Modelling the Relationship of Accurate and Fluent Word Recognition in Primary School

Panagiotis Karageorgos, Bettina Müller & Tobias Richter

University of Würzburg

Manuscript accepted for publication in Learning and Individual Differences (2019)

Author Note

The research reported in this article was supported by the German Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF, grant 01GJ1402B).

Address for correspondence:

Panagiotis Karageorgos

University of Würzburg

Department of Psychology IV

Röntgenring 10

97070 Würzburg, Germany

E-mail: panagiotis.karageorgos@uni-wuerzburg.de

Abstract

This study investigated whether word-reading speed starts increasing only after German fourth graders (*n* = 826) have reached a basic level of word-reading accuracy. Moreover, we examined for 170 readers with lower reading abilities (below percentile rank 50 in both word reading and reading comprehension) in an experimental pre-post control-group design whether a word-reading intervention has differential effects depending on the level of accuracy a child has reached before the intervention. The results based on the full sample suggest that a specific level of word-reading accuracy seems to be required before word-reading speed starts improving. Further analyses with the trained readers showed positive treatment effects on word-reading accuracy for readers below the accuracy level, on word-reading speed regardless of their accuracy and on reading comprehension for readers above the accuracy level. The results suggest that a sufficient level of word-reading accuracy is an important precondition for the development of fluent reading as well as the effectiveness of reading interventions at the word level in German fourth graders.

Keywords: primary school children, reading development in German, word-reading accuracy and word-reading speed, reading comprehension, syllable-based intervention

To be able to comprehend written texts, readers need to master various cognitive processes, which range from recognizing written words, comprehending sentences, to establishing coherence between multiple sentences (Daneman, 1991; Perfetti, Landi, & Oakhill, 2005; Vellutino, Fletcher, Snowling, & Scanlon, 2004). Word level processes play a pivotal role for reading comprehension (Perfetti & Hart, 2002). If these processes function inefficiently, fewer cognitive resources will be available for higher levels of processing at the sentence or text level and reading comprehension can be impaired (Perfetti, 1985). Thus, efficient word recognition, which is characterized by accurate and fluent word reading, is an important precondition for good reading comprehension.

This study examined the development and training of word reading in German fourth graders. According to a study of Danish children by Juul, Poulsen, and Elbro (2014), children must first reach a basic level of word recognition accuracy before their fluency can increase. Our first aim was to determine whether such a basic accuracy threshold also exists in German fourth graders. A second aim was to investigate whether a reading intervention at the word level has different effects, depending on the accuracy level of the participants before the intervention. If a basic accuracy threshold exists, a word reading training may be expected to improve reading fluency only in children who have already reached this threshold. In the following, we briefly explain the development of word reading skills and provide a focused overview of interventions that have been developed to increase reading accuracy and fluency. The theories and findings associated with these interventions form the background of the hypotheses tested in our study.

The development of accurate and fluent word reading

Phonological recoding

The development of word reading mainly rests on the development of two types of processes: phonological recoding and orthographical decoding (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Martens & de Jong 2006; Zoccolotti et al. 2005). In phonological recoding, readers rely on grapheme-phoneme correspondence rules to convert each letter into a phoneme and then assemble a word (Coltheart et al., 2001; Müller & Richter, 2014). This process needs more time than orthographic decoding, strains phonological working memory, and can only function correctly for regular words that follow the spelling pattern of the language (Müller & Richter, 2014; Snowling & Hulme, 2005). Hence, skilled readers use phonological recoding mostly for unknown or infrequent words (Share, 1999, 2004).

Orthographical decoding

In orthographical decoding, the letter clusters activate the word’s stored representation in the orthographic lexicon, and then the stored representation activates the word’s node in the phonological lexicon, which in turn activates the word’s phonemes (Coltheart, 2005). Orthographic decoding can only be used for familiar words, i.e. words that already exist in the mental lexicon and are easily accessible (Coltheart et al., 2001). This more efficient processing allows readers to bypass the conversion of each letter into phonemes. Consequently, readers have more cognitive resources at their disposal for higher-order comprehension processes at the sentence and text level (Perfetti & Hart, 2002; Pikulsky & Chard, 2005), which in turn fosters text comprehension (e.g., Fuchs, Fuchs, Hosp, & Jenkins, 2001; Kim, Petcher, & Foorman, 2015).

The importance of orthographical decoding in the German language

In reading development, phonological recoding serves as a bridge to accurate reading and allows for some degree of fluency. However, children need to acquire and routinize orthographical decoding skills to read fluently and eventually comprehend texts without errors. Nonetheless, phonological recoding remains "an alternative mechanism for automatic translation of orthographic information into a sublexical phonological code" even in skilled adult readers (Grainger, Lété, Bertand, Dufau, & Ziegler, 2012, p. 289). In the process of routinization, readers also gradually learn to make use of morphemes and syllables to segment a word (Coltheart et al., 2001; Klicpera, Gasteiger-Klicpera, & Schabmann, 1993). The orthography of the language makes a difference for the slopes of the developmental trajectories. In a language with a transparent orthography such as German, one year of formal instruction is usually sufficient to recognize words accurately via phonological recoding, whereas in opaque orthographies like Danish that lack one-to-one letter-sound correspondence this development can be twice as slow (Seymour, Aro, & Erskine, 2003). Landerl and Wimmer (2008) found in an 8-year longitudinal study with German students that poor readers were able to decode 72% of the words and 61% of the pseudowords correctly at the end of Grade 1 (see also Gangl et al., 2018). Similarly, a study by Richter, Isberner, Naumann and Kutzner (2012) found that the accuracy of word recognition and pseudoword recognition increased monotonically from the end of Grade 2 to the end of Grade 4. In the same period, the average speed of recognizing words or pseudowords increased in a log-linear fashion but large individual differences in the word-reading speed remained even at the end of Grade 4 (cf. Landerl & Wimmer, 2008, number of syllables read per minute, end of Grade 1: 19 to 176, end of Grade 4: 59 to 285). This raises the question how the accuracy and fluency of word reading are related. Does fluency develop even in readers who are not yet accurate readers?

Theory and research suggests that this is not the case. Apparently, readers need to be able to recognize words with a sufficient degree of accuracy before their reading speed can develop, too. If readers struggle to recognize words, the limited cognitive capacity will be overtaxed and, thus, word-reading fluency will also be impaired (Castles, Rastle, & Nation, 2018; LaBerge & Samuels, 1974). Furthermore, accurate phonological recoding processes are needed for building up high-quality representations of orthographical word forms in the mental lexicon, which, in turn, is a prerequisite for fluent reading. These general assumptions are in line with findings obtained by Juul et al. (2014) from a sample of Danish beginning readers. These authors argue that readers should attain a basic accuracy of word recognition before reading fluency can increase. Scatterplots of beginning readers' accuracy and speed scores showed a curvilinear distribution in which word recognition speed starts increasing after participants have reached a 70% accuracy score. To investigate this relationship statistically, the authors calculated correlations between word recognition accuracy and speed for participants below and above the accuracy threshold. The correlations for the participants below the threshold were not significant, indicating that the speed of word recognition starts increasing only after readers have reached a certain level of word recognition accuracy. Moreover, according to the same authors, the longer readers can read with this accuracy (basic accuracy achievement time), the more time they have at their disposal to develop and improve their word recognition speed.

In sum of the reviewed literature, the development of accurate and fluent word reading skills is a complex interplay of phonological and orthographical processes (Coltheart et al., 2001). For children learning to read in German, accurate and fluent orthographical decoding skills seems to be important for efficient word recognition. Differences between good and poor readers in these skills occur early and remain stable (Richter et al., 2012). Moreover, it seems necessary to develop a basic accuracy level first before fluency can increase (Juul et al., 2014).

Trainings to foster poor word recognition

Evidence-based word reading interventions for poor readers in primary schools are usually divided into phonics trainings and reading fluency trainings (Galuschka, Ise, Krick, & Schulte-Körne, 2014). *Phonics trainings* teach children to read words by making use of the grapheme-phoneme correspondence rules, whereas *reading fluency trainings* focus on the ability to read accurately and fast by making use of the repeated reading method (Suggate, 2016). Several systematic reviews and meta-analyses suggest positive effects of phonics trainings on word recognition skills (e.g., Ehri, Nunes, Stahl, & Willows, 2001; Galuschka et al., 2014; McArthur et al. 2018; Suggate, 2010) and of repeated reading on reading fluency and comprehension skills (e.g., Meyer & Felton, 1999; Stevens, Walker, & Vaughn, 2017; Suggate, 2016; Therrien, 2004).

In the German language, the treatment effect of phonics trainings on word-reading accuracy is very clear and robust (e.g. Galuschka & Schulte-Körne, 2015). In contrast, the effect of phonics training on word-reading fluency is less clear and rather fragile. For example, Klatte et al. (2014) investigated the effects of a computer-based phonics training, which mainly focused on fostering the grapheme-phoneme correspondence rules and phonemic awareness. They found a treatment effect on word-reading accuracy two months after the end of the training, but no treatment effect on word-reading fluency. A positive treatment effect of a phonics training on word-reading accuracy has also been reported for a sample of German first graders (Klatte, Steinbrink, Bergström, & Lachmann, 2016).

However, it seems plausible to assume that word-reading fluency might profit from phonics trainings that go beyond teaching grapheme-phoneme-correspondences and make use of syllabication, the process of segmenting a word into syllables. For example, the disyllabic German word *Ratte* (rat) would be segmented into *Rat.te* because of the double consonants /tt/ (Dudenredaktion, 2017). The trisyllabic word *Ameise* (ant) would be *A.mei.se* because the diphthong /ei/ is never divided (Dudenredaktion, 2017). Ritter and Scheerer-Neumann (2009) developed a phonics training with the primary focus on syllabication. In a sample of German third and fourth graders who were identified as poor readers, they found positive treatment effects on both word-reading accuracy and word-reading fluency. Furthermore, intervention studies conducted in other languages also hint on the beneficial effect of repeated reading of syllables to improve the accuracy and fluency of word recognition for poor reading children in Grades 2 to 4 (e.g., Finish: Heikkilä, Aro, Närhi, Westerholm, & Ahonen, 2013; Huemer, Aro, Landerl, & Lyttinen, 2010; Dutch: Berends, & Reitsma, 2006; Wentink, Van Bon, & Schreuder, 1997; French: Ecalle, Magnan, & Calmus, 2009). McArthur et al. (2015) studied the effects of a reading intervention with poor English readers aged from 7 to 12 years. The intervention consisted of two components, a phonics training that made use of syllabication and a sight word training (repeated reading of words). The results showed gains in word-reading accuracy and fluency in both conditions. Hence, it appears that training of syllabication is a key component for fostering word-reading fluency. Moreover, training phonics before sight words led to significantly larger gains in accuracy than training sight words before phonics (McArthur et al., 2015). This result is consistent with the results of Juul et al. (2014), suggesting that the phonics training enabled children to reach the basic accuracy level that is needed before reading fluency could be increased through repeated reading.

The present study

The present study followed two aims. First, against the background of the extensive and stable range in word-reading fluency of German readers from Grade 1 through 8 (cf. Landerl & Wimmer, 2008), we hypothesized that an accuracy-before-fluency pattern as described in Juul et al. (2014) would also be found in German-speaking fourth graders. Specifically, we were interested in whether these children would need to attain a basic accuracy of word recognition before their reading fluency could increase. We used orthographical decoding skills as a measure of accurate and fluent word recognition skills because of their importance for reading fluency and their strong relationship with text comprehension throughout primary school in German readers (Knoepke, Richter, Isberner, Naumann, & Neeb, 2014; Richter et al., 2012). Several studies suggest that poor reading skills in transparent orthographies, such as German, are associated with deficits in readers' orthographic decoding route (e.g., Martens & de Jong, 2006; Protopapas, Sideridis, Mouzaki, & Simos, 2007; Richter et al., 2012; Zoccolotti et al. 2005). In contrast, the phonological recoding is slow but reliable even in poor readers (e.g., Mayringer & Wimmer, 2000; Ziegler, Conrad, Ma-Wyatt, Ladner, & Schulte-Körne, 2003). German readers have already received a large dose of spelling instruction at the beginning of

Grade 4, and they should be able to read books appropriate for their age. Thus, in comparison to the sample of beginning readers studied by Juul et al., deficits in orthographical decoding skills should be more obvious at this time. We expected a curvilinear relationship between accuracy and speed of orthographical decoding, with a positive relationship of accuracy and speed only in children who excel a basic accuracy threshold (Hypothesis 1).

The second aim of this study was to implement the basic accuracy level in a longitudinal study that investigates the effectiveness of a word recognition training that focuses on accurate and fast reading of frequent German syllables. The study was guided by the question of whether the training effects on word reading and text-based reading comprehension depend on the children’s basic accuracy level before the intervention. We expected children in the treatment group who scored below the basic accuracy level before the intervention to read words more accurately at posttest compared to children in the control group below the basic accuracy level (Hypothesis 2).

According to Juul et al. (2014), the basic accuracy achievement time explains a significant amount of variance in the fluency of word recognition. Danish readers who scored above the basic accuracy level had more time to read fluently than the readers below the basic accuracy level. The word recognition training implemented in this study provided ample opportunities for becoming more fluent in word reading (see the description of the training in the Method section below), but according to the notion of a basic accuracy level, only the reading fluency of children who scored above this level should benefit from these opportunities. Two related hypotheses were derived from this assumption. First, we expected children in the treatment group who scored above the basic accuracy level at pretest to read words more fluently after the intervention at posttest compared to untrained children in the control group who scored above the basic accuracy level at pretest (Hypothesis 3a). Furthermore, within the treatment group, children who scored above the basic accuracy level were expected to read words more fluently than children below the basic accuracy level (Hypothesis 3b).

Apart from the treatment effects on the accuracy and fluency of word recognition, the lexical quality hypothesis (Perfetti & Hart, 2002) predicts that a word recognition training that improves fluent word reading should also indirectly promote reading comprehension by making more cognitive resources available for higher-order comprehension processes. Thus, we expected an indirect treatment effect on reading comprehension through word-reading fluency for the children in the treatment group who were already above the basic accuracy level at pretest compared to same-skilled children in the control condition who were above the basic accuracy level at pretest, too (Hypothesis 4).

Method

Design and Procedure

The data were obtained in two longitudinal studies investigating the effects of different reading trainings in primary school. Study 1 took place in the school year 2014-2015 and Study 2 in the school year 2015-2016. Both studies were based on an experimental pre/post-test design with randomization at the class level and were conducted in the urban areas of Giessen and Kassel (Germany). Participants in both studies were first screened with subtests of two standardized German reading tests: ProDi-L (Richter, Naumann, Isberner, Neeb, & Knoepke, 2017) for word recognition processes and ELFE 1-6 (Lenhard & Schneider, 2006) for reading comprehension in Study 1 and the revised ELFE II (Lenhard, Lenhard, & Schneider, 2017) in Study 2. Children with pretest scores below percentile rank 50 on the class norms of both word recognition (mean composite score of the ProDi-L subtests phonological recoding, orthographical decoding, and access to word meanings) and reading comprehension were selected as participants for the intervention. If there were more than seven participants in one class, they were randomly assigned in groups of four to six. The groups were randomly allocated at the class level to either the treatment or the control condition to avoid drawing false conclusions due to regression toward the mean (Trochim, Donnelly, & Arora, 2016).

The groups in the treatment condition received the intervention after the pretest, and the groups in the control condition (wait-list) received the same intervention after the posttest. The intervention occurred twice a week for 45 min per training session in addition to the regular school curriculum and lasted up to 3 months. The training sessions were conducted by student research assistants (psychology undergraduates or prospective teachers) who received standardized instructions and were supervised in regular intervals by the authors. Each training group of four to six children was assigned to a trainer and each session was conducted in a group setting. The children’s reading processes were assessed again after the final training session with ProDi-L and ELFE 1-6 or ELFE II respectively.

For the purpose of this study, data from both intervention studies were combined. Note that parts of the data from Study 1 were previously reported with a different analytic focus (Müller, Richter, Karageorgos, Krawietz, & Ennemoser, 2017). In the previous study, we examined the effects of the intervention by comparing the integrated posttest scores of orthographical decoding (computed as the ratio of accuracy and response time) and reading comprehension posttest scores of the 43 children in the treatment group with the 32 children in the wait-list group. The results showed a significant treatment effect on word recognition but not on reading comprehension. Furthermore, we investigated if the treatment and wait-list groups differ significantly from 44 untreated good readers in word recognition and reading comprehension at posttest. The word recognition performance of the treatment group was still significantly worse than that of the untrained good readers, but no significant difference occurred for reading comprehension.

In the current study, we included word-reading accuracy and word-reading speed as separate scores and used a larger aggregated sample of two intervention studies to examine whether children below and above a basic accuracy level respond differently to the intervention.

Participants

The participating 826 fourth graders in the aggregated data set attended 21 schools. All parents gave informed written consent to participation in the research. Treatment group allocation (random assignment) comprised 105 children in the treatment group and 65 children in the wait-list group (see Table 1 for the composition of the sample per study). The 170 children with reading skills below percentile rank 50 were on average 9.41 years old (SD = 0.75) and the proportion of girls and boys was nearly equal in the aggregated sample, χ^2 (1) = 1.30, p = .254 (Table 1). Parents of 22 children from the treatment group and 13 from the wait-list group reported that their child's first language was another language other than German. Parents of 61 children from the treatment group and 33 children from the wait-list group omitted information about first language.

One-way analyses of variance (ANOVA) found no statistically significant pretest differences between the aggregated treatment and wait-list groups in age, $F(1, 79) = 0.01$, p = .921, accuracy of orthographic decoding, $F(1, 167) = 0.58$, $p = .447$, speed of orthographic decoding, $F(1, 160) = 0.26$, $p = .612$, and reading comprehension, $F(1, 167) = 3.53$, $p = .062$.

To examine differences in the development of the treatment groups of Study 1 and 2 from pretest to posttest we conducted ANOVAs with the difference scores of accuracy and speed of orthographical decoding and reading comprehension at posttest minus their corresponding pretest scores as outcome variables (cf. Judd, Kenny, & McClelland, 2001). The results indicate no significant differences in the development of orthographical decoding accuracy, $F(1, 93) = 1.18$, $p = .280$, orthographical decoding speed, $F(1, 85) = 0.04$, $p = .839$, and reading comprehension, $F(1, 93) = 1.85$, $p = .547$. In sum, average test scores were comparable in the samples of both studies.

Measures

Orthographical decoding. A lexical decision task with 16 items presented in randomized order was used to assess the accuracy and speed of orthographical decoding (subtest of the German-speaking computerized instrument ProDi-L; Richter et al., 2017). The children's task was to decide whether a string of letters was a real word or a pseudoword using two response keys (dichotomous response format: yes/no). Half of the items were real German words and the other half orthographically and phonologically legal pseudowords. Parallel versions of the 16 items were used at pre- and posttest. All items varied systematically in length, frequency, and number of orthographical neighbors. The pseudowords varied in their similarity to actual German words. For each item, response accuracy and response latency were measured.

Lexical decision tasks are commonly used to assess the accuracy and speed of word recognition. The speed of responses in a lexical decision task has been shown to correlate strongly with word naming speed (a measure of word reading fluency that involves reading aloud) and reading fluency on the text level, as well as with reading comprehension in a sample of adult readers with reading problems (Katz et al., 2012). Likewise, strong correlations with reading comprehension for both accuracy and speed of lexical decision with the same tasks as in the present study have also been shown in samples of elementary school children (Richter et al., 2012, 2013). The accuracy of lexical decision responses for common words tends to be quite high by the end of primary school but the remaining variability is nevertheless strongly related to reading comprehension, over and above the contributions of the speed of lexical decisions (e,g,, for evidence in German primary school children: Richter et al., 2012). Yap, Balota, Sibley, and Ratcliff (2012) used data from the English Lexicon Project to show that even in adult readers, where the accuracy of lexical decisions is higher than 90%, the remaining individual differences still reflect systematic variance (not just response error) and are strongly related, for example, to vocabulary knowledge (r = .622).

Based on Juul et al. (2014), we transformed the sum of correct responses into a percentage score representing the accuracy of orthographical decoding and calculated a words-per-minute score as an indicator of orthographical decoding speed. The words-per-minute score was calculated by multiplying the number of items with valid responses by 60000 ms and the product divided by the overall latency measured in ms. For example, a child who responded to 16 items in 16000 ms received a score of 60 words per minute. No score was computed for participants with more than 10% missing values. The test-retest reliability over a 5-month interval was computed as the intraclass correlation of pre- and posttest scores of the complete data of 538 untrained children by using the R-package irr (Gamer, Lemon, Fellows, & Singh, 2019). A two-way mixed-effects model based on a mean-rating and absolute agreement was used (Price et al., 2015; Koo & Li, 2016; McGraw & Wong, 1996). According to the rules of thumb provided by Cicchetti (1994), the estimated test-rest reliability was acceptable for the accuracy percentage score (ρ_I = .62, 95% CI [.54, .68]) and good for the words-per-minute score (ρ_I = .77, 95% CI [.66, .84]).

Reading comprehension. Reading comprehension skills were assessed with the computerized subtest text comprehension of the ELFE 1-6 (Lenhard & Schneider, 2006) in Study 1 and the revised ELFE II (Lenhard et al., 2017) in Study 2. The test consists of short, narrative and expository texts (two to five sentences) with four multiple-choice items presented in randomized order at both measurement points. The items assess the children’s ability to identify information in texts, generate anaphoric references and to make inferences across sentences in the text. In ELFE 1-6, 20 texts were presented; in ELFE II, 26 texts. The sum of correct responses within 7 min processing time was transformed into a percentage score of accuracy. Again, the test-retest reliability over a 5-month interval was computed as the intraclass correlation of pre- and posttest comprehension scores of the complete data of 538 untrained children by using the R-package irr (Gamer et al., 2019). A two-way mixed-effects model based on a mean-rating and absolute agreement was used (Price et al., 2015; Koo & Li, 2016; McGraw & Wong, 1996). The estimated test-rest reliability was good, $\rho_I = .83$, 95% CI [.61, .91].

Word Recognition Training

We used a new syllable-based word-reading training (Müller, Otterbein-Gutsche, & Richter, 2018) that combines the phonics and reading fluency approaches. The training was based on the 500 most frequent German syllables in texts typically read by 9- to 12 year-old children (selected from the data base childLex, Schroeder, Würzner, Heister, Geyken, & Kliegl, 2015), comprised of 24 sessions and was divided in two phases. In the first phase (Sessions 1 – 15), the training focused on grapheme-phoneme correspondence rules within syllables. The exercises consisted of analyzing the syllabic structure of words by marking syllables with arcs during reading, finding the vowel nucleus within each syllable, and combining prefixes and stems. In the second phase (sessions 16-24), the training focused mainly on providing extensive practice of accurate and fast reading. Several syllable and word reading games were used, for example the “syllable race” (children drew successively a card from the top of a deck, read the word on the card, and moved their token on the board one square for each syllable while reading the word loudly) or a kind of flashcard reading (very briefly presentation of words on cards). The complexity of the material in all exercises was increased successively beginning with single words with regular spelling and a maximum length of four syllables up to irregularly spelled words with a maximum length of eight syllables as well as sentences and short texts.

The rationale behind capitalizing on syllabic reading was to strengthen the mental representations of frequent syllables and orthographic representations of words consisting of these syllables, because poor readers often have trouble identifying the syllabic structure of words (Colé, Magnan, & Grainger, 1999; Scheerer-Neumann, 1981). Moreover, the accurate phonological pronunciation of consonant clusters was important to help children master the complex syllabic structure of German (167 possible syllable types with different combinations of vowels and consonants, and different vowel lengths; Seymour et al., 2003; Colé et al., 1999; Scheerer-Neumann, 1981). Thus, the aim of the intervention was to improve children’s accuracy and speed of word recognition by moving them from slow letter-by-letter decoding to faster holistic reading via the extraction of syllabic units (consolidated alphabetic phase, Ehri, 2005).

The training was designed for 24 sessions. However, in Study 1 the exercises were implemented faster than expected after three weeks of intervention Thus, for Study 1 we reduced the number of sessions to 16 by combining two consecutive sessions within one. In Study 2, a revised version of the intervention was used with the same exercises and structure as in Study 1 but with extended word material, which was still based on the 500 most frequent syllables.

Results

Given the hierarchical structure of the data (students nested within schools), the intra-class correlations (ICC) with random intercept multilevel models (Raudenbush & Bryk, 2002) were estimated with the R-packages lme4 (Bates, Maechler, Bolker, & Walker, 2015) and lmerTest (Kuznetsova, Brockhoff, & Christensen, 2019). The low ICC ($\rho < .05$) for the words-per-minute score at pretest indicated no clustering effect in the data. Thus, we proceeded with regular multiple regression models with listwise deletion.

The significance level for all significance tests was set at .05 (one-tailed, as all hypotheses are directional). To examine the assumptions of linearity, normality, and homoscedasticity of linear models, the standardized residuals were plotted against the unstandardized predicted values. None of the assumptions underlying linear models were violated in any of the models. The assumptions of non-multicollinearity of the predictors and the independence of residuals were also supported (Cohen, Cohen, West, & Aiken, 2003, Chapters

4 and 10). As ProDi-L is a computerized, reaction-time based test, data might be biased due to participants clicking randomly. To detect these individuals, two outlier identification rules were used in all models. First, all cases with high discrepancy values (cutoff: ± 2.00 for the externally studentized residuals) and high global influence (cutoff: + 1.00 for Cook's *D*) were excluded from the analysis (Cohen et al., 2003, Chapter 10). With this approach, two outliers were identified and excluded. Second, the data were examined for cases with accuracy values below 50% at each measuring time, which is the probability of guessing correctly an item with dichotomous format (Urbina, 2014, Chapter 6), and words-per-minute values below 1 standard deviation below the mean at pretest, which would be an indication for anomalous responding. None of the participants matched the second outlier identification rule.

Accuracy-before-Speed Pattern

To investigate whether German readers must first reach a basic accuracy level in orthographical decoding before their orthographical decoding speed starts to increase (Hypothesis 1), the words-per-minute score at pretest was plotted against the accuracy percentage score at pretest within the whole screening sample (N = 824 participants, 2 outliers; Figure 1). The plot shows a u-shaped curve.

We conducted a curve-fitting analysis with power polynomial regressions up to the third degree (Cohen et al., 2003, Chapter 6) to test whether a linear, quadratic, or cubic model fits best to describe the relationship between accuracy and orthographical decoding speed. The words-per-minute score at pretest was included as the outcome variable, and the accuracy percentage score at pretest was used as predictor. For model comparison, we used a likelihood ratio test taking R^2 (the proportion of explained variance by the model) and Akaike's information criterion (AIC; Akaike, 1973) into account (see Table 2 for the parameter estimates). Note that the AIC compares model fit by taking into account the number of predictors. Smaller AIC values indicate a better fit.

The three measures showed that the third-degree polynomial regression, an s-shaped cubic function with two bends, had the best fit.[1] This indicates that the relationship between accuracy and speed is not linear. Instead, a vertex appears in the relationship, indicating that orthographical decoding speed increased only after accuracy achieved a certain level. The second bend indicates an accuracy level for which orthographical decoding speed theoretically reaches a maximum. However, this accuracy level would be at 110%, a value that does not exist. Thus, in line with the accuracy-before-speed pattern, the accuracy-speed relationship is characterized by only one vertex. To investigate this first vertex at which the curve starts rising, we computed the minimum value (*W*) of the third-degree polynomial regression as recommended by Cohen et al. (2003, Chapter 6.2):

$$W = \frac{-B_2 \pm \sqrt{B_2^2 - 3B_1B_3}}{3B_3} \qquad (1)$$

In Equation 1, *B* refers to the unstandardized coefficients of the model, and the subscripted numbers indicate the power degree of the polynomial predictors, $B_1 = 0.430$, $B_2 = 0.006$ and $B_3 = -0.0004$. The result showed that 71% accuracy is the minimum value (*W*) before the curve starts rising (Figure 1). Hence, progression in orthographical decoding speed could mainly be observed after children achieved a 71% accuracy level. The accuracy-before-speed pattern is also substantiated by the correlations between accuracy and the words-per-minute score for readers below and above the basic accuracy level. No significant correlation was found for readers below the basic accuracy level ($r = -.19$, $p = .111$), whereas the correlation was significant and positive for readers scoring above this level ($r = .28$, $p < .001$). Furthermore, to confirm this relationship we also used the R-package segmented (Muggeo, 2017) to estimate the

[1] The exclusion of one influential data point with high fluency and low decoding accuracy led to a better fit for the quadratic model. However, the bend of the quadratic model was nearly identical to that of the cubic model and all further analyses led to the same results.

threshold via the piecewise regression approach with continuity requirement. The breakpoint was calculated by using an automatic iterative procedure to fit segmented linear regressions models without specifying an initial guess for the threshold. After two iterations the piecewise regression identified a breakpoint at 73% (SE = 2.95%) accuracy, which is pretty close to the threshold that we found in the curvilinear analysis. In sum, as expected, a basic accuracy level appears to exist in German that readers have to master before their orthographical decoding speed increases.

Treatment Effects on Accuracy of Word Recognition

To investigate whether the effect of the word-reading intervention depended on the children's basic accuracy level before the intervention, the 168 participants (excluding the 2 outliers) of the treatment and wait-list group were divided into four groups according to the 71% cutoff criterion of the curvilinear analysis: (1) treatment group above the basic accuracy level, (2) treatment group below the basic accuracy level, (3) control group above the basic accuracy level, and (4) control group below the basic accuracy level. Note that both criteria of the curvilinear and piecewise regression analyses led to the same segmentation into groups. Descriptive statistics for the four groups can be found in Table 3.

A multiple regression model was used to analyze whether children in the treatment group below the basic accuracy level read more accurately than children in the control group below the basic accuracy level at posttest (Hypothesis 2). The percentage score of the accuracy of orthographic decoding at posttest was used as the outcome variable. Three dummy-coded variables with the treatment group below the basic accuracy level as reference category were entered as predictors in the multiple regression model. The accuracy percentage score at pretest was entered as covariate (centered) to control for pre-training differences. All predictors were entered simultaneously into the model.

In line with Hypothesis 2, the results showed a significant treatment effect for the children below the basic accuracy level, B = -11.25, SE = 4.13, p = .007, ΔR^2 = .04, indicating that their orthographical decoding accuracy increased compared to the accuracy of the untrained children below the basic accuracy level (see Table 4 for the parameter estimates).

Treatment Effects on Speed of Word Recognition

To investigate whether children in the treatment group above the basic accuracy level achieved higher words-per-minute scores at posttest than untrained participants above the basic accuracy level (Hypothesis 3a) and trained participants below the basic accuracy level (Hypothesis 3b), we ran a multiple regression model. The words-per-minute score at posttest was used as outcome. Three dummy-coded variables with the treatment group above the basic accuracy level as reference category were entered as predictors. The words-per-minute score at pretest was entered as covariate (centered) to control for pre-training differences. All predictors were entered simultaneously into the model.

The results showed a significant treatment effect for the children above the basic accuracy level who received the treatment, B = -4.21, SE = 2.02, p = .039, ΔR^2 = .02, indicating that the improvement of the orthographical decoding speed was steeper for the children above the basic accuracy level who received the treatment than for the untrained children above the basic accuracy level (see Table 5 for the parameter estimates). Hence, in line with Hypothesis 3a, orthographical decoding speed improved for trained children above the basic accuracy level compared to untrained children above the basic accuracy level. However, no significant difference was found in the speed of word recognition at posttest between the treatment groups above and below the basic accuracy level, B = - 0.86, SE = 2.63, p = .744.

Treatment Effects on Reading Comprehension

We ran a third regression model to examine whether trained readers above the basic accuracy level scored higher on reading comprehension at posttest than the other three groups.

The percentage score of the accuracy of reading comprehension was used as outcome variable. Three dummy-coded variables with the treatment group above the basic accuracy level as reference category were entered as predictors in the multiple regression model. The reading comprehension percentage score at pretest was also entered as covariate (centered) to control for pre-training differences. Moreover, the ELFE test version (dummy coded) was entered as covariate to control for the different versions used in Study 1 (ELFE 1-6) and Study 2 (ELFE II). All predictors were entered simultaneously into the model. The results showed significant treatment effects for the trained children who were above the basic accuracy level compared to the other three groups (see Table 5 for the parameter estimates). Thus, children who were already above the basic accuracy level at pretest and then received the treatment scored higher in reading comprehension at posttest than the other three groups.

To examine whether the higher reading comprehension outcomes for the trained readers above the basic accuracy level were mediated by word-reading speed (Hypothesis 4) we ran a mediation analysis using the PROCESS macro for SPSS (Hayes, 2013, Chapter 4). Only readers above the basic accuracy level were included in this analysis. Reading comprehension at the posttest was included as an outcome variable and the words-per-minute score at the posttest as mediator. One dummy-coded variable that compared the untrained group (1) to the treatment group (0) was entered as predictor. Reading comprehension and words-per-minute pretest scores were entered as covariates (centered). Moreover, the ELFE test version (dummy coded) was also entered as covariate to control for the different versions used in Study 1 (ELFE 1-6) and Study 2 (ELFE II). A bootstrap analysis with 5000 samples (Hayes, 2013; Shrout & Bolger, 2002) revealed significant effects of the treatment condition on word-reading speed and a significant effect of word-reading speed on reading comprehension (Figure 2). However, the indirect effect on reading comprehension through word-reading speed narrowly failed to reach significance when the bias-corrected bootstrap interval was taken into consideration, Est. = -.80, 90% CI [-

2.10, 0.05] (Figure 2). Thus, we cannot conclude that the effect on reading comprehension was mediated through the orthographical decoding speed although the overall pattern of results suggests such a relationship.

Discussion

In the present study, the first purpose was to investigate whether German fourth graders have to reach a basic accuracy level at orthographical decoding before they start reading words fluently. Furthermore, to our knowledge, this is one of the first experimental studies based on a pre/post-design to examine whether the effects of a word-level reading intervention depend on students' accuracy level in orthographical decoding.

The results provided support for our first hypothesis that German fourth graders should first reach a basic accuracy level before their word-reading speed starts improving. This finding is consistent with the model of automaticity in reading proposed by LaBerge and Samuels (1974). In the first stages of reading, word recognition is to some extent accurate but still slow because of inefficient word recognition processes. Only after readers become more skilled and are able to recognize words accurately as a single unit, can they achieve full automaticity. Undoubtedly, the absolute value of the basic accuracy level is not meaningful, despite the striking similarity of the values obtained by Juul et al. (2014) and in this study (70% vs. 71% & 73%). Rather, the level is likely to vary depending on the properties of the words used as items, the task and the estimation method (as indicated by the small deviation between curvilinear analysis and piecewise regression). However, the ProDi-L items were generated by taking the skills of the typical German readers into consideration (Richter et al., 2012). Thus, we assume that a similar basic accuracy level can also be found with other word items. Nonetheless, the primary finding is that a basic accuracy level seems to exist in German primary school children and those children must achieve it before their word-reading speed increases.

In line with our predictions, the word-reading intervention fostered word-reading accuracy for poor German reading fourth graders below this basic accuracy level. Readers who received the intervention were taught the principles of segmenting regular words into syllables in the first phase, which could have served as a bridge from phonological recoding to orthographical decoding processes (Klicpera et al., 1993). Hence, the focus of the first phase was mainly to improve word-reading accuracy while the focus of the second phase was to put this new knowledge accurately and quickly into practice by reading regular and irregular words built from the frequently used syllables, become more familiar with reading via orthographical decoding, and store new words in their mental lexicon. This result is also in line with findings from other similar training approaches, in which positive effects on word-reading accuracy were shown in the German language (Klatte et al., 2014, 2016; Ritter & Scheerer-Neumann, 2009).

The results also revealed another significant treatment effect for the trained readers above the basic accuracy level. The intervention augmented their word-reading speed compared to untrained readers above the accuracy level. It appears that the development of word-reading speed can be trained and accelerated through the word-reading intervention (McArthur et al, 2015). Our result is also consistent with previous intervention studies showing that phonics trainings that focus on syllabication (Ritter & Scheerer-Neumann, 2009) and trainings that focus on repeated reading of frequent syllables (Heikkilä et al., 2013) and infrequent syllables (Huemer et al., 2010) increase the reading speed in children who take part in these interventions.

Contrary to our expectations, the comparison of word-reading speed in trained readers above and below the basic accuracy level showed no difference between these two groups. One plausible explanation for this finding could be that the reading speed of readers above the basic accuracy level reached a plateau at some point during the intervention (Breznitz, 2006; Heikkilä et al., 2013). In addition, readers below the basic accuracy level reached the basic accuracy level during the intervention. Presumably, they began making gains in word-reading speed after reaching that point and eventually caught up to the readers above the basic accuracy. However, its plausibility notwithstanding, this interpretation is speculative at this point and should be tested directly in future studies that monitor the development of students during the intervention.

Finally, the comparison of the reading comprehension of trained readers above the basic accuracy level with the other groups showed significant treatment effects. The word-reading intervention not only had positive effects on word recognition but also on reading comprehension. This finding is in accordance with the lexical quality hypothesis (Perfetti & Hart, 2002). It appears that the more accurately and fast the words were identified the more cognitive resources were available for higher levels of processing. However, the mediation analysis with bias-corrected bootstrap, failed to establish a significant indirect treatment effect on reading comprehension through word-reading speed for the comparison of the children above the basic accuracy level (Hypothesis 4), even though the effect failed to reach significance by a narrow margin. According to the guidelines of Fritz and MacKinnon (2007), to achieve a power of .80 for a simple mediation model with bias-corrected bootstrap at least 148 participants would be necessary. Hence, the failure to establish a significant mediation effect could be due to the combination of our complex mediation model (three covariates) and our small sample for this kind of analysis (106 children). Nevertheless, surprisingly the direct treatment effect on reading comprehension remained significant. It seems plausible to assume that the training fostered reading comprehension through a process that we did not measure.

In sum, the results indicate that German fourth graders must achieve a word-reading basic accuracy level before their word-reading speed starts improving. Moreover, the word-reading intervention fosters word-reading accuracy for poor German-reading fourth graders who are below this basic accuracy level and word-reading speed, regardless of the basic accuracy level.

Limitations

Our results are encouraging, given that the orthography of the German language is transparent and that most children already can read accurately by the end of the first grade (Landerl & Wimmer, 2008). However, some limitations must be considered. The words-per-minute and reading comprehension scores in the mediation model were measured concurrently. Hence, the lack of temporal precedence of the mediator weakens the conclusions about a causal relationship between these two variables. Furthermore, for more than half of the sample, information about the children's first language were missing. Thus, we cannot be certain how many of the children were German-native speakers and how many read in their non-native language. Even though language minority learners are equivalent with monolingual children in word-reading accuracy (e.g., Lesaux, Koda, Siegel, & Shanahan, 2006), they still lag behind their native-speaking peers in reading comprehension (e.g., Proctor, August, Carlo, & Snow, 2006; Lesaux, Crosson, Kieffer, & Pierce, 2010).

Another limitation is that the basic accuracy achievement time was not included in the models. The pre/post-test design used in this study included only two measuring times. Thus, the estimation of the basic accuracy achievement time was not possible. Future studies could adapt a pre/post-test design and assess word-reading accuracy and speed at more measuring times to examine whether the same basic accuracy level holds at different measurement point, that is, whether children’s word-reading speed starts to increase when a specific accuracy level is reached during the intervention (but not sooner). Furthermore, it would be possible to investigate whether the effect of the word-reading intervention is moderated by the basic accuracy achievement time. If such a study would include a sufficiently large sample of children who learned to read in their non-native language, the role of the first language and its possible effects on basic accuracy level could also be clarified.

Finally, the lexical decision task used in this study assesses the accuracy and speed of orthographic decoding and may thus be regarded as a silent word-reading task. However, it deviates from the read-aloud task (or word naming task) that has been used by Juul et al. (2014) to assess accuracy and speed of word reading. Both tasks involve word recognition but presumably tap into the underlying cognitive processes to different degrees (see, e.g., Yap et al., 2012). Most notably, the lexical decision task used in the present study puts an emphasis on lexical processes and word recognition via orthographic representations, whereas reading words aloud draws on phonological representations in the mental lexicon or efficient phonological recoding. It is reassuring that a similar accuracy-before-speed pattern occurs with both types of word-reading tasks but future studies should use both tasks within the same sample to determine commonalities and differences in a stringent manner. In particular, it would be worthwhile to test whether individuals reach the accuracy threshold, which is required before speed starts to develop, at the same time in both tasks.

Conclusion

Despite its limitations, the findings of this study are novel and promising. If similar findings were obtained for other grade levels, the basic accuracy level would become an important factor to consider when tailoring reading interventions to the educational needs of individual learners. For example, interventions for poor readers below the basic accuracy level should first focus on augmenting their accuracy until the basic accuracy level has been reached (e.g. training grapheme-phoneme correspondence rules within the syllables) and then proceed with training word-reading speed (e.g. repeated reading of words, sentences and shorts texts). In contrast, interventions for poor readers above the basic accuracy level should focus directly on word-reading speed. Replication and clarification of these findings would not only result in an increase in the efficiency of interventions but also the allocation of the available funds, which are usually scarce in research projects. Apart from that, it would also be possible to integrate these findings and exercises in the first Grades of the regular class curriculum. That way they would find direct application in the classroom context without additional costs.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csaki (Eds.), *Proceedings of the Second International Symposium on Information Theory* (pp. 267–281). Budapest: Akademiai Kiado.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48.

Berends, I. E., & Reitsma, P. (2006). Remediation of fluency: Word specific or generalised training effects? *Reading and Writing, 19*, 221–234.

Breznitz, Z. (2006). *Fluency in reading: Synchronization of processes*. Mahwah, NJ: Lawrence Erlbaum.

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest, 19*, 5–51.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*, 284–290.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences.* Mahwah, NJ: Erlbaum.

Colé, P., Magnan, A., & Grainger, J. (1999). Syllable-sized units in visual word recognition: Evidence from skilled and beginning readers in French. *Applied Psycholinguistics, 20*, 507–532.

Coltheart, M. (2005). Modeling reading: The dual-route approach. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (p. 209–226). Oxford, UK: Blackwell.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*, 204–256.

Daneman, M. (1991). Individual differences in reading skills. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, p. 512–538). Mahwah, NJ: Erlbaum.

Dudenredaktion (2017). *Duden: Die deutsche Rechtschreibung* [*Duden: The German spelling*]. Berlin, Germany: Dudenverlag.

Ecalle, J., Magnan, A. & Calmus, C. (2009). Lasting effects on literacy skills with a computer-assisted learning using syllabic units in low-progress readers. *Computers Education, 52*, 554–561.

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading, 9*, 167–188.

Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the national reading panel's meta-analysis. *Review of Educational Research, 71*, 393–447.

Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science, 18*, 233–239.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5*, 239-256.

Galuschka, K., Ise, E., Krick, K., & Schulte-Körne, G. (2014). Effectiveness of treatment approaches for children and adolescents with reading disabilities: a meta-analysis of randomized controlled trials. *PloS one, 9*, e89900. doi:10.1371/journal.pone.0089900

Galuschka, K., & Schulte-Körne, G. (2015). Evidenzbasierte Interventionsansätze und forschungsbasierte Programme zur Förderung der Leseleistung bei Kindern und Jugendlichen mit Lesestörung – Ein systematischer Review [Evidence-based intervention approaches and research-based programs to foster reading performance in children and adolescents with reading disorder - A systematic review]. *Zeitschrift für Erziehungswissenschaften, 18,* 473–487.

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2019). irr: Various Coefficients of Interrater Reliability and Agreement. R package version 0.84.1 [software] Available from https://CRAN.R-project.org/package=irr.

Gangl, M., Moll, K., Jones, M. M., Banfi, C., Schulte-Körne, G., & Landerl, K. (2018). Lexical reading in dysfluent readers of German. *Scientific Studies of Reading, 22,* 24–40.

Grainger, J., Lété, B., Bertand, D., Dufau, S., & Ziegler, J. C. (2012). Evidence for multiple routes in learning to read. *Cognition, 123,* 280–292.

Hayes, A. F. (Ed.). (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York: Guilford Press.

Heikkilä, R., Aro, M., Närhi, V., Westerholm, J., & Ahonen, T. (2013). Does training in syllable recognition improve reading speed? A computer-based trial with poor readers from second and third grade. *Scientific Studies of Reading, 17,* 398–414.

Huemer, S., Aro, M., Landerl, K., & Lyytinen, H. (2010). Repeated reading of syllables among finnish-speaking children with poor reading skills. *Scientific Studies of Reading, 14,* 317–340.

Judd, C. M., Kenny, D. A., & McClelland, G. H. (2001). Estimating and testing mediation and moderation in within-subject design. *Psychological Methods, 6*, 115–134.

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

Katz, L., Brancazio, L., Irwin, J., Katz, S., Magnuson, J., & Whalen, D. H. (2012). What lexical decision and naming tell us about reading. *Reading and Writing, 25*, 1259-1282.

Kim, Y.-S., Petcher,Y, & Foorman, B. (2015). The unique relation of silent reading fluency to end-of-year reading comprehension: Understanding individual differences at the student, classroom, school, and district levels. *Reading and Writing, 28*, 131–150.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractc Medicine, 15*, 155–163.

Klatte, M., Steinbrink, C., Bergström, K., & Lachmann, T. (2016). "Lautarium" - ein computerbasiertes Trainingsprogramm für Grundschulkinder mit Lese-Rechtschreibschwierigkeiten ["Lautarium" – A Computer-based training program for primary-school children with reading and writing difficulties]. In M. Hasselhorn und W. Schneider (Eds.), *Förderprogramme für Vor- und Grundschule* (pp. 115–141). Göttingen, Germany: Hogrefe.

Klatte, M., Steinbrink, C., Prölß, A., Estner, B., Christmann, C., & Lachmann, T. (2014). Effekte des computerbasierten Trainingsprogramms „Lautarium“ auf die phonologische Verarbeitung und die Lese-Rechtschreibleistungen bei Grundschulkindern [Effects of a computer-based training program “Lautarium” on phonological processing and reading/writing performance of primary-school children]. In G. Schulte-Körne (Ed.), *Legasthenie und Dyskalkulie: Neue Methoden zur Diagnostik und Förderung* (pp. 127–144). Bochum, Germany: Winkler.

Klicpera, C., Gasteiger-Klicpera, B., & Schabmann, A. (1993). *Lesen und Schreiben – Entwicklung und Schwierigkeiten: Die Wiener Längsschnittuntersuchungen über die Entwicklung, den Verlauf und die Ursachen von Lese- und Rechtschreibschwierigkeiten in der Pflichtschulzeit* [Reading and writing - Development and difficulties: The Viennese longitudinal studies on the development, the course and the causes of reading and writing difficulties in compulsory education]. Bern, Switzerland: Huber.

Knoepke, J., Richter, T., Isberner, M.-B., Naumann, J., & Neeb, Y. (2014). Phonological recoding, orthographic decoding, and comprehension skills during reading acquisition. *Zeitschrift für Erziehungswissenschaft, 17*, 447–471.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2019). lmerTest: Tests in linear mixed effects models. R package version 3.1-0 [software] Available from https://CRAN.R-project.org/package=lmerTest.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*, 293–323.

Landerl, K., & Wimmer, H. (2008). Development of word reading fluency and spelling in a consistent orthography: An 8-year follow-up. *Journal of Educational Psychology, 100,* 150–161.

Lenhard, W., Lenhard, A., & Schneider, W. (2017). *ELFE II: Ein Leseverständnistest für Erst- bis Siebtklässler* [ELFE II: A reading comprehension test for children in grades one to six]. Göttingen, Germany: Hogrefe.

Lenhard, W., & Schneider, W. (2006). *ELFE 1-6: Ein Leseverständnistest für Erst- bis Sechstklässler* [ELFE 1-6: A reading comprehension test for children in grades one to six]. Göttingen, Germany: Hogrefe.

Lesaux, N. K., Crosson, A. C., Kieffer, M. J., & Pierce, M. (2010). Uneven profiles: Language minority learners' word reading, vocabulary, and reading comprehension skills. *Journal of Applied Developmental Psychology, 31*, 475–483.

Lesaux, N. K., & Koda, K., Siegel, L., & Shanahan, T. (2006). Development of Literacy. In D. August & T. Shanahan (Eds.), *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth* (pp. 75–122). Mahwah, NJ, US: Lawrence Erlbaum.

Martens, V. E. G., & de Jong, P. F. (2006). The effect of word length on lexical decision in dyslexic and normal reading children. *Brain and Language*, *98*, 140–149.

Mayringer, H., & Wimmer, H. (2000). Pseudoname learning by German-speaking children with dyslexia: Evidence for a phonological learning deficit. *Journal of Experimental Child Psychology*, *75*, 116–133.

McArthur, G., Castles, A., Kohnen, S., Larsen, L., Jones, K., Anandakumar, T., & Banales, E. (2015). Sight word and phonics training in children with dyslexia. *Journal of Learning Disabilities*, *48*, 391–407.

McArthur, G., Sheehan, Y., Badcock, N. A., Francis, D. A., Wang, H. C., Kohnen, S., . . . Castles, A. (2018). Phonics training for English-speaking poor readers. *Cochrane Database of Systematic Reviews*, *11*, CD009115.

McGraw, K. O., & Wong S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*, 30–46.

Meyer, M. S., & Felton, R. H. (1999). Repeated reading to enhance fluency: Old approaches and new directions. *Annals of Dyslexia, 49*, 283–306.

Muggeo, V. M. R. (2017). segmented: Regression models with break-points / Change-points estimation. R package version 0.5-3.0 [software] Available from https://CRAN.R-project.org/package=segmented.

Müller, B., Otterbein-Gutsche, G., & Richter, T. (2018). Leseförderung mit Silben und Sprachsystematik: Konzeption eines Trainingsprogramms [Systematic reading improvement using syllables: An intervention program]. *Psychologie in Erziehung und Unterricht*, *65*, 52–57.

Müller, B., & Richter, T. (2014). Lesekompetenz [Reading competence]. In J. Grabowski (Ed.), *Sinn und Unsinn von Kompetenzen: Fähigkeitskonzepte im Bereich von Sprache, Medien und Kultur* (p. 29–49). Leverkusen: Budrich.

Müller, B., Richter, T., Karageorgos, P., Krawietz, S., & Ennemoser, M. (2017). Effects of a word-reading reading intervention in poor-reading fourth graders. *Frontiers Psychology, 8*, 1635.

Perfetti, C. A. (1985). *Reading ability*. New York, NY: Oxford University Press.

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189-213). Amsterdam, The Netherlands: John Benjamins.

Perfetti, C. A., Landi, N., & Oakhill J. (2005). In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (p. 227–247). Oxford, UK: Blackwell.

Pikulski, J., & Chard, D.J. (2005). Fluency: Bridge between decoding and reading comprehension. *The Reading Teacher, 58*, 510–519.

Proctor, C. P., August, D., Carlo, M. S., & Snow, C. (2006). The intriguing role of Spanish language vocabulary knowledge in predicting English reading comprehension. *Journal of Educational Psychology, 98*, 159–169.

Protopapas, A., Sideridis, G. D., Mouzaki, A., & Simos, P. G. (2007). Development of lexical mediation in the relation between reading comprehension and word reading skills in Greek. *Scientific Studies of Reading, 111*, 165–197.

Price, R. B., Kuckertz, J. M., Siegle, G. J., Ladouceur, C. D., Silk, J. S., Ryan, N. D., . . . Amir, N. (2015). Empirical recommendations for improving the stability of the dot-probe task in clinical research. *Psychological Assessment, 27*, 365–376.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Richter, T., Isberner, M.-B., Naumann, J., & Kutzner, Y. (2012). Prozessbezogene Diagnostik von Lesefähigkeit bei Grundschulkindern [Process-orientedassessment of reading skills in primary school children]. *Zeitschrift für Pädagogische Psychologie, 26*, 313–331.

Richter, T., Isberner, M.-B., Naumann, J. & Kutzner, Y. (2013). Lexical quality and reading comprehension in primary school children. *Scientific Studies of Reading*, *17*, 415–434.

Richter, T., Naumann, J., Isberner, M.-B., Neeb, Y., & Knoepke, J. (2017). *ProDi-L: Prozessbasierte Diagnostik von Lesefähigkeiten bei Grundschulkindern [ProDi-L: Process-oriented assessment of reading skills in primary school children; computerized test]*. Göttingen, Germany: Hogrefe.

Ritter, C., & Scheerer-Neumann, G. (2009). *PotsBlitz: Das Potsdamer Lesetraining [PotsBlitz: The Potsdam reading training]*. Köln, Germany: ProLog.

Scheerer-Neumann, G. (1981). The utilization of intraword structure in poor readers: Experimental evidence and a training program. *Psychological Research, 43*, 155–178.

Schroeder, S., Würzner, K.-M., Heister, J., Geyken, A., & Kliegl, R. (2015). childLex: A lexical database of German read by children. *Behavior Research Methods, 47*, 1085–1094.

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94*, 143–174.

Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology, 72*, 95–129.

Share, D. L. (2004). Orthographic learning at a glance: On the time course and developmental onset of self-teaching. *Journal of Experimental Child psychology, 87*, 267–298.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7*, 422–445.

Snowling, M. J., & Hulme, C. (2005). *The science of reading: A handbook*. Oxford: Blackwell.

Stevens, E. A., Walker, M. A., & Vaughn, S. (2017). The effects of reading fluency interventions on the reading fluency and reading comprehension performance of elementary students with learning disabilities: A synthesis of the research from 2001 to 2014. *Journal of Learning Disabilities, 50*, 576–590.

Suggate, S.P. (2010). Why what we teach depends on when: Grade and reading intervention modality moderate effect size. *Developmental Psychology, 46*, 1556–1579.

Suggate, S.P. (2016). A meta-analysis of the long-term effects of phonemic awareness, phonics, fluency, and reading comprehension interventions. *Journal of Learning Disabilities, 49*, 77–96.

Therrien, W. J. (2004). Fluency and comprehension gains as a result of repeated reading: A meta-analysis. *Remedial and Special Education, 25*, 252–261.

Trochim, W. M., Donnelly, J. P., & Arora, K. (2016). *Research methods: The essential knowledge base.* Boston, MA: Cengage Learning.

Urbina, S. (2014) *Essentials of psychological testing*. New York: John Wiley & Sons.

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry, 45*, 2–40.

Wentink, H. W. M. J., Van Bon, W. H. J., & Schreuder, R. (1997). Training of poor readers' phonological decoding: Evidence for syllable-bound processing. *Reading and Writing, 9*, 163–192.

Yap, M. J., Balota, D. A., Sibley, D. E., & Ratcliff, R. (2012). Individual differences in visual word recognition: Insights from the English Lexicon Project. *Journal of Experimental Psychology: Human Perception and Performance, 38*, 53-79.

Ziegler, J. C., Perry, C., Ma-Wyatt, A., Ladner, D., & Schulte-Körne, G. (2003). Developmental dyslexia in different languages: Language-specific or universal? *Journal of Experimental Child Psychology, 86*, 169–193.

Zoccolotti, P., De Luca, M., Di Pace, E., Gasperini, F., Judica, A., & Spinelli, D. (2005). Word length effect in early reading and in developmental dyslexia. *Brain and Language, 93*, 369–373.

Tables

Table 1

Characteristics of the Samples and Mean Values for Accuracy and Speed of Orthographical Decoding and Reading Comprehension at Pretest by Treatment Condition

	Screening sample	Treatment group	Wait-list control group
N Total sample	826	105	65
n Study 1	309	43	32
n Study 2	517	62	33
Number of females	388	56	28
Age in years	9.21 (0.69)	9.44 (0.72)	9.36 (0.80)
Orthographical decoding, ProDi-L			
Accuracy, Percentage Score			
M (*SD*)	85.78 (11.55)	77.46 (12.87)	78.94 (11.24)
Speed, Words per Minute			
M (*SD*)	33.22 (12.04)	28.78 (13.24)	27.79 (9.76)
Reading Comprehension (ELFE, *T*-values)			
M (*SD*)	48.12 (9.78)	41.12 (7.69)	39.22 (6.23)

Note. Word recognition = 16-item lexical decision task (subtest of ProDi-L, Richter et al., 2017). Reading comprehension = *T*-value of the sum of correct responses compared to the class norms (Study 1: 20-item ELFE 1-6, Lenhard & Schneider, 2006; Study 2: 26-item ELFE II, Lenhard et al., 2017).

Table 2

Fit Indices of Polynomial Models with the Words-per-Minute Score as Outcome and Orthographical Decoding Accuracy as Predictor

Model	R^2	ΔR^2	log-likelihood	χ^2	AIC
1 Linear	0.05	0.05	-2628.63	38.00**	5263.27
2 Quadratic	0.07	0.02	-2620.15	16.97**	5248.30
3 Cubic	0.08	0.01	-2618.01	4.28*	5246.03

Note. All metric predictors were centered.

* p < .05. ** p < 01.

Table 3

Summary of Means and Standard Deviations of Accuracy and Speed of Orthographical Decoding and Reading Comprehension by Treatment Condition (Treatment group vs. Wait-List Group) Above and Below the Basic Accuracy Level of 71% at Pre- and Posttest

	Orthographical decoding accuracy					Orthographical decoding speed					Reading comprehension				
	Pretest		Posttest		Cohen's *d*	Pretest		Posttest		Cohen's *d*	Pretest		Posttest		Cohen's *d*
	M (*SD*)	*N*	*M* (*SD*)	*N*		*M* (*SD*)	*N*	*M* (*SD*)	*N*		*M* (*SD*)	*N*	*M* (*SD*)	*N*	
Experimental group above	83.12 (7.32)	77	84.77 (10.86)	71	0.18	27.88 (9.89)	76	35.18 (13.15)	67	0.63	37.39 (19.13)	76	55.20 (22.43)	70	0.86
Experimental group below	63.5 (8.96)	25	79.89 (13.32)	23	1.46	29.18 (16.13)	23	34.94 (13.38)	22	0.39	31.31 (10.50)	25	39.31 (17.44)	23	0.56
Wait-list group above	83.62 (7.02)	50	85.73 (10.76)	46	0.23	27.80 (9.70)	47	30.20 (11.66)	46	0.22	31.69 (12.57)	50	42.37 (15.32)	46	0.77
Wait-list group below	63.33 (8.14)	15	68.75 (20.80)	14	0.35	27.77 (10.33)	14	33.91 (12.13)	13	0.55	28.21 (11.31)	15	34.44 (11.25)	15	0.55

Note. Above = above the basic accuracy level; below = below the basic accuracy level.

Table 4

Parameter Estimates for Multiple Regression Analysis with the Accuracy of Word Recognition as Outcome Variables at the Posttest, Treatment Condition as Predictor, and Pretest Scores as Covariates.

	Orthographical decoding accuracy		
	B	*SE*	ΔR^2
Intercept	84.85**	3.24	
Trained below (reference group) vs. trained above	-1.56	3.92	.00
Trained below (reference group) vs. untrained above	-0.76	4.10	.00
Trained below (reference group) vs. untrained below	-11.25*	4.13	.04
Pretest scores	0.32*	0.13	.03
Goodness of fit	R^2 = .17, F(4,149) = 7.64, p < .001		

Note. The pretest scores were centered. The predictors representing treatment conditions were dummy-coded (trained below as reference group).

* p < .05; ** p < .01 (one-tailed).

Table 5

Parameter Estimates for Multiple Regression Analyses with the Speed of Word Recognition and Reading Comprehension as Outcome Variables at the Posttest, Treatment Condition as Predictor, and Corresponding Pretest Scores as Covariates.

	Orthographical decoding speed			Reading comprehension		
	B	*SE*	ΔR^2	*B*	*SE*	ΔR^2
Intercept	34.89**	1.27		51.77**	2.14	
Trained above (reference group) vs. trained below	-0.86	2.63	.00	-10.26*	3.50	.03
Trained above (reference group) vs. untrained above	-4.21*	2.02	.02	-8.01*	2.80	.03
Trained above (reference group) vs. untrained below	-1.81	3.23	.00	-13.04*	4.15	.03
Pretest scores	0.66**	0.08	.33	0.80**	0.08	.34
Elfe Version				0.43	2.58	.00
Goodness of fit	R^2 = .36, F(4,136) = 19.39, $p < .001$			R^2 = .51, F(5,147) = 30.93, $p < .001$		

Note. The pretest scores were centered. The predictors representing treatment conditions were dummy-coded (trained above as reference group). The test version of the Elfe test was entered as dummy-coded predictor.

* $p < .05$; ** $p < .01$ (one-tailed).

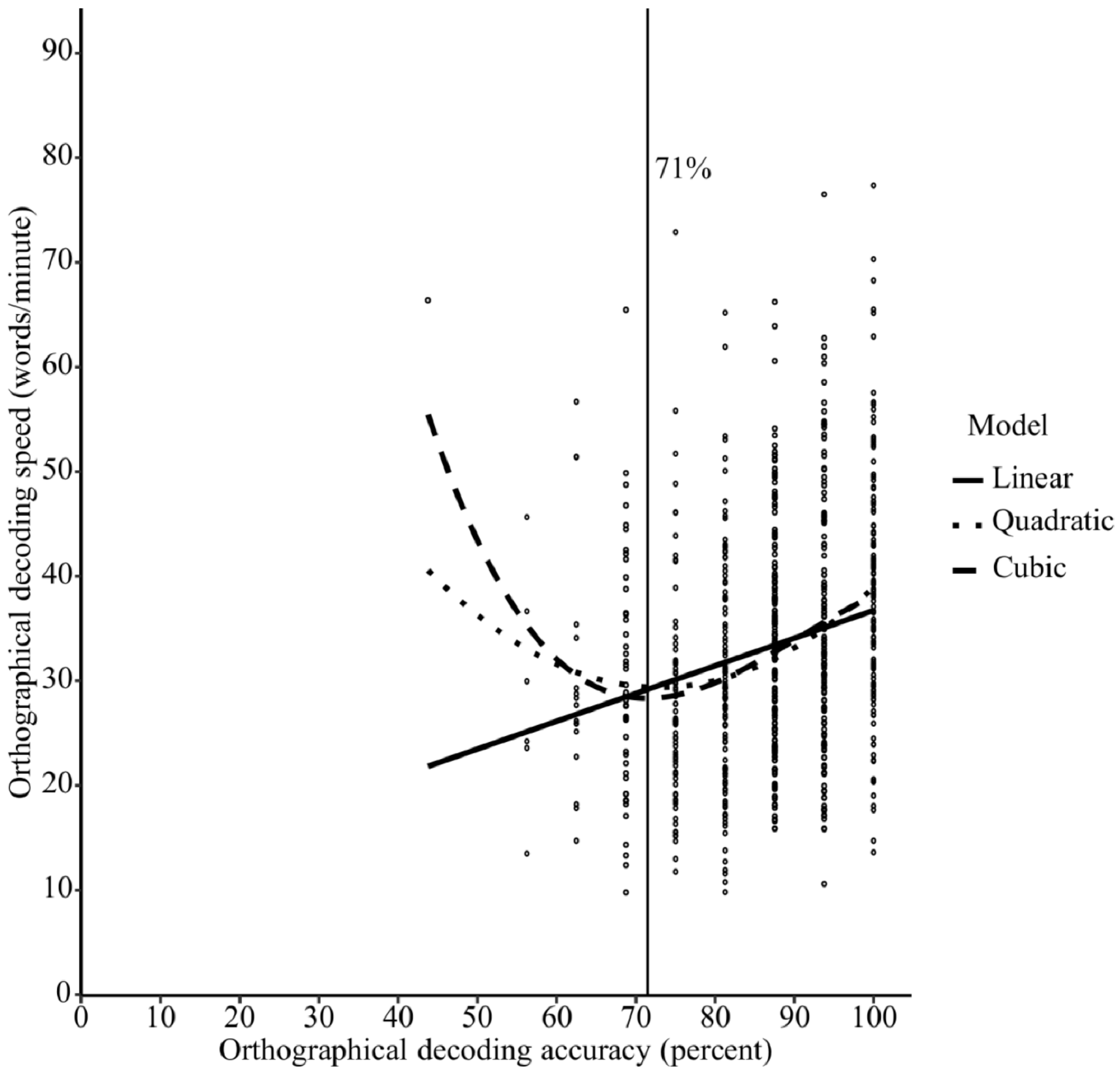

Figure 1. Polynomial curve fitting points up to third-degree for orthographical decoding accuracy (percentage score of the sum of correct answers in orthographical decoding) and orthographical decoding speed (words-per-minute score based on the accuracy and latency in orthographical decoding; $N = 824$).

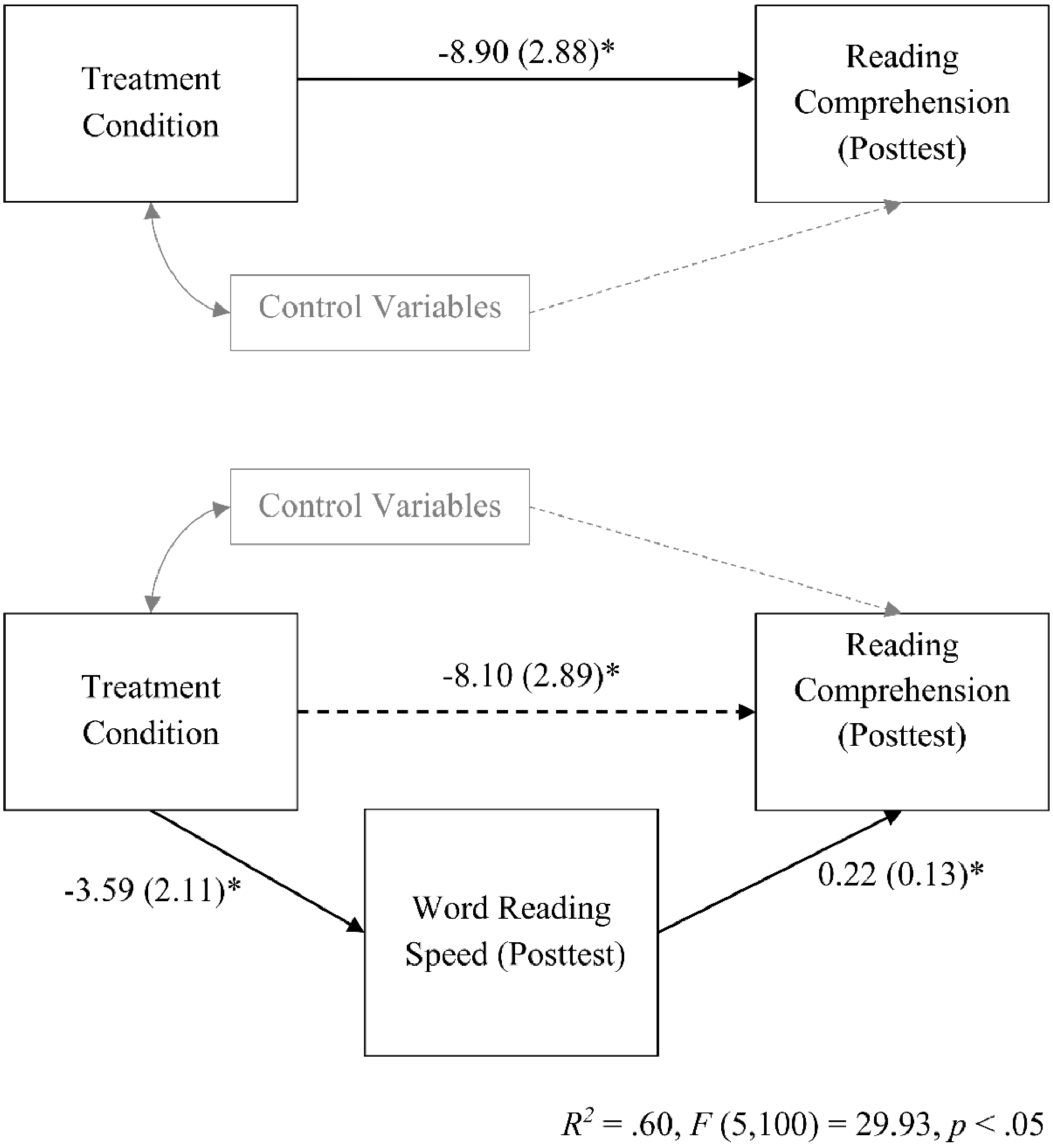

Figure 2. Mediation model for the comparison of the treatment group above the basic accuracy level with the untrained children above the basic accuracy level (dummy-coded, treatment group above the basic accuracy level as reference group) on reading comprehension at posttest with orthographical decoding speed at posttest as mediator. Unstandardized regression weights with associated standard errors in parentheses (* $p < .05$, one-tailed).

Ethics statement

According to the ethical guidelines of the German Society for Psychology (DGPs) and regulations of the local ethics committee, prior review by an Institutional Review Board is not mandatory for research that provides signed informed consent from study participants; collects data anonymously; and has no foreseeable negative impact on participants. Furthermore, this study was carried out in accordance with the recommendations of "Ministry of Education and Cultural Affairs, Hesse, Germany Hessisches Kultusministerium)" with written informed consent from all subjects. The parents of all subjects gave written informed consent in accordance with the Declaration of Helsinki. Principals of the schools participating in the study gave written consent after the school conference (i.e., the majority of teachers agreed to realize the study in their school). The protocol was approved by the "Ministry of Education and Cultural Affairs, Hesse" (cf. Education Act of Hesse, section 84).