for efficient word recognition (Ehri, 1995, 2005).

In the current study, we investigated the effects of a word recognition intervention based on the syllable as a major unit of reading processing in German second graders. In the following sections, we start with a brief sketch of how word recognition processes develop in primary school. We then argue that syllables play an important role in the development of word reading skills. The discussion of these approaches is used to derive hypotheses on how the syllable-based reading intervention examined in this study should affect word recognition processes and reading comprehension skills.

1.1 Syllables as Processing Unit in Word Recognition

The accuracy of word recognition for children learning to read in German is already close to ceiling at the end of Grade 1. Even poor readers manage to read a large proportion of regular (pseudo)words accurately at this point (Landerl & Wimmer, 2008; Huemer, Landerl, Aro, & Lyytinen, 2008), but huge individual differences in the efficiency of word recognition persist throughout primary school (Landerl & Wimmer, 2008; Richter, Isberner, Naumann, & Kutzner, 2012). Given the transparency of German orthography, children can
translate each letter of a (pseudo)word into a spoken representation by applying grapheme-
phoneme correspondence rules and by blending them into a phonological representation
(phonological recoding). However, this letter-by-letter phonological recoding is inefficient
and error prone, and it makes high demands on readers’ working memory (Coltheart et al.,
2001). A more efficient way to read words is by recognizing syllables that are holistically
translated into phonological representations. Ehri (1995, 2005) coined the term *consolidated
alphabetic phase* for the phase of reading development that follows the full alphabetic phase
in which grapheme-phoneme connections are formed. This consolidated alphabetic phase is
categorized by reliance on “letter sequences that symbolize blends of graphophonemic
units” (Ehri, 2014, p. 150), which most often are syllables (or components of syllables, such
as frequent multi-letter onsets and rhymes). According to this theory, the consolidated
alphabetic phase replaces the full alphabetic phase and is a regular phase in word reading
development. Syllable-based phonological recoding changes into the use of morphemes and
then eventually into the use of orthographic word forms for word recognition (i.e., word
recognition by lexical access) after children have encountered words through reading
practice. Thus, the recognition of (sub)lexical units serves as a bridge between the slow
phonological recoding and the faster direct recognition of whole words.

When learning to read in German, the syllable is likely to be crucial for helping
children to make the first step towards holistic word reading in Grade 2, whereas morphemes
might become more relevant in Grade 4 (Hasenhäcker & Schroeder, 2017). Successful
developing readers can thereby recognize an increasing number of words directly and
efficiently by mapping (sub)lexical units or whole word forms directly onto their lexical
entry to gain faster access to word meanings (meaning access).

Several findings hint at the role that syllables might play in the development of word
reading processes. For example, in a comparison of 16 European languages, Seymour et al.
(2003) found that the efficiency of pseudoword reading in Grade 1 depended on the syllable structure in contrast to the efficiency of reading real familiar words that could be recognized presumably via activating orthographic word forms in the mental lexicon. This finding suggests that recognizing words directly by accessing their lexical representation might play a role in recognizing familiar words even in Grade 1 because the direct recognition of words does not depend on syllabic complexity. In another study of poor-reading second graders learning to read in Finnish, a transparent orthography with simple syllabic structure, word recognition time increased with the number of syllables, implying that the poor readers rely primarily on phonological recoding (Hautala, Aro, Eklund, Lerkkanen, & Lyytinen, 2012). In contrast, good readers in this study responded equally fast to bisyllabic words compared to length-matched trisyllabic words, which suggests a reliance on holistic word reading via the lexical representations of orthographic word forms. Colé, Magnan, and Grainger (1999) investigated children at the end of Grade 1 who learned to read in French, which has an opaque orthography but a simple syllabic structure. The good readers in this study detected words faster when the word’s initial syllable appeared before the target word compared to a condition with a mismatching syllable. Same-aged poor readers, however, did not show this syllable compatibility effect. In sum, poor readers utilize the orthographic and syllabic structure of written words to a lesser extent than good readers. One possible explanation offered by Ehri (1995, 2005) is that poor readers are less able to perceive the regularity of syllables than good readers, making it more difficult for them to advance from the full to the consolidated alphabetic phase.

In line with this conjecture, studies investigating reading words with graphical syllable segmentation (e.g., inserting double spaces or hyphenation at syllable boundaries) hint on beneficial effects of graphical segmentation for poor readers’ accuracy (Scheerer-Neumann; 1981; Vellutino et al., 2004). In contrast, hyphenation has been shown to slow
down word recognition and sentence integration processes in good Finnish first- and second-grade readers (Häikiö et al., 2015, 2016). The authors concluded that marking syllables with hyphens disrupts word recognition in children who learn to read in a language with unambiguous syllable boundaries such as Finnish. Some children seem to use orthographic information already after two months of reading instruction and probably access more than one syllable simultaneously (Grainger & Ziegler, 2011). However, poor readers in Grade 2 benefit from the signaling of syllable boundaries, possibly because they still rely heavily on phonological recoding and because hyphenation increases the efficiency of word recognition by facilitating syllable-based instead of letter-based phonological recoding.

In sum, the role of syllables as processing units in word recognition changes with reading proficiency and likely depends on a language’s orthographic depth. The syllabic structure of German is complex and ambiguous (167 possible clusters of vocal and consonants, different vowel lengths, numerous initial consonant clusters; Seymour, et al., 2003; Würzner & Schroeder, 2015). This complexity poses problems for poor readers’ recognition of written words, leading to further deficits in the development of higher reading processes (Knoepke, Richter, Isberner, Naumann, & Neeb, 2014).

1.2 Fostering Word Recognition Skills in Poor Readers Using the Syllable

The importance of efficient word recognition skills for reading comprehension and the relevance of syllables as a developmental bridge between single-letter phonological recoding and direct word recognition through lexical access suggest that a systematic, syllable-focused training of word reading can remediate reading problems in primary school. Supporting this assumption, intervention studies have shown beneficial effects of repeated reading of frequent and infrequent syllables, and syllable segmentation on poor readers’ accuracy and fluency of word recognition in several orthographies such as Dutch (Berends & Reitsma, 2006; Wentink, Van Bon, & Schreuder, 1997), Finnish (Heikkilä, Aro, Närhi, &
Results of German readers also suggest that repeated syllable reading might be an effective way to foster word recognition skills. In a study by Huemer, Aro, Landerl, and Lyttinen (2008), a repeated reading intervention of high-frequency, word-initial consonant clusters with a vowel ending that partly corresponded with syllables improved reading speed and reading accuracy of words that included the practiced sublexical items for poor-reading second and fourth graders. However, no transfer effect on overall reading fluency occurred. In an intervention study with poor-reading fourth graders, we found an effect on word-reading fluency after a 16-session syllable-based reading intervention (Müller, Richter, & Karageorgos, 2017; see Ritter, 2010, for similar results with German third and fourth graders). It should be noted that phonological recoding and direct recognition of words through lexical access are highly correlated in German primary school children (.51 for accuracy, .65 for the degree of routinization of the underlying processes, Richter et al., 2012). The strong relationship suggests that deficits in phonological recoding are probably associated with deficits in building and using lexical representations of words and their orthographic forms. Thus, a word recognition intervention in German should convey association rules between graphemes and phonemes and spelling patterns to stress the alphabetic principle as well as components of repeated reading of syllables and words (Galuschka, Ise, Krick, & Schulte-Körne, 2014).

McArthur et al. (2015) investigated whether the effects of an intervention consisting of grapheme-phoneme correspondence rules and repeated reading of words depends on the order of the two treatment components. One group of poor English readers received the grapheme-phoneme intervention, followed by the fluency treatment. Another group received the two interventions in reverse order. Both conditions led to gains in word reading accuracy.
and speed, but training grapheme-phoneme rules first caused a larger gain in reading accuracy. This result is in line with results of cross-sectional and longitudinal studies showing that the fluency of word recognition mainly develops after reading accuracy has achieved a certain level (Karageorgos, Müller, & Richter, 2019; Fuchs, Fuchs, Hosp, & Jenkins, 2001; Juul, Poulsen, & Elbro, 2014) and that syllables become the salient unit of coding only after a child has mastered the alphabetic principle (Colé et al., 1999).

1.3 Efficient word recognition and reading comprehension

Good word reading skills, especially the efficient recognition of written words, are a precondition of good reading comprehension. Arguably, written texts can only be understood when the words in the text are recognized and understood reliably. However, the degree of routinization also plays a major role because of limited cognitive resources. If word recognition processes function inefficiently, fewer resources can be spent on higher level reading processes (such as reading strategies or inferences) that hamper reading comprehension (see, e.g., verbal efficiency theory, Perfetti, 1985). In fact, the efficiency of word recognition skills usually explains a large proportion of interindividual variance in reading comprehension (e.g., more than 50% in primary school children learning to read in German, Richter, Isberner, Naumann, & Neeb, 2013). In sum, written words must be recognized accurately but also fast (i.e., demanding few cognitive resources) to prepare the grounds for good reading comprehension.

1.3 The current study

The main objective of our study was to investigate treatment effects of a newly developed syllable-based reading intervention for German second graders with poor word reading skills. For the purpose of this study, poor readers are defined broadly as readers with poor to average reading skills that show room for improvement (but do not necessarily fulfill the much stricter criteria of a reading disorder). The intervention consisted of exercises on
grapheme-phoneme correspondence in the context of frequent syllables and repeated reading of syllables and words. The following research questions guided the study:

(1) Does the syllable-based intervention improve the phonological recoding of unfamiliar words and their ability to recognize words directly via lexical access?

(2a) Does reading comprehension also benefit from the syllable-based intervention?

(2b) Does an increase in word recognition skills mediate the expected positive effect on reading comprehension?

(3) Do the trajectories of word reading development differ between poor readers who received the training and good readers without specific training?

Against the background of the ambiguous syllabic structure in German and poor readers’ difficulties in reliably identifying syllabic units, we hypothesized that an intervention providing sufficient opportunities to practice syllable reading should strengthen the mental representations of syllables and frequent words. Therefore, both phonological recoding and direct word recognition should become more efficient, that is, more accurate and faster for children in the treatment condition. Considering the relevance of word recognition skills for higher-order comprehension processes (e.g., Perfetti, 1985; Perfetti & Hart, 2002), we further assumed a treatment effect on the poor readers’ comprehension skills mediated by the improved word recognition processes. Research Question 3 was based on a meta-analysis showing that children with deficits in word recognition processes often exhibit a Matthew effect (Pfose, Hattie, Dörfler, & Artelt, 2014). That is, the gradient of their increase is less steep than the gradient in children without these deficits. Hence, we explored whether the syllable-based reading training would change the typical pattern of trajectories by accelerating the development of poor readers compared to untrained good readers.
2 Method

2.1 Design and Procedure

We used an experimental pre/posttest/follow-up design with treatment and wait-list group and randomization at the class level (Figure 1). Children were first screened with the standardized German-speaking reading tests ProDi-L (Richter, Naumann, Isberner, Neeb, & Knoepke, 2017) for the efficiency of visual word recognition and ELFE (Lenhard & Schneider, 2006) for reading comprehension, and with subtests of the CFT 1-R (Weiß & Osterland, 2013) to assess their general intellectual abilities. From each class, children with the lowest scores in word recognition (all below percentile rank 50 on the ProDi-L norms for Grade 2) and average intelligence scores were chosen as participants. The word recognition score was computed as a mean composite score of the three ProDi-L subtests phonological recoding, direct word recognition, and access to word meanings when at least two subtests had no missing values. Children were clustered in groups of four to six. If a sufficient number of poor readers could not be recruited from one class, we ranked the children from two classes together. The groups of four to six children were randomly allocated at the class level to either the treatment or the wait-list group.

The groups in the treatment condition \(n = 14\) received the intervention between pre- and posttest. The groups in the wait-list condition \(n = 12\) received the same intervention between posttest and follow-up (experimental design with switching replications, cf. Shadish, Cook, & Campbell, 2002, pp. 146-147). The treatment consisted of 24 sessions and occurred twice a week for 45 minutes per training session in addition to the regular school curriculum. Most schools integrated the training sessions in the so-called individual learning time slots (i.e., a slot in the schedule when pupils work on different content depending on their strength and weaknesses) or offered the training as an additional subject for the poor readers. No limitations were placed on the content of the regular curriculum that poor
readers in the wait-list condition received between pre- and posttest. The training sessions were conducted by student research assistants who received standardized instructions and were supervised by the authors. Reading skills were assessed again at posttest and follow-up with the ProDi-L and ELFE.

At each measurement occasion, all children with parental permission for participation in the study were tested. To compare the development of reading skills between poor and good readers in Grade 2, we used a sample of untreated children with word recognition skills above percentile rank 50 at pretest (ProDi-L composite score, norms for Grade 2). These good readers received no intervention beyond the regular school curriculum.

2.2 Participants

We screened 549 second graders from 13 primary schools (35 classes) in the urban areas of Kassel and Fulda, Germany. The average age of the participants was 7.7 years ($SD = 7$ months). In total, 79 children with poor word recognition skills (25 females) were allocated to the treatment group. The means in the three word-recognition subtests in the treatment group were all at least 1.5 standard deviations below the average according to the ProDi-L norms for Grade 2 (see Table 1 for descriptive statistics of the raw scores). According to their parents, 25 had learned another language than German as their first language. First language information was missing for 36 children. The wait-list group comprised 71 poor readers (40 females; mean values in all word-recognition subtests were 1.5 standard deviations below the average according to the ProDi-L norms for Grade 2; see Table 1). Among the children in the wait-list group were 13 non-native speakers. First language information was missing for 46 children in this group because parents did not provide this information. Significant differences were found in mean intelligence scores ($min = 0$, $max =$
45) between children in the treatment group ($M = 20.05, SD = 7.78$) and the wait-list condition ($M = 23.99, SD = 8.64$), $F(1, 140) = 8.14, p < .01, \eta^2 = .06$. 

At the posttest, data from 13 children were missing (8 from the treatment condition, 5 from the wait-list group) because of illness or school change. At the follow-up test, the data of 69 children were missing (43 children in the treatment condition, 26 in the wait-list group) for the same reasons and because one school terminated the participation of their second graders in the study after the posttest. According to the teachers in this school, the testing was more time-consuming than they expected and could not be implemented into the class schedule for a third time in one school year.

Moreover, a sample of 36 good readers (19 females) with a percentile rank above 50 on the ProDi-L norms for Grade 2 (Table 1) and without missing values at pretest was selected as another comparison group of untreated good readers.\(^1\) These good readers showed significant higher mean intelligence scores ($M = 28.67, SD = 8.83$) than children in the treatment group, $F(1, 106) = 25.87, p < .001, \eta^2 = .20$, and children in the wait-list group, $F(1, 98) = 6.40, p < .05, \eta^2 = .06$.

- TABLE 1 ABOUT HERE -

2.3 Variables

2.3.1 Measurement of the efficiency of word recognition. The efficiency of visual word recognition processes was assessed with three subtests of the German-speaking

\(^1\) The relatively small number of good readers might be explained by the fact that the norming studies were done in the second half of Grade 2, whereas our study was conducted at the beginning of the school year. Consequently, the level of word recognition skills could have been underestimated. We rechecked all analyses with a sample of good readers based on the norms for Grade 1. These analyses led to the same results.
computer-based instrument ProDi-L that captures accuracy and response time for each item. (Richter et al., 2017). Each subtest has a dichotomous response format (yes/no). The accuracy (i.e., mean number of correct responses per subtest) reflects the reliability of the reading process targeted by each subtest. The response time, measured from the item’s onset to the press of the response button, reflects the degree of reading routinization. Mean response times were calculated based on the log-transformed response times of all items per subtest when at least three items per scale had valid responses. Reaction times that were three standard deviations or more below the item-specific mean or more than two standard deviations from the person-specific mean were recoded as missing values. Integrated test scores were computed for each subtest as a quotient of the mean accuracy divided by the mean log-transformed response time to reflect the efficiency of the targeted word recognition processes. The higher the integrated test value, the more efficient the focal reading process can be performed (Richter et al., 2017).

For each of the three points of measurement, different but parallelized versions of the subtests were used. All three subtests reached acceptable internal consistencies at all measurement points (i.e., Cronbach’s α ≥ .65, see Table A1 in the appendix for the reliabilities). *Phonological recoding* was assessed with a phonological comparison task based on 16 pairs of pseudowords at each point of measurement. The first pseudoword was presented auditorily followed by the second one presented visually on the screen. The children were required to decide whether the written pseudoword matched (e.g., nipu - nipu) or mismatched (e.g., kosawa - kosuwa) the spoken one. The pseudowords were congruent with phonological and orthographical rules of German and consisted of one to four open syllables with a simple vowel-consonant structure. *Direct word recognition* via lexical representations was assessed with a lexical decision task. The children were required to decide whether a string of letters formed a real word (e.g., Wasser / water) or a pseudoword
(e.g., Wassa). Sixteen items were used at each point of measurement, half of which were real words and the other half (orthographically and phonologically legal) pseudowords. These items varied systematically in length, frequency, and number of orthographical neighbors. The pseudowords varied in their similarity to actual German words. Access to word meaning was measured with a categorization task with 10 items. In each item, a category name was presented auditorily followed by a visually presented target word. Children were asked to decide whether the target word matched the category (e.g., category: Obst / fruits, target: Apfel / apple) or mismatched (e.g., category: Gewürze / spice, target: Suppe / soup).

2.3.2 Measurement of reading comprehension. The subtest text comprehension of the standardized reading test ELFE 1-6 (computer version, Lenhard & Schneider, 2006) was used to assess reading comprehension at the text level. The test assesses the ability to identify information in a text, generate anaphoric references across sentences, and draw local and global inferences. The test consists of 20 short, mostly narrative texts presented in randomized order with four multiple-choice items each. For example, a short five-sentence story about an elf performing magic and the ingredients needed for it was presented, followed by the question “Which title fits best with the story?” with four headings to choose from. The internal consistency of the reading comprehension scale was acceptable at all three measurement points (Cronbach’s α ≥ .65, see Table A1 in the appendix).

2.4 Syllable-based Word Recognition Training

The materials and standardized instructions of the intervention were designed by the authors in collaboration with a learning therapist (Müller, Richter, & Otterbein-Gutsche, in press). We developed a 24-session training based on word materials that were systematically selected according to the frequency of German written syllables in texts typically read by 6- to 8-year-old children (according to the childLex database, Schroeder, Würzner, Heister, Geyken, & Kliegl, 2015). All materials consisted of the 500 most frequent syllables.
Children were taught to divide words into syllables by swinging their arms while speaking a word aloud (Reuter-Liehr, 2008), mark syllables with arcs during silent reading, and find the vowel nucleus within each syllable. The rationale behind these tasks was to assist children in visually dividing words into syllables, which made syllables a salient processing unit.

In the first part of the training (Session 1-15) regularly spelled words with a maximum length of four open syllables were used to train phonological recoding in the syllable context. Special consideration was given to the correct pronunciation (e.g., differentiation between plosive sounds). From session eight onward, the words contained consonant clusters in the onset of the syllables. These consonant clusters were marked in the first exercises to alert participants to the onset of the syllable as a phonological unit because poor readers often fail to recognize syllable onsets (Stahl & Murray, 1994). In addition to words, few pseudowords were used because they do not have a stored representation in the mental lexicon and can only be pronounced accurately by using phonological recoding (Wentink et al., 1997). The second part (Session 16-24) mainly trained direct word recognition by providing children extensive practice in reading accurately and fast. The materials and exercises became increasingly complex by including, for example, irregular words, consonant clusters, diphthongs, or affixes. The number of syllables per word was escalated to eight, and sentences and short texts were also read in later sessions.

The 24 sessions followed the same tripartite structure and were conducted with the whole group. First, the whole group swung their arms while speaking words aloud. After each word, one of the children marked the syllables of the word with arcs on the board while reading the word aloud syllable by syllable. Second, children worked on a task independently (e.g., marking vowels, finding consonant clusters within words after the clusters of interest were pronounced with the group, combining prefixes and stems to words). In this phase, children read worksheets silently. Whispering while reading was
allowed. After reading the worksheets, each child read one word aloud syllable by syllable and marked vowels and syllable arcs on the board. During the whole session, the student research assistant who taught the group gave feedback on word pronunciation. They also checked the children’s worksheets and gave individual corrective feedback and support when necessary. Third, the whole group or pairs of students played one of the several syllable-based games. All games were based on reading single words aloud and mainly used the focal words from the session. The games included, for example, “syllable race” (i.e., moving a game character on a board according to the number of syllables in a word), “syllable bingo” (i.e., words were presented orally by the student research assistant and children were required to find the appropriate prefix for each word on their worksheet), or flash card reading (i.e., very brief presentation of words on cards to encourage fast decoding, Wentink et al., 1997; cf. Müller et al., 2017).

Overall, the scope of the intervention was the application and practice of syllable reading to strengthen the mental representation of syllables and frequent words to foster poor readers’ phonological recoding and direct word recognition skills. Word recognition was first trained for isolated words and later for words integrated in texts. The only rule that was taught explicitly was that every syllable contains a vowel. Thus, syllables are created around a vowel nucleus.

The median number of sessions attended by students was 22 (Range = 5-24). There was no special strategy to support the children when they missed a session. However, within the exercises, the words used in prior sessions were always repeated as well as rules and techniques for working with the words. Given that all treatment sessions followed the same structure, we assumed that getting back into the intervention after missing a session would be easy.
2.5 Fidelity Measures

To assess the integrity of the treatment, the student research assistants rated the implementation and difficulty of the exercises on a four-point rating scale (ranging from 1 to 4) after each session. Descriptive statistics were calculated when at least half of the treatment sessions were rated (higher values indicate higher agreement). The results indicate that the exercises were mainly implemented as described in the training manual in both conditions (treatment group: $M = 2.96, SD = 0.50$; wait-list group: $M = 2.79, SD = 0.77$) and that the difficulty of the exercises was in line with students’ reading skills (treatment group: $M = 3.01, SD = 0.62$; wait-list group: $M = 2.73, SD = 1.04$).

Moreover, the behaviors of four student research assistants selected at random who taught a wait-list group were recorded at the same session. The videos were coded by two independent judges on a 5-point rating scale to decide whether the parts of the session were implemented as described in the manual (0 = not assessable to 4 = full agreement). The treatment integrity ratings of the judges showed high interrater-reliability ($ICC_{unjust} = .88$) and indicated that the exercises were mainly implemented as intended ($M = 2.94, SD = 1.57$). The two judges rated the research assistants to provide valuable feedback and support to the children ($M = 3.50, SD = 0.88$), with moderate interrater reliability ($ICC_{unjust} = .55$). Note that the ICCs might be low because the ratings are close to ceiling with small variances. Discipline and cooperation was rated as substantial ($M = 2.79, SD = 1.20$) with good interrater reliability ($ICC_{unjust} = .80$). On average, 95% of the total time in each session was used to implement the exercises. In sum, the fidelity measures suggest an acceptable integrity of the treatment.

2.6 Data Analysis

Given that the treatment groups were formed at the class level, we first tested for the presence of clustering effects in the data. We calculated the intra-class correlations (ICC)
with random intercept multilevel models for all three word-recognition subtests (time of measurement nested within students and students nested within classes, Raudenbush & Bryk, 2002). We assumed from the low ICCs ($\rho \leq .05$) that the clustering effects had no impact on the data. Thus, we proceeded with regular multiple regression analyses.

Significance level was set at .05 (one-tailed because all hypotheses were directional, e.g., Maxwell & Delaney, 2000, p. 144). The visual inspection of standardized residuals versus unstandardized predicted values revealed that the assumptions of linearity, normality, homoscedasticity, non-multicollinearity of the predictors and independence of residuals were not violated in any of the models (Cohen, Cohen, West, & Aiken, 2003, Chapter 4 and 10).

In each model, about one to eight cases with high discrepancy values (cutoff: $\pm 2.0$ for the externally studentized residuals) and high global influence (cutoff: 1.0 for Cook’s $D$) were excluded from the analysis (Cohen et al., 2003, Chapter 10).

To address Research Questions 1 and 2a, we ran separate regression models for each subtest of word recognition and for reading comprehension at posttest and follow-up. All analyses were performed with the raw scores. The dummy-coded treatment condition (with the wait-list group as reference) was used as predictor and word recognition processes and reading comprehension were used as outcome variables ($z$-standardized). To control for pre-training differences, the pretest score corresponding to the outcome variable and intellectual ability were entered as covariate ($z$-standardized), which has more statistical power compared to repeated-measures analysis of variance (Rausch, Maxwell, & Kelley, 2003). All predictors were entered simultaneously into the models. We checked all analyses for interactions to investigate whether the treatment effects varied depending on the students’ pretest word recognition scores (moderated multiple regression analysis with interaction terms of treatment condition and the subtests of word recognition at pretest as predictors, e.g., Aiken & West, 1991). However, no significant interaction effects occurred. Thus, we
report the results of multiple regressions with the pretest values of the outcomes as predictor but without interactions. This approach is equivalent to an analysis of covariance with the pretest score as covariate. As effect sizes for significant effects in the multiple regressions, we report the increment of explained variance ($\Delta R^2$), the quotient of the explained variance $R^2$ divided by the unexplained variance (Cohen, 1988, Chapter 9)

$$f^2 = \frac{R^2}{(1 - R^2)}$$

and Cohen’s $d$ (difference between posttest means of the treatment and the wait-list group adjusted for pretest differences, divided by the pooled standard deviation). Post hoc power analysis was performed with G*Power (linear multiple regression: fixed model; Faul, Erdfelder, Buchner, & Lang, 2009).

We performed a multiple imputation and compared the results obtained with the imputed data files with those of the regression analysis with listwise deletion to check whether the results of the analysis were systematically biased by the missing data (33.32% of all data points). We assumed the missing values to be missing at random (Rubin, 1976), mainly caused by illness, change of school, or early termination of the study. No systematic differences were found in word recognition skills at pretest between the children from the school that dropped the follow-up and children from the other schools. Thus, the assumption seems justified that the missing values of students from this school were also random. In addition, from one to eight values were classified as outliers in each model. The variables included in the imputation models were intellectual ability (CFT-1), the integrated values of the three word-recognition subtests (ProDi-L), and the reading comprehension scores (ELFE) for pre- and posttest and follow-up. Five datasets were estimated with the automatic imputation method. All parameter estimates except the phonological recoding at follow-up differed by less than one tenth from the results of the original data set, without any changes.
in significance. This comparison indicated that no systematic bias existed from the missing values. Thus, we report the results of the regression analysis with the original data and listwise deletion. However, for phonological recoding at the follow-up test, the results for the parameter estimates with listwise deletion and with multiple imputation are reported.

To address Research Question 2b, a mediation analysis (Hayes, 2013, Chapter 4) was conducted to test the assumption that an increase in word recognition mediates the expected treatment effect on reading comprehension. Reading comprehension at the posttest was used as the outcome variable, a composite score of the three posttest word recognition processes at the posttest was used as mediator (mean of the three ProDi-L subtests, z-standardized), and treatment condition was used as predictor controlling for pretest differences in reading comprehension, word recognition processes, and intellectual ability. Given the fact that outcome and mediation variables were both measured at the same time, the lack of temporal precedence of the mediator potentially weakens conclusions about a causal relationship between word recognition and reading comprehension. Therefore, we also estimated an alternate mediational model to test whether the reversed relationship holds (Moscovitch, Hofmann, Suvak, & In-Albon, 2005) by entering word recognition as the outcome variable and reading comprehension as the mediator. This allows the comparison of the regression weights of both mediation models to identify the causal sequence between the development of word recognition and reading comprehension to rule out alternative explanations of the relationship.

To address Research Question 3, we compared the development of word recognition skills of the poor readers in the treatment condition with an untrained group of children from the same classes who scored above average on word recognition skills at pretest as a reference group. We ran multiple regression models with the difference score of word recognition at posttest minus the pretest scores and with word recognition at follow-up
minus the pretest (composite scores, \(z\)-standardized) as the outcome variables (Judd, Kenny, & McClelland, 2001). The treatment condition (dummy coded with the good readers as reference) and intellectual ability were entered as predictors.

### 3 Results

Intercorrelations for all variables separated by treatment condition are summarized in Table 2.

- **TABLE 2 ABOUT HERE** -

#### 3.1 Comparison of the Treatment Group and the Wait-list Group

**3.1.1 Effects on word recognition.** To investigate whether the syllable-based training had positive effects on the efficiency of word recognition (Research Question 1), we compared the poor readers in the treatment group and the wait-list group at the posttest. As expected, the results of the multiple regression analysis with the processes of word recognition at posttest as outcomes indicate a significant treatment effect for phonological recoding and direct word recognition (see Table 3 for the estimates). No treatment effect occurred for meaning access.

- **TABLE 3 ABOUT HERE** -

The follow-up test was conducted three months after the intervention in the treatment condition and immediately after the wait-list group received the same intervention. No significant effects were found for direct word recognition \((B = .28, SE = .23, ns)\) and meaning access \((B = -.13, SE = .23, ns)\). Thus, after children in the wait-list group were also trained, any differences between the treatment and the wait-list group in these two variables disappeared. However, the model revealed a significant treatment effect for phonological recoding \((B = .46, SE = .20, p < .05, \Delta R^2 = .02)\), suggesting that the efficiency of phonological recoding further improved in the treatment condition even after the intervention had ended. The same analysis with the multiple imputation datasets, however,
showed no significant treatment effect for phonological recoding at follow-up \( (B = .30, SE = .22, ns) \). Hence, this effect should be interpreted with caution.

3.1.2 Effects on reading comprehension. To investigate treatment effects on the text level (Research Question 2a), we ran a multiple regression model with the posttest reading comprehension as the outcome variable. The analysis revealed a significant treatment effect in line with our expectations (Table 3). Moreover, there was no significant treatment effect on reading comprehension at the follow-up \( (B = .13, SE = .19, ns) \). On average, children in the wait-list group reached the same level of reading comprehension than children in the treatment condition after both had received the intervention (Table 1).

3.1.3 Mediation analysis. We assumed that the syllable-based intervention improved the routinization of the poor readers’ basic reading processes, making cognitive resources available for comprehension processes at the text level (Research Question 2b). A bootstrap analysis with 5000 samples (Shrout & Bolger, 2002; Hayes, 2013) revealed a significant indirect effect from the treatment condition via word recognition on reading comprehension \( (Est. = .10, 95\% \text{ CI } [.01, .24]) \). Figure 2 provides the estimates for the mediation model and illustrates that the direct effect from the treatment condition on reading comprehension disappears after including word recognition as a mediator in the model.

In the alternative mediational model with word recognition as the outcome variable and reading comprehension as the mediator, a bootstrap analysis with 5000 samples showed a significant indirect effect from the treatment condition via reading comprehension on word recognition \( (Est. = .08, 95\% \text{ CI } [.01, .21]) \). However, the indirect effect was numerically smaller than the effect from the treatment condition via word recognition on reading comprehension. Moreover, the direct effect from the treatment condition on word recognition remained significant after including reading comprehension as a mediator in the
model \((B = .32, SE = .18, p < .05, \text{one-tailed})\). In sum, these results support the assumption that the syllable-based training seems to improve reading comprehension by improving the efficiency of word recognition.

3.2 Comparison of the Treatment Group with the Untreated Group of Good Readers

The results of the multiple regression analysis comparing the development of word recognition skills of the poor readers in the treatment condition with an untrained group of good readers as the reference group (Research Question 3) revealed a significant treatment effect \((B = .55, SE = .15, p < .001, \Delta R^2 = .14)\). Thus, the slope of the efficiency of word recognition between post- and pretest seems to be steeper in the poor readers who received the intervention. The analysis of the difference score of word recognition between the follow-up and the pretest also revealed a similar effect \((B = .99, SE = .23, p < .001, \Delta R^2 = .28)\). Hence, the efficiency of word recognition processes increased faster in poor readers even after the training had ended (Figure 3).

In a second multiple regression model, we compared the development of word recognition skills of the poor readers in the wait-list condition with the untreated group of good readers. Significant effects were found for the difference scores of word recognition at posttest minus pretest \((B = .36, SE = .15, p < .05, \Delta R^2 = .06)\) and the difference between follow-up and pretest \((B = .54, SE = .17, p < .05, \Delta R^2 = .16)\). Poor readers in the wait-list group showed a faster development in visual word recognition than untreated children with above-average reading skills.

4 Discussion

The goal of this study was to investigate the effects of a syllable-based reading training based on the 500 most frequent German syllables in 6- to 8-year-old children for poor readers in Grade 2. Standardized measures of reading skills were collected before and
after a 24-session intervention within treated and untreated poor and good readers to compare the reading development above one school year. The liberal criterion of percentile rank 50 in the word recognition measure at pretest was used as the cut-off to distinguish between poor (i.e., below-average) and good (i.e., above-average) readers (cf. Ise et al., 2012).

The reading intervention improved the efficiency of poor readers’ phonological recoding and direct word reading processes. Children in the treatment condition outperformed the same-skilled children in the control group at the posttest in the efficiency of both types of processes. Given the missing interactions, those effects seem to be independent of the students’ pretest word recognition skills. The largest improvement was obtained for phonological recoding, which was explicitly practiced in the first part of the intervention and, as a by-product, exercised in the second part via the accurate pronunciation of words. The treatment effect on direct word recognition via lexical representations can be interpreted as a shift from letter-by-letter processing to the recognition of syllables or in terms of Ehri’s framework (1995, 2005) reaching the consolidated alphabetic phase. Thus, in line with previous studies investigating treatment effects of syllable-based interventions for poor readers in Grade 2 (Berends & Reitsma, 2006; Heikkilä et al., 2013; Wentink et al., 1997), when syllables are the salient processing unit, the intervention seems to foster reading via orthographic comparisons also in German.

At the follow-up test, after the wait-list group had also received the intervention, the average word reading skills of children in the treatment and the wait-list group were at the same level (Table 1), with the exception of a small treatment effect on phonological recoding. However, given the lack of effect in the analysis based on multiple imputation and an effect only in the analysis with listwise deletion, the result might be due to a biased parameter estimation because of missing values at the follow-up.
In line with our expectation, we found a significant treatment effect on text-based reading comprehension at the posttest that was fully mediated by the improvements in word recognition. The fact that single word decoding was embedded in sentence and text reading activities in the second part of the intervention might have promoted this transfer effect (cf. McCandliss et al., 2003). Reliable and easily accessible representations of words in the mental lexicon allow readers to allocate their working memory capacity to higher-order comprehension processes (Perfetti & Hart, 2002). The indirect effect from the treatment condition via word recognition on reading comprehension is in line with this assumption. In contrast, the access to word meanings was not improved by the intervention because most likely the intervention excluded any kind of vocabulary training. Integrating morphological awareness exercises in the intervention might be a promising method to foster lexical access. Morphemes, in contrast to syllables, are meaningful units that are relevant for the development of word reading skills beyond Grade 2 (e.g., Deacon & Kirby, 2004; Hasenhäcker & Schroeder, 2017; Nagy, Berningner, & Abbott, 2006).

The comparison of the development of word reading skills in good and poor readers indicates that the participation in the intervention has the potential to counter the Matthew effect. The pattern of trajectories showed the steepest increase in word reading from pre- to post- and follow-up test for poor readers in the treatment condition (following a compensatory pattern, Pfost et al., 2014). However, the untrained poor readers in the waitlist group also showed a faster development than the good readers, even without an intervention. Similar effects were reported in previous intervention studies with German second graders (Müller, Krizan, Hecht, Richter, & Ennemoser, 2013) and third graders (Groth, Hasko, Bruder, Kunze, & Schulte-Körne, 2013). On average, the steepest increase in the efficiency of word recognition processing can be observed in Grades 1 and 2 when children move from letter-by-letter reading to direct word recognition (Richter et al., 2012).
The comparison of good and poor readers in the present study suggests that the good readers had taken this step already in Grade 1, resulting in efficient word-reading skills already at the pretest in these children (Landerl & Wimmer, 2008). Poor readers, in contrast, seem to take the step of further routinization later in Grade 2, leading to a steeper gradient. However, the syllable-based reading intervention seems to further accelerate the development of word recognition skills. Children in the treatment condition showed the strongest growth in word recognition skills from the pretest to the follow-up test.

In sum, the results show that the shallow orthography and complex syllabic structure of German provide a favorable linguistic context for a training that capitalizes on the sublexical units of syllables. A distinctive feature of the intervention used in this study is the purposefully selected word material that contained the 500 most frequent syllables to enable poor readers to form and consolidate orthographic representations of the most frequent multi-letter representations and routinize access to these representations. An intervention study with poor Finish-speaking readers in Grades 2 and 3 also suggests positive effects of teaching infrequent syllables (Heikkilä et al., 2013). Investigate effects of the syllable-based treatment using a combination of high and less frequent syllables would be informative. According to Grainger et al. (2011, 2012), skilled readers recognize words using the frequency of letter combinations. The results of the Finish-speaking sample suggest that even poor readers are able to reach this level of word recognition.

We should emphasize, however, that our results do not allow for the conclusion that making syllables the salient unit for processing definitely improved word reading skills. The training included ample opportunities to practice reading words accurately and fast. Further research should include a control group with children who read the same words as those used in the treatment but without attention or instruction on syllabic reading to isolate potential effects of the focus on
syllables. In this context, applying the same training in a group of good readers could be fruitful, because comparing the results for the poor readers to this group of readers, who presumably make efficient use of syllables and larger linguistic structures for word recognition, might shed light on the role of syllables in producing the treatment effect. Likewise, the components of the intervention might be investigated in more detail to gain an insight into the processes that produce the treatment effect (e.g., effects of reading words independently vs. playing syllable-based games in groups, and effects of reading isolated words vs. reading these words embedded in sentences and short texts). Implementing a learning progress assessment, for example, could help to monitor the participants’ progress in reading skills during the intervention (Förster, Kawohl, & Souvignier, 2018). In addition, assessing children’s ability to spell multisyllabic words before and after the intervention might also be fruitful to detect possible transfer effects on children’s spelling ability.

5 Conclusion

The current study further underscores the potential of word-level reading interventions for helping poor readers. The criterial tests used to evaluate the training effects adequately assessed the efficiency of word-recognition processes by combining reaction time and accuracy. Based on the results from using these measures, syllable-based reading training can enable poor readers with inefficient word-recognition processes to recognize words more accurately and with less cognitive effort. The intervention seems to strengthen the mental representations of syllables and orthographic representations that consist of these syllables. A noteworthy finding from the study is that transfer effects on text-based reading comprehension were fully mediated by improvements in word-reading skills.
References


Heikkilä, R., Aro, M., Närhi, V., & Westerholm, J. (2013). Does training in syllable recognition improve reading speed? A computer-based trial with poor readers from


Tables

Means and Standard Deviations for the Raw Scores of all Variables at all Measurement Occasions by Treatment Condition compared to the Class Norms.

<table>
<thead>
<tr>
<th>Treatment condition</th>
<th>Grade 2 norms according to the manual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treatment group</td>
</tr>
<tr>
<td></td>
<td>M</td>
</tr>
<tr>
<td>Word recognition</td>
<td></td>
</tr>
<tr>
<td>1. Phonological recoding t1</td>
<td>261.39</td>
</tr>
<tr>
<td>2. Phonological recoding t2</td>
<td>349.77</td>
</tr>
<tr>
<td>3. Phonological recoding t3</td>
<td>354.37</td>
</tr>
<tr>
<td>4. Direct word recognition t1</td>
<td>269.79</td>
</tr>
<tr>
<td>5. Direct word recognition t2</td>
<td>336.29</td>
</tr>
<tr>
<td>6. Direct word recognition t3</td>
<td>374.52</td>
</tr>
<tr>
<td>7. Meaning access t1</td>
<td>281.20</td>
</tr>
<tr>
<td>8. Meaning access t2</td>
<td>310.31</td>
</tr>
<tr>
<td>9. Meaning access t3</td>
<td>356.66</td>
</tr>
<tr>
<td>Reading comprehension</td>
<td></td>
</tr>
<tr>
<td>10. ELFE t1</td>
<td>3.45</td>
</tr>
<tr>
<td>11. ELFE t2</td>
<td>5.44</td>
</tr>
<tr>
<td>12. ELFE t3</td>
<td>7.04</td>
</tr>
</tbody>
</table>

Note. Word recognition = subtest of ProDi-L, integrated values as quotient of accuracy and reaction time (Richter et al., 2017), Phonological recoding: min = 0, max = 494.99, Direct word recognition: min = 0, max = 506.09, Meaning access: min = 0, max = 507.75, for each measurement different but parallelized versions of the subtests were used, descriptives in the manual are based on the ProDi-L full version with all items per subtest; Reading comprehension = subtest of Elfe, sum of correct answers, min = 0, max = 20 (Lenhard & Schneider, 2006); t1 = pretest; t2 = posttest; t3 = follow-up; Sample sizes of the norming studies: ProDi-L N = 323 second graders, ELFE N = 421 second graders.
Table 2

Intercorrelations for all Variables at all Measurement Occasions by Treatment Condition (Treatment Group below the Diagonal, Wait-list Group above the Diagonal).

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Word recognition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. phonological recoding t1</td>
<td>—</td>
<td>.43*</td>
<td>.23</td>
<td>-.02</td>
<td>.00</td>
<td>.04</td>
<td>.07</td>
<td>.06</td>
<td>.27</td>
<td>.00</td>
<td>.03</td>
<td>.12</td>
<td>71</td>
</tr>
<tr>
<td>2. phonological recoding t2</td>
<td>.26*</td>
<td>—</td>
<td>.23</td>
<td>.06</td>
<td>.04</td>
<td>.00</td>
<td>-.03</td>
<td>.12</td>
<td>.23</td>
<td>.00</td>
<td>.17</td>
<td>.20</td>
<td>61</td>
</tr>
<tr>
<td>3. phonological recoding t3</td>
<td>.05</td>
<td>.15</td>
<td>—</td>
<td>.17</td>
<td>.26</td>
<td>.13</td>
<td>-.06</td>
<td>.29*</td>
<td>.40*</td>
<td>.10</td>
<td>.36**</td>
<td>.11</td>
<td>54</td>
</tr>
<tr>
<td>4. direct word recognition t1</td>
<td>-.03</td>
<td>.17</td>
<td>.32*</td>
<td>—</td>
<td>.12</td>
<td>.25</td>
<td>-.25*</td>
<td>.11</td>
<td>-.02</td>
<td>-.01</td>
<td>.28*</td>
<td>.11</td>
<td>71</td>
</tr>
<tr>
<td>5. direct word recognition t2</td>
<td>-.09</td>
<td>.04</td>
<td>.41**</td>
<td>.39**</td>
<td>—</td>
<td>.30</td>
<td>.07</td>
<td>.28</td>
<td>-.17</td>
<td>-.17</td>
<td>.43**</td>
<td>.28*</td>
<td>60</td>
</tr>
<tr>
<td>6. direct word recognition t3</td>
<td>-.11</td>
<td>.01</td>
<td>.50**</td>
<td>.24</td>
<td>.32</td>
<td>—</td>
<td>.01</td>
<td>.18</td>
<td>-.12</td>
<td>-.13</td>
<td>.38*</td>
<td>.22</td>
<td>41</td>
</tr>
<tr>
<td>7. meaning access t1</td>
<td>-.18</td>
<td>.03</td>
<td>-.02</td>
<td>-.27*</td>
<td>-.06</td>
<td>.14</td>
<td>—</td>
<td>.04</td>
<td>.09</td>
<td>-.09</td>
<td>-.11</td>
<td>.01</td>
<td>71</td>
</tr>
<tr>
<td>8. meaning access t2</td>
<td>-.01</td>
<td>.10</td>
<td>.12</td>
<td>.17</td>
<td>.07</td>
<td>-.03</td>
<td>.16</td>
<td>—</td>
<td>.32*</td>
<td>.14</td>
<td>.30*</td>
<td>.23</td>
<td>61</td>
</tr>
<tr>
<td>9. meaning access t3</td>
<td>.10</td>
<td>.17</td>
<td>.25</td>
<td>-.02</td>
<td>.17</td>
<td>.30</td>
<td>.29</td>
<td>.45*</td>
<td>—</td>
<td>.25</td>
<td>-.09</td>
<td>.18</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Reading comprehension</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. ELFE t1</td>
<td>-.06</td>
<td>.04</td>
<td>-.08</td>
<td>.07</td>
<td>.04</td>
<td>-.06</td>
<td>.07</td>
<td>-.01</td>
<td>.25</td>
<td>—</td>
<td>.12</td>
<td>.06</td>
<td>69</td>
</tr>
<tr>
<td>11. ELFE t2</td>
<td>-.20</td>
<td>.25*</td>
<td>.25</td>
<td>.29*</td>
<td>.09</td>
<td>.17</td>
<td>-.03</td>
<td>.20</td>
<td>-.11</td>
<td>.18</td>
<td>—</td>
<td>.56**</td>
<td>60</td>
</tr>
<tr>
<td>12. ELFE t3</td>
<td>-.14</td>
<td>.33*</td>
<td>.39**</td>
<td>.31*</td>
<td>.48**</td>
<td>.54**</td>
<td>-.02</td>
<td>.08</td>
<td>.15</td>
<td>.27*</td>
<td>.35*</td>
<td>—</td>
<td>52</td>
</tr>
</tbody>
</table>

n

Note. Word recognition = subtest of ProDi-L, integrated values as quotient of accuracy and reaction time (Richter et al., 2017); Reading comprehension = subtest of Elfé, sum of correct answers (Lenhard & Schneider, 2006); t1 = pretest; t2 = posttest; t3 = follow-up.

* p < .05. ** p < .01.
Table 3

**Parameter Estimates for Linear Models with the Efficiency of Word Recognition Processes and Reading Comprehension as Outcome Variables at the Posttest and Treatment Condition, Pretest Scores, and Intellectual Ability as Predictors**

<table>
<thead>
<tr>
<th></th>
<th>Word recognition</th>
<th>Reading comprehension</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Phonological recoding</td>
<td>Direct word recognition</td>
</tr>
<tr>
<td>Intercept</td>
<td>-.39** .11</td>
<td>-.15 .12</td>
</tr>
<tr>
<td>Treatment vs. wait-list group (dummy coded 1 vs. 0)</td>
<td>.77** .16 .08</td>
<td>.29* .17 .01</td>
</tr>
<tr>
<td>Pretest scores</td>
<td>.48** .08 .23</td>
<td>.34** .09 .13</td>
</tr>
<tr>
<td>Intellectual ability</td>
<td>.14 .08 .02</td>
<td>.16* .09 .02</td>
</tr>
<tr>
<td>Goodness of fit</td>
<td>$R^2 = .33, F(3,120) =$</td>
<td>$R^2 = .16, F(3,126) =$</td>
</tr>
<tr>
<td></td>
<td>19.48, $p &lt; .001$</td>
<td>7.91, $p &lt; .001$</td>
</tr>
<tr>
<td>Effect size (Cohen’s $d$)</td>
<td>0.83</td>
<td>0.29</td>
</tr>
<tr>
<td>Post-hoc power ($1 - \beta$)</td>
<td>.99</td>
<td>.99</td>
</tr>
</tbody>
</table>

*Note.* All outcomes and metric predictors were $z$-standardized.

* $p < .05$. ** $p < .001$ (one-tailed).
Figure 1. Sampling and flow of the subjects through the pretest-posttest design with follow-up.
Figure 2. Mediation model for treatment condition (dummy coded, wait-list group as reference) on reading comprehension with word recognition ($z$-standardized, composite score of phonological recoding, direct word recognition, and access to word meanings) as mediator. Unstandardized regression weights with associated standard errors in parentheses (* $p < .05$).

$R^2 = .11, F (5,116) = 2.92, p < .05$
Figure 3. Trajectories of visual word recognition skills (z-standardized, composite score of phonological recoding, direct word recognition, and access to word meanings) by treatment condition.
### Appendix

Table A1

*Reliabilities (Cronbach’s α) for all Measures at all Measurement Occasions in Comparison to Reliabilities in Grade 2 in the Norming Studies*

<table>
<thead>
<tr>
<th></th>
<th>Pretest</th>
<th>Posttest</th>
<th>Follow-up</th>
<th>Manual</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Word recognition (ProDi-L)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phonological recoding</td>
<td>.82</td>
<td>.87</td>
<td>.69</td>
<td>.97</td>
</tr>
<tr>
<td>Direct word recognition</td>
<td>.81</td>
<td>.80</td>
<td>.80</td>
<td>.99</td>
</tr>
<tr>
<td>Meaning access</td>
<td>.79</td>
<td>.78</td>
<td>.78</td>
<td>.95</td>
</tr>
<tr>
<td><strong>Reading comprehension</strong></td>
<td>.65</td>
<td>.78</td>
<td>.83</td>
<td>.94</td>
</tr>
</tbody>
</table>

(ELFE 1-6)

*Note.* Word recognition = subtest of ProDi-L, reliabilities based on reaction times, for each measurement different but parallelized versions of the subtests were used. Reliabilities in the manual are based on the ProDi-L full tests with all items per subtest (Richter et al., 2017); Reading comprehension = subtest of ELFE, reliabilities based on accuracy (Lenhard & Schneider, 2006). Sample sizes of the norming studies: ProDi-L $N = 323$ second graders, ELFE $N = 421$ second graders. Note that the number of items used in the word recognition subtests taken from ProDi-L was only one-third of the number of items in the full tests, for which reliabilities are reported in the ProDi-L manual.