The Role of Word-Recognition Accuracy in the Development of Word-Recognition Speed and Reading Comprehension in Primary School: A Longitudinal Examination

Panagiotis Karageorgos[a], Tobias Richter[a], Maj-Britt Haffmans[b], Julia Schindler[a], and Johannes Naumann[c]

[a] University of Würzburg, Department of Psychology IV, Educational Psychology, Röntgenring 10, 97070, Würzburg, Germany

[b] University of Kassel, Department of Psychology, Developmental Psychology, Holländische Str. 36-38, 34127, Kassel, Germany

[c] University of Wuppertal, Department of School of Education, Research Methods in Educational Science, Rainer-Gruenter-Straße 21, 42119, Wuppertal, Germany

Accepted for publication in the journal *Cognitive Development* (2020)

Corresponding author: Panagiotis Karageorgos

University of Würzburg, Department of Psychology IV

Röntgenring 10

97070 Würzburg, Germany

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

phone: +49 931/ 31 80874

E-mail addresses: panagiotis.karageorgos@uni-wuerzburg.de (P. Karageorgos), tobias.richter@uni-wuerzburg.de (T. Richter), maj-britt.haffmans@uni-kassel.de (M.-B. Haffmans), julia.schindler@uni-wuerzburg.de (J. Schindler), j.naumann@uni-wuppertal.de (J. Naumann).

Abstract

Recent studies suggest that readers need to reach a certain word-recognition accuracy threshold first before word-recognition speed starts to improve. In a longitudinal study, 1,095 German primary school children were followed from Grades 1 to 4. Word-recognition accuracy and speed were assessed at the end of Grades 1 to 4 and reading comprehension at the end of Grades 2 to 4. The growth curves of word-recognition speed and reading comprehension were hypothesized to be steeper for children who achieved a basic word-recognition accuracy of 71% compared to children who failed to reach this threshold by the end of Grade 1. Multilevel growth models revealed that the improvement of word-recognition speed and reading comprehension was more pronounced for children who reached the critical threshold by the end of Grade 1. Moreover, the overall pattern was that children who reached the basic word-recognition accuracy in later grades showed flatter trajectories of word-recognition speed and reading comprehension over the primary school years. These findings suggest that good word-recognition accuracy lays the foundation for the development of word-recognition speed and reading comprehension in German primary school children.

Keywords: primary school children, reading development in German, word-recognition accuracy, word-recognition speed, reading comprehension

Introduction

Learning to read is a laborious process that starts at younger ages and improves with time and effort. Starting from basic reading processes at the word level, such as efficiently recognizing written words, readers must also become proficient in higher levels of cognitive processing such as integrating words syntactically and semantically and establishing a coherent mental model of the text (Kintsch, 1998; Perfetti, Landi, & Oakhill, 2005; Schindler & Richter, 2018; Torgesen, 2000; Verhoeven & Perfetti, 2008). Efficient (i.e., fast and accurate) visual word recognition saves cognitive resources which then become available for higher-order cognitive processes needed to achieve good text comprehension (*verbal efficiency hypothesis*, Perfetti, 1985). In contrast, inefficient visual word recognition is likely to impair text comprehension (Perfetti, 1985; Pikulski & Chard, 2005). Thus, fluent and reliable word recognition is a fundamental precondition for good reading comprehension (Perfetti & Hart, 2002; Gough & Tunmer 1986; Hoover & Gough 1990).

In a longitudinal study with Danish children, Juul, Poulsen and Elbro (2014) found that readers' word-recognition accuracy needed to reach a certain threshold before their word-recognition speed began to improve. Their results also suggest that the time when children reach this threshold in reading development (i.e., the basic accuracy achievement time) is also an important predictor of the development of word-recognition speed. In line with Juul et al. (2014), a study with German fourth graders by Karageorgos, Müller and Richter (2019) supported the existence of such a basic accuracy threshold. The present study used longitudinal data to investigate the effect of basic accuracy achievement time on the development of word-recognition speed and reading comprehension in German primary school children.

The Development of Word Reading Processes

According to the *dual route cascaded model* (DRC) proposed by Coltheart, Rastle, Perry, Langdon, and Ziegler (2001), written words can be recognized via two main routes:

phonological recoding and direct word recognition via orthographical representations of word forms (see also Bowey & Muller, 2005; Ehri, 2005; Perfetti & Stafura, 2014). Recognizing words via phonological recoding requires readers to acquire the alphabetic principle and apply the grapheme-phoneme correspondence rules of their specific language to recode each grapheme into its corresponding phoneme (Coltheart et al., 2001; Ziegler & Goswami, 2005). This process is laborious, it strains working memory, and it is prone to errors (Müller & Richter, 2014; Snowling & Hulme, 2005). However, phonological recoding enables beginning readers to recognize new or unfamiliar words and to acquire orthographic representations of word forms, which are then stored in *sight vocabulary* (Ehri, 2005; Grainger et al., 2012; Share, 1999, 2004). These orthographic representations are the basis for fast word recognition via orthographical decoding (Grainger, Lété, Bertand, Dufau, & Ziegler, 2012; Kyte & Johnson, 2006; Share, 1995, 1999).

In orthographical decoding, orthographic representations of written words are directly retrieved from the mental lexicon without laboriously recoding them letter-by-letter (Coltheart et al., 2001). This alternative route is usually faster and more efficient than phonological recoding (Acha & Perea, 2008). When orthographical decoding is sufficiently automatized, readers recognize words effortlessly, which saves cognitive resources for higher levels of processing at the sentence and text levels (Bowey & Muller, 2005; Klauda & Guthrie, 2008; Perfetti & Hart, 2002), which are necessary for reading comprehension (e.g., Kendeou, van den Broek, Helder, & Karlson, 2014; Kintsch, 1998; Torgesen, 2000).

Another well-known model that describes the development of visual word reading and has many similarities to the DRC model is the *three-phase developmental model* proposed by Frith (1985, 1986). According to Frith, a reader must go through three basic phases in word recognition (logographic, alphabetic, and orthographic) to achieve the level of a skilled reader. In the logographic phase, readers depend on graphic distinguishing features, such as the first letter or other cues, to recognize familiar words. Beginning readers must master this strategy before they can move on to the alphabetic phase. In the alphabetic phase, readers use the grapheme-phoneme correspondence rules to recode words letter-by-letter. Only after children master the alphabetic phase can they proceed to the orthographic phase. In this last phase, readers can recognize words as a whole without the use of alphabetic skills. Hence, the alphabetic and orthographic phases correspond to the phonological recoding and orthographical decoding skills described in the DRC model.

Even though both models support the notion that at some point a shift occurs from phonological recoding skills to orthographical decoding skills, this shift is described differently in each model (Knoepke, Richter, Isberner, Naumann, & Neeb, 2014). Whereas the three-phase model supports the notion that strategies are attained in a serial manner and each strategy builds on top of the existing one, the DRC model assumes that both phonological and orthographical skills are relevant in visual word recognition and that skilled readers utilize both ways, whichever one is more efficient. The assumption of the parallel use of phonological and orthographical skills in word recognition is also in accordance with several studies showing phonological and orthographical priming effects for skilled readers (e.g., Booth, Perfetti, & MacWhinney, 1999; Ziegler, Bertrand, Lété, & Grainger, 2014). Nevertheless, although the role of phonological recoding is pivotal at the early stages of reading, fluent reading and text comprehension require reliable and fluent orthographical decoding skills (Castles, Rastle, & Nation, 2018). With increasing reading skills, reliance on phonological recoding in word reading decreases, although it continues to have a robust influence on text comprehension (Acha & Perea, 2008; Grainger et al., 2012; Ziegler et al., 2014).

The Role of Visual Word Recognition in German

Noticeable variability exists across orthographic systems as to when beginning readers recognize written words efficiently, that is, reliably and fluently (Seymour, Aro, & Erskine, 2003; Spencer & Hanley, 2003). In languages with a relatively shallow orthography such as

German, phonological recoding skills are usually acquired within the first year of formal reading instruction, whereas readers of languages with an opaque orthography (with ambiguous grapheme-phoneme mappings) such as Danish need twice as long (Landerl & Wimmer, 2008; Seymour et al., 2003). In an eight-year longitudinal study by Landerl and Wimmer (2008), even poor German readers (fluency scores one standard deviation below the group mean) correctly recognized 72% of written words presented to them and 61% of pseudowords at the end of Grade 1, although large individual differences were found in word-recognition speed (see also Gangl et al., 2018). Even though their word-recognition fluency improved throughout the duration of the study, individual differences between skilled and poor readers remained stable by the end of Grade 8. A similar finding was reported by Richter, Isberner, Naumann, and Kutzner (2012). These authors found a monotonous increase in the accuracy of word and pseudoword recognition across primary school years, with ceiling effects at the end of Grade 4, whereas the average recognition speed increased log-linearly and individual differences remained stable by the end of Grade 4. These findings indicate that skilled and poor German readers differ not so much in word-recognition accuracy but rather in word-recognition speed (Mayringer & Wimmer, 2000; Wimmer, 1996; Ziegler et al., 2003). Apparently, good readers depend on efficient phonological recoding and orthographical decoding skills to recognize words, whereas poor readers get stuck in the phonological phase (e.g., Ehri, 2005).

An important point is that word-recognition accuracy and speed might not develop independently. According to Juul et al. (2014), readers have to reach a certain word-recognition accuracy-threshold before word-recognition speed begins to improve. If readers stay below this threshold, the reliable recognition of words is likely to make high demands on cognitive resources. Consequently, word-recognition speed is impaired (Castles et al., 2018; LaBerge & Samuels, 1974). Juul et al. (2014) showed that word-recognition speed improved only when children reached a 70% accuracy threshold. Significant positive correlations between word-recognition accuracy and word-recognition speed were found above this basic accuracy threshold but not below. Basic accuracy-achievement time also had a significant effect on the development of word-recognition speed. The sooner a child reaches the threshold, the more time they have at their disposal to improve their word-recognition speed. In line with Juul et al.'s (2014) findings, Karageorgos et al. (2019) found evidence in favour of a 71% accuracy threshold in German fourth graders. In this study, response accuracy and speed in a lexical decision task were used to determine the relationship of word-recognition accuracy and speed. Based on cross-sectional data, the basic accuracy threshold was determined and validated by applying curve-fitting analyses and piecewise regression models. Moreover, Karageorgos et al. showed that the effects of an intervention developed to foster poor readers' word-recognition skills had differential effects on word-recognition accuracy and speed, depending on whether children had already reached the basic accuracy threshold or not. However, a systematic longitudinal analysis of the relationship between word-recognition accuracy and speed that examines the role of basic accuracy-achievement time for the development of word-recognition speed and reading comprehension is still lacking.

The Present Study

The aim of the present study was to investigate the development of word-recognition speed in relation to word-recognition accuracy and reading comprehension in a longitudinal study with German primary school children. Although poor German readers usually can read words accurately to some extent through phonological recoding, they still score lower than good readers in word-recognition speed (e.g., Ziegler, Perry, Ma-Wyatt, Ladner, & Schulte-Körne, 2003). According to Knoepke et al. (2014), direct word recognition via the lexical route by orthographical decoding has a stronger relationship to reading speed and reading comprehension in German primary school children. We expected children who reach the 71% accuracy threshold by the end of Grade 1 to show a steeper overall growth curve for word-recognition speed up to Grade 4 compared to children who fail to reach the threshold by then.

Moreover, we expected the overall growth curve for word-recognition speed to be flatter for readers who reach the basic accuracy threshold at later grade levels. Consequentially, children who do not reach the basic accuracy threshold until Grade 4 should show the least growth in word-recognition speed over the primary school years. In addition, children reaching the threshold at earlier grades as a result of steeper and earlier development of reading speed were expected to recognize words faster at the end of Grade 4 than children reaching the threshold at later grades and children not reaching the threshold.

Furthermore, according to the lexical quality hypothesis (Perfetti & Hart, 2002), efficient (i.e., accurate and fast) visual word recognition facilitates reading comprehension by freeing cognitive resources necessary for higher-order reading processes. Therefore, children who reach the threshold of basic accuracy by the end of Grade 1 were also expected to show a more pronounced growth curve for reading comprehension up to Grade 4 compared to children who fail to reach the threshold by then. Again, we expected the growth curve for reading comprehension to be flatter for readers reaching the basic accuracy threshold in later grades. Children failing to reach the basic accuracy threshold by Grade 4 should show the least growth in reading comprehension over the primary school years. Finally, children who take less time to reach the threshold were expected to comprehend texts better at the end of Grade 4 than children who reach this threshold at later grades and children who fail to reach the threshold.

Method

Design and Procedure

Data were collected in the context of a longitudinal study with two cohorts of primary school children and four measurement points at the end of Grades 1 to 4. Word-recognition accuracy and speed were assessed at all four measurement points. Reading comprehension was assessed at the end of Grade 2 to Grade 4. The first cohort started primary school in 2011 and the second cohort in 2012 (see descriptive statistics in Table 1). The study was conducted in the urban areas of Frankfurt am Main and Kassel, Germany. In the process of recruitment, principals of primary schools in both areas were contacted by mail or phone. Parents of children in schools whose principals and teachers expressed interest to participate in the study received further information material only after the school council gave its' consent.

- INSERT TABLE 1 ABOUT HERE –

- INSERT TABLE 2 ABOUT HERE –

Participants

A total of 1,095 children (510 females, 524 males, 61 missing gender information) from 68 classes participated in the study. Out of the 1,095 children, 62 children participated only at the first measurement point, 67 children joined the study after Grade 1, and 692 children had no missing values at any of the measurement points. The missing values in our data set are due to random factors such as illnesses or relocation of the participants (see Table 2 for the attrition rates). The children's parents or guardians gave informed written consent to participation in the study. At the end of Grade 1, the children were on average 7.44 years old (*SD* = 0.45). Parents of 656 children reported that German was their child's first language, parents of 87 children reported that their children grew up bilingually with German as the first language, and parents of 246 children reported that their children spoke German as a second language. Parents of 106 children provided no information about their child's first language. Sociodemographic information was collected via parents' and teachers' questionnaires.

Measures

Word-recognition Accuracy and Speed. Word-recognition accuracy and speed were assessed with the computerized lexical decision subtest of the ProDi-L test battery (Richter, Naumann, Isberner, Neeb, & Knoepke, 2017; see also Richter, Isberner, Naumann, & Kutzner, 2013). Children were presented with 16 words (e.g., *Traktor* [tractor]) and pseudowords (e.g., *Spinfen*) in randomized order. Their task was to decide whether the presented letter string was a real word or not by using two response keys (*yes*/*no*).

Pseudowords were orthographically and phonologically legal letter strings and varied in their similarity to actual German words. Pseudowords similar to actual words were constructed by changing the first character of an existing word (e.g, *Name* → *Bame*). Pseudowords dissimilar to actual words were constructed by combining the syllables of two existing words with irregular spellings. For example, the pseudoword *Chilance* was constructed by combining the first syllable of the word Chili and the second and third syllables of the word Balance. The pseudowords also included pseudohomophones (1-3 per measurement point), which sound like real words but have a different orthographical form (e.g., *Heckse* instead of *Hexe*/*witch*). These items cannot be solved via the application of phoneme-grapheme translation rules but require direct word recognition via the lexical route. Seven items in the first measurement point, nine items in the second measurement point and eight items in the last two measurement points were regular and irregular real German words. Different but parallel words and pseudowords were used at all four measurement points. Apart from the slight difference in the proportions of words and pseudowords in the first and second item set (which was due to an error), the item sets were strictly parallelized according to the item features, mean accuracy and mean response time of each item set which were obtained in another cross-sectional study (Richter et al., 2013).

The word stimuli were systematically varied in frequency and number of orthographical neighbors. They had an average frequency of 1.25 (*SD* = .87), retrieved from the CELEX German lemma lexicon (metric: Mannheim written frequency, logarithmic; Baayen, Piepenbrock, & Gullikers, 1995; Baayen, Piepenbrock, & van Rijn, 1993), an average length of 5.62 (*SD* = 1.56) characters and on average 1.75 (*SD* = 2.46) orthographical neighbours. The pseudowords were matched in length and frequency to the word stimuli. Pseudowords were based on words with an average frequency of 1.03 (*SD* = .66), they had an average length of 6.31 (*SD* = 2.16) characters and on average 1.69 (*SD* = 3.25) orthographical neighbours. In order to examine whether words and pseudowords differed in frequency, length, and orthographical neighbours across the measurement points we ran three separate analyses of variance. The results indicated no significant differences (for all comparisons, $p > .17$) between words and pseudowords across the measurement points. These results suggest that largely parallel items were used at each measurement point.

Following the ProDi-L manual, two criteria were applied to identify and remove outliers. Logarithmic latencies that were three standard deviations below or above the mean logarithmic latency for the item in the norming sample were coded as missing. The idea behind this criterion is that very short response times are likely to indicate an irregular response, such as clicking through items without reading them, and thus they should not be included in further analyses. Likewise, very long response times are likely due to disturbances, mind wandering, etc. Furthermore, for each child, response times that deviated more than two standard deviations from the average of the individual logarithmic response times were also coded as missing. Further data preparation was performed separately for each measurement point according to the procedure reported by Karageorgos et al. (2019). The sum of correct responses was transformed into proportions representing word recognition accuracy. Furthermore, a words-per-minute score was calculated as an indicator of word-recognition speed. The number of correct and incorrect responses to words and pseudowords was multiplied by 60,000 ms and then divided by the overall latency across all items measured in ms. A child, for example, who responded to 10 items in 10,000 ms received a score of 60 words per minute. Words-per-minute scores were not computed for participants with more than 10% missing values (due to the outlier removal criteria discussed above) at the relevant measurement point. Thus, word-recognition speed scores were missing for 449 of 4,380 data points. The test-retest reliability between measurement points was computed as the intraclass correlation of word-recognition scores at the end of each school year for a total of 692 children (those with complete data sets) using the R-package *irr* (Gamer, Lemon, Fellows, & Singh, 2019). A two-way mixed-effects model for mean rating and absolute agreement was used for computing the ICC (Koo & Li, 2016; McGraw & Wong, 1996; Price et al., 2015). According to the interpretation guidelines proposed by Cicchetti (1994), the estimated test-retest reliability (i.e., stability) was good for the accuracy score, $\rho_I = .624$, $F(691, 32.9) = 3.48$, $p < 0.001$, 95% CI [.40, .75], and fair for the words-per-minute score, $\rho_I = .50$, $F(543, 6.7) = 4.46$, $p = 0.005$, 95% CI [.01, .73].

Reading Comprehension. Reading comprehension was assessed with the computerized text comprehension subtest of the standardized German reading test ELFE 1-6 (Lenhard & Schneider, 2006). At the end of each school year (starting at Grade 2), children were presented with the same 13 texts (two to five sentences) in randomized order and were asked to answer one to three multiple-choice questions about the contents of these texts (literal comprehension and inference questions). The sum of correct responses was transformed into proportions representing text comprehension. Again, the test-retest reliability between measurement points was computed as the intraclass correlation of comprehension scores for a total of 658 children with complete data sets for the comprehension scores. A two-way mixed-effects model for mean rating and absolute agreement was used (Price et al., 2015; Koo & Li, 2016; McGraw & Wong, 1996). The estimated test-rest reliability was good, $\rho_I = .69$, $F(657, 7.8) = 5.92$, $p = 0.005$, 95% CI [.21, .85].

Results

Equivalence of Cohorts

To examine whether the two cohorts developed differently over time, multilevel growth models (measurement points nested within students) with random intercepts were estimated with the R-packages *lme4* (Bates, Maechler, Bolker, & Walker, 2019) and *lmerTest* with Satterthwaite adjustments to denominator degrees of freedom (Kuznetsova, Brockhoff, & Christensen, 2019). Separate models were estimated for the three dependent variables: word-recognition accuracy, word-recognition speed, and reading comprehension. Measurement point (four measurement points coded -1.5, -0.5, 0.5, and 1.5) was included as a fixed and random effect. Cohort was included as a dummy-coded predictor variable (with Cohort 1 serving as reference group). Furthermore, the interaction of measurement point with cohort was also included in the models to test for differences in growth rates between cohorts. Finally, we included in the models a dummy-coded variable for children's first language (other language than German as reference group) and a dummy-coded variable for gender (females as reference group) to control for individual differences. Significance tests were based on a Type I error probability of .05 (two-tailed).

The interaction term of measurement point with cohort was not significant in any of the models (word-recognition accuracy: $B = 0.15$, $SE = 0.35$, $p = .675$; word-recognition speed: $B = -0.27$, $SE = 0.39$, $p = .501$; reading comprehension: $B = 0.87$, $SE = 0.79$, $p = .273$), indicating that both cohorts developed similarly over time. Therefore, the two cohorts were combined for the following analyses.

Growth-curve Analyses for Word-recognition Speed and Reading Comprehension: Time Point of Reaching the Basic Accuracy Threshold as Moderator

To investigate whether the development of visual word-recognition speed and reading comprehension varies as a function of the time when children achieve the basic accuracy threshold, the 1,095 children were divided into five groups according to the 71%, 95% CI [64.75, 78.10] cut-off criterion which was identified by Karageorgos et al. (2019) in another sample of German fourth graders: (1) criterion achieved at the end of Grade 1, (2) criterion achieved at the end of Grade 2, (3) criterion achieved at the end of Grade 3, (4) criterion achieved at the end of Grade 4, and (5) criterion not achieved at the end of Grade 4. Descriptive statistics for the five groups can be found in Table 3.

Given the longitudinal design and the hierarchical structure of our data (repeated measures nested within students and students nested within classes), we estimated quadratic multilevel growth curve models for word-recognition speed and reading comprehension as dependent variables with the R-packages lme4 (Bates et al., 2019) and lmerTest with

Satterthwaite adjustments to denominator degrees of freedom (Kuznetsova et al., 2019). Please note that because of a technical issue, no information is available about the assignment of classes to schools for about half of the sample (i.e., for the schools located in the urban area of Frankfurt). Thus, we opted to include the random effect of classes only, accounting for clustering effects within classes, but not the random effect of schools. Nevertheless, we fitted an unconditional model for the other half of the sample, which included data from schools in the urban area of Kassel, to estimate the magnitude of potential clustering effects in schools. The ICCs for word-recognition speed ($\rho = .01$) and reading comprehension ($\rho = .05$) in the unconditional model for the data from this subset indicated little clustering effects due to schools. Similarly, low ICCs for word-recognition speed ($\rho = .02$) and reading comprehension ($\rho = .06$) in the unconditional model with the full data set indicated little clustering effects due to classes.

We therefore estimated two-level growth models with repeated measurement points nested within students but not within classes. Time (four measurement points coded as -1.5, -0.5, 0.5, and 1.5) and squared time were included as fixed and random effects. Four dummy-coded word-recognition accuracy variables were included as fixed effects in the model with children who reached the accuracy threshold by the end of Grade 1 (Group 1) serving as the reference category, children who reached the threshold by the end of Grade 2 (Group 2), children who reached the threshold by the end of Grade 3 (Group 3), children who reached the threshold by the end of Grade 4 (Group 4), and children who did not reach the threshold by the end of Grade 4 (Group 5). Interaction terms of the four word-recognition accuracy variables with time and squared time were included in the model. Finally, we included in the models a dummy-coded variable for first language (other language than German as reference group) and a dummy-coded variable for gender (females as reference group) to control for individual differences in these variables. Outlier analysis (Baayen, 2008, Chap. 7) identified 0.8% of data points in Model 1 with word-recognition speed as dependent variable and 0.3%

of the data points in Model 2 with reading comprehension as dependent variable that deviated more than 2.5 standard deviations from the mean of the residuals and were thus excluded from the analysis. Significance tests were based on a Type-I error probability of .05 (one-tailed, because all hypotheses were directed).

- INSERT TABLE 3 ABOUT HERE –

Effects on Word-recognition Speed. The results for word-recognition speed as dependent variable are shown in Table 4 (left column). The model showed a significant main effect of time on word-recognition speed, $B = 10.59$, $t(756.93) = 43.99$, $p < 0.001$, and significant main effects for the comparison of the reference group (children who reached the threshold by the end of Grade 1) with children that reached the threshold at the end of Grade 2, $B = -3.91$, $t(851.58) = -5.01$, $p < 0.001$, at the end of Grade 3, $B = -7.73$, $t(863.83) = -6.59$, $p < 0.001$, at the end of Grade 4, $B = -9.54$, $t(1011.70) = -5.11$, $p < 0.001$ and with children who did not reach the threshold by the end of Grade 4, $B = -9.22$, $t(1019.35) = -3.52$, $p < 0.001$, while keeping all other predictors in the model constant. However, no significant main effect of squared time on word-recognition speed was found, $B = -0.11$, $SE = 0.19$, $p = 0.56$. Moreover, time and group variables were involved in significant two-way interactions. All interaction effects of time and squared time with the dummy-coded group variables were significant (except for the interaction of time with Group 2), indicating earlier and more pronounced improvement of word-recognition speed for Group 1 children compared to children from Group 3, Group 4 and Group 5. Accordingly, the simple slopes of time decreased monotonically from Group 1 to Group 5 (Table 5, left columns). Furthermore, multiple Bonferroni corrected pairwise comparisons for the final measurement point with the R-package *emmeans* (Lenth, Singmann, Love, Buerkner, & Herve, 2019) showed that Group 1 ($M = 48.40$, $SE = 0.75$) achieved higher scores at the end of Grade 4 than Group 2 ($M = 45.50$, $SE = 0.81$), Group 3 ($M = 42.10$, $SE = 1.41$), Group 4 ($M = 37.40$, $SE = 2.13$) and Group 5 ($M = 37.10$, $SE = 3.71$). Finally, Group 2 also achieved higher scores than Group 4 (Table 6).

In sum, children's word-recognition speed improved over time, but the improvement was more pronounced for children who reached the threshold by the end of Grade 1 compared to children who reached the threshold after Grade 3 and children who had not reached the threshold by the end of Grade 4 (Figure 1). Moreover, children that reached the threshold by the end of Grade 1 recognized words faster at the end of Grade 4 than children that reached the threshold at later grades and children who had not reached the threshold.

- INSERT FIGURE 1 ABOUT HERE –

- INSERT TABLE 4 ABOUT HERE –

Effects on Reading Comprehension. The results for reading comprehension as the dependent variable are shown in Table 4 (right column). The model revealed significant main effects of time, $B = 15.50$, $t(788.91) = 27.04$, $p < 0.001$, and squared time, $B = -3.69$, $t(747.10) = -4.60$, $p < .001$, on reading comprehension and significant main effects for the comparison of Group 1 children with children from the other four groups, while keeping all other predictors in the model constant. Moreover, all predictor variables were involved in significant two-way interactions (except for the interactions of time and squared time with Group 2 and squared time with Group 3 and Group 5), indicating earlier and more pronounced improvement of reading comprehension for Group 1 children compared to children from Groups 3 to 5. Accordingly, the simple slopes of time estimated within each group decreased monotonically from Group 1 to Group 5 (Table 5, right-hand columns). Again multiple Bonferroni-corrected pairwise comparisons at the end of Grade 4 showed that Group 1 ($M = 73.00$, $SE = 1.25$) achieved higher scores than Group 2 ($M = 66.00$, $SE = 1.34$), Group 3 ($M = 52.90$, $SE = 2.31$), Group 4 ($M = 49.50$, $SE = 3.61$), and Group 5 ($M = 38.70$, $SE = 5.29$). Similarly, Group 2 also performed better than Groups 3-5 at the end of Grade 4 (Table 6).

In sum, these findings suggest an overall improvement of word-recognition speed and reading comprehension over time, both of which start earlier and are more pronounced for children who reach the basic accuracy threshold already in Grade 1 compared to children who reach the threshold in later grades and children who fail to reach the threshold by the end of Grade 4 (Figure 2). Furthermore, differences in reading comprehension between children who reach the threshold by the end of Grade 1 or by the end of Grade 2 and children who reach the threshold at later grades apparently remain stable by the end of primary school. Finally, reading comprehension for children who reach the threshold at later grades seems to improve more linearly across the primary school years compared to word-recognition speed.

- INSERT TABLE 5 ABOUT HERE –

- INSERT FIGURE 2 ABOUT HERE –

Discussion

The aim of the present study was to investigate the relationship between the development of visual word-recognition accuracy, word-recognition speed, and reading comprehension across the primary school years. The study tested the hypothesis that achieving a basic accuracy threshold early is a critical precondition for improvement in word-recognition speed and reading comprehension. To that aim, a longitudinal study was conducted with primary school children who were followed from Grade 1 to Grade 4.

In line with our expectations, the improvement of word-recognition speed across primary school years was more pronounced for children who reached the basic accuracy threshold (at least 71% of the written words in a lexical decision task were recognized correctly) by the end of Grade 1 compared to children who reached this threshold after Grade 3 and children who failed to reach this threshold by the end of Grade 4. These findings are in line with the theory of reading automaticity proposed by LaBerge and Samuels (1974). German beginning readers can recognize written words with high accuracy already after a short period of formal instruction (about one year, Landerl & Wimmer, 2008). However, they are still behind skilled readers in terms of word-recognition speed because of the inefficiency of their word recognition processes. To achieve reading automaticity, they need to be able to recognize words as a single unit. The findings are also in line with Juul et al. (2014) who suggest that the growth curve of word-recognition speed varies as a function of the time needed to reach the basic accuracy threshold. Hence, children who reached the basic accuracy threshold at Grade 1 apparently had more time at their disposal to develop their word-recognition speed and reach reading automaticity compared to children below the threshold. This interpretation is also substantiated by the multiple pairwise comparisons at the end of Grade 4. None of the groups that reached the threshold after Grade 3 and the group that failed to reach the threshold were able to catch up with the group that reached the threshold at Grade 1. This group probably reached automatized reading. The differential trajectories of word-recognition speed led to considerable differences at the end of primary school. However, there was no significant difference in the rate of change between Group 1 and Group 2. This could be due to the lack of more measurement points. Children were tested at the end of each Grade. Therefore, we only have a rough estimation of the time point at which children achieved the basic accuracy. It could be possible that the month interval between children achieving the basic accuracy threshold by the end of Grade 1 and children achieving the basic accuracy threshold by the end of Grade 2 was too short. As a result, these children developed similarly throughout the study. Despite its plausibility, this assumption should be investigated in further studies with more measurement points. Nevertheless, Group 2 was not able to catch up to Group 1 and their differences remained stable until the end of Grade 4.

The results also provide evidence in support of our second hypothesis regarding the development of reading comprehension. Reading comprehension of children who reached the basic accuracy threshold by the end of Grade 1 developed significantly better than reading comprehension of children who reached this threshold by the end of Grade 3, by the end of Grade 4, and children who failed to reach the threshold by the end of Grade 4. This finding is consistent with the lexical quality hypothesis (Perfetti, 1985; Perfetti & Hart, 2002). A likely explanation is that word-recognition speed improved quickly for those children who showed high word-recognition accuracy by the end of Grade 1. Thus, word recognition was highly efficient (reliable and fast) for those children, which saved cognitive resources for higher-level cognitive processes of reading comprehension. However, we found no differences in reading comprehension between children who reached the accuracy threshold by the end of Grade 1 and children who reached the threshold by the end of Grade 2. This result could be due to our study method. The first measurement point of reading comprehension corresponded to the end of Grade 2. By the end of Grade 2, both groups had already reached the basic accuracy threshold and started making gains on word-recognition speed. Presumably, if we had measured reading comprehension scores at the end of Grade 1 and included them in our model, this developmental pattern would probably be analogous to the comparisons with the other groups, and most likely these two groups would not have developed so similarly. Despite its plausibility, this assumption still needs empirical support by future studies. Nevertheless, even though the two groups developed similarly, group differences remained stable by the end of Grade 4. Furthermore, none of the other groups were able to catch up with the groups that reached the threshold at the end of Grade 1 or Grade 2. This result is also in line with the lexical quality hypothesis of Perfetti & Hart (2002). The two groups that crossed the basic accuracy threshold early had more cognitive resources at their disposal for engaging in and augmenting the higher-level processes necessary to achieve good reading comprehension.

In sum, the results of the present study strongly suggest that German primary school children’s word-recognition speed only begins to improve after they reach a certain basic word-recognition accuracy threshold. Furthermore, the time needed to achieve this threshold also seems to be important. The sooner the children reach the basic accuracy threshold, the larger the gains for the development of word-recognition speed and reading comprehension.

Limitations and Suggestions for Future Research

The results of the present study need to be interpreted with its limitations in mind. First, given the lack of a thumb rule for the calculation of the threshold in a longitudinal design, we chose to use the basic accuracy threshold found by Karageorgos et al. (2019) instead of calculating it from our sample data. To calculate the threshold in our longitudinal design, we would have had to run four curve-fitting models for each Grade with words-per-minute scores as dependent variables and polynomials of the accuracy scores as predictors in order to determine the turning point for each model. This would have resulted in four similar but not exactly equal thresholds. Consequently, there would have been different criteria for group allocation in each grade, and therefore group comparisons would not have been possible.

Another limitation of our study is that reading comprehension was not assessed at the end of Grade 1. The absence of one measurement point in combination with the complexity of our models led to convergence difficulties for the model with reading comprehension as dependent variable when squared time was included as a random effect. Thus, we cannot be sure whether the inclusion of squared time as a random effect in the model would have altered the effects.

Furthermore, even though our results are in line with the results of Juul et al. (2014), there are some fundamental differences between the methods used in the two studies that have to be mentioned. First, Juul et al. used a reading-aloud task to assess word reading accuracy and word reading speed, whereas we used a lexical decision task. Although both tasks assess word reading processes, reading aloud (or word naming) emphasizes word recognition via phonological recoding whereas a lexical decision task emphasizes on word recognition via orthographical representations. Second, the study of Juul et al. investigated the development of word reading from the end of kindergarten to the end of Grade 2 by testing Danish children eleven times. In our study, we investigated the development from the end of Grade 1 to the end of Grade 4 by testing German children four times. Thus, Juul et al. were able to predict an approximate number of days that it took a child to achieve the basic accuracy threshold in the first two Grades, whereas our study did not allow for such a precise estimate due to the larger spacing of the measurement points.

Despite these differences, the potential relevance of a basic accuracy threshold for the development of word recognition speed has now been established in two languages with different orthographies, Danish (Juul et al., 2014) and German, suggesting that the basic accuracy threshold is a general phenomenon that occurs in other alphabetic languages such as English as well. Of course, this conjecture needs to be backed up by future research, which ideally compares reading development in different languages. While the basic accuracy threshold might be a general phenomenon that might be found across alphabetic languages, it might still be the case that accuracy is more important in some languages than others and reading speed starts to develop earlier in some languages compared to others. For example, Lange-Küttner (2005) has shown that normally schooled German children (without preschool education) in Grade 1 read familiar words slower (and novel scrambled words faster) than British children, with no differences in accuracy. A potential explanation for this finding are different roles of phonology and word structure in learning to read in German vs. English (see also Lange-Küttner & Krappmann, 2011), which might also affect the exact position of the basic accuracy threshold in reading development.

Finally, although there are strong theoretical arguments supporting the possibility of a causal effect of achieving the basic accuracy threshold on word reading speed and reading comprehension, the results of our study are correlative and, therefore, do not provide unequivocal support for such an interpretation. Even though the basic accuracy level might be necessary for the development of word reading speed and reading comprehension, many other factors are important drivers of individual reading development, as well (e.g., Bowey, 2005).

How the basic accuracy threshold acts in concert with these other factors is a question for future research to address.

Conclusion

Our findings are in line with and extend the findings by Juul et al. (2014) and Karageorgos et al. (2019) regarding the relationship of word reading accuracy and fluency in reading development. The results suggest that the time needed to reach the basic accuracy threshold plays an important role in the development of children's word-recognition speed and reading comprehension in German. Furthermore, the differential trajectories of word-recognition speed and reading comprehension lead to considerable differences at the end of Grade 4. Therefore, identifying readers who are below the basic accuracy threshold at Grade 1 would be a judicious educational objective. Early identification of these children would allow for planning and implementing interventions that focus on fostering word-recognition accuracy. The hope—to be confirmed by future studies—is that such interventions will support and accelerate the development of fluent reading to improve comprehension.

Acknowledgements

Data collection was funded by the German Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF, grants 01GJ0985, 01GJ0986, 01GJ1206A, and 01GJ1206B). 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. Researchers who would like to inspect the items of the ProDi-L used in this study are invited to send an e-mail to the first author.

References

Acha, J., & Perea, M. (2008). The effects of length and transposed-letter similarity in lexical decision: Evidence with beginning, intermediate, and adult readers. *British Journal of Psychology, 99*, 245–264.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX lexical database* [CD-ROM 2nd Release]. Philadelphia. PA: Linguistic Data Consortium.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database* [CD-ROM]. Philadelphia. PA: Linguistic Data Consortium.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2019). Lme4: Linear mixed-effects models using 'Eigen' and S4. R package version 1.1-21 [software] Available from https://cran.r-project.org/web/packages/lme4/index.html

Booth, J. R., Perfetti, C. A., & MacWhinney, B. (1999). Quick, automatic, and general activation of orthographic and phonological representations in young readers. *Developmental Psychology, 35*, 3–19.

Bowey, J. A. (2005). Predicting individual differences in learning to read. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 155–172). Oxford, UK: Blackwell.

Bowey, J. A., & Muller, D. (2005). Phonological recoding and rapid orthographic learning in third graders’ silent reading: A critical test of the self-teaching hypothesis. *Journal of Experimental Child Psychology, 92*, 203–219.

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.

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

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.

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

Feingold, A. (2019). Time-varying effect sizes for quadratic growth models in multilevel and latent growth modelling. *Struct Equ Modeling, 26,* 418–429.

Frith, U. (1985). Beneath the surface of developmental dyslexia. In K. Patterson, J. Marshall, & M. Coltheart (Eds.), *Surface Dyslexia, Neuropsychological and Cognitive Studies of Phonological Reading* (pp. 301–330). London: Erlbaum.

Frith, U. (1986). A developmental framework for developmental dyslexia. *Annals of Dyslexia*, 36, 69–81.

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/web/packages/irr/index.html

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.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*, 6–10.

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

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal*, *2*, 127–160.

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

Karageorgos, P., Müller, B., & Richter, T. (2019). Modelling the relationship of accurate and fluent word recognition in primary school. *Learning and Individual Differences*, *76*.

Kendeou, P., van der Broek, P., Helder, A., & Karlsson, J. (2014). A cognitive view of reading comprehension: Implications for reading difficulties. *Learning Disabilities Research & Practice, 29*, 10–16.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, Cambridge University Press.

Klauda, S. L., & Guthrie, J. T. (2008). Relationships of three components of reading fluency to reading comprehension. *Journal of Educational Psychology, 100*, 310–321.

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.

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

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/web/packages/lmerTest/index.html.

Kyte, C. S., & Johnson, C. J. (2006). The role of phonological recoding in orthographic learning. *Journal of Experimental Child Psychology, 93*, 166–185.

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.

Lange-Küttner, C. (2005). Word structure effects in German and British reading beginners. *Zeitschrift für Pädagogische Psychologie, 19*, 207-218.

Lange-Küttner, C. & Krappmann, (2011). Ein modalitätsspezifisches Wortgedächtnis ist ausreichend: Wordgedächtnismodalität bei Leseanfängern und neuronalen Netzen [One modality-specific memory for words is sufficient: Word memory modality in beginning readers and neuronal networks]. *Zeitschrift für Pädagogische Psychologie, 25*, 19-37.

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2019). emmeans: Estimated marginal means, aka least-squares means. R package version 1.4.3.01 [software] Available from https://cran.r-project.org/web/packages/emmeans/index.html.

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.

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.

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

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.

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). The acquisition of reading comprehension skill. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227–247). Oxford, UK: Blackwell.

Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading, 18,* 22–37.

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

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.

Richter, T., Isberner, M.-B., Naumann, J., & Kutzner, Y. (2012). Prozessbezogene Diagnostik von Lesefähigkeit bei Grundschulkindern [Process-oriented assessment 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.

Schindler, J., & Richter, T. (2018). Reading comprehension: Individual differences, disorders, and underlying cognitive processes. In D. Ravid & A. Bar-On (Eds.), *Handbook of* *communication disorders: Theoretical, empirical, and applied linguistic perspectives* (pp. 503-524). Berlin, Germany: De Gruyter.

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. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition, 55*, 151–218.

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.

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

Spencer, L. H., & Hanley, J. R. (2003). Effects of orthographic transparency on reading and phoneme awareness in children learning to read in Wales. *British Journal of Psychology, 94*, 1–28.

Torgesen, J. K. (2000) Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research & Practice, 15*, 55–64.

Verhoeven, L., & Perfetti, C. (2008). Advances in text comprehension: Model, process and development. *Appl. Cognit. Psychol., 22*, 293–301.

Wimmer, H. (1996). The nonword reading deficit in developmental dyslexia: Evidence from children learning to read German. *Journal of Experimental Child Psychology*, *61*, 80–90.

Ziegler, J. C., Bertrand, D., Lété, B., & Grainger, J. (2014). Orthographic and phonological contributions to reading development: Tracking developmental trajectories using masked priming. *Developmental Psychology, 50*, 1026–1036.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, *131*, 3–29.

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.

Table 1

Characteristics of the Samples and Mean Values for Word-recognition Accuracy and Speed, and Reading Comprehension for all Measurement Points

					Word recognition		Reading comprehension
Measurement point	$n_{cohort\ 1}$	$n_{cohort\ 2}$	$n_{females}$	Age *M* (*SD*)	Accuracy *M* (*SD*)	Speed *M* (*SD*)	*M* (SD)
t1	426	602	474	7.44 (0.45)	66.31 (14.92)	17.26 (14.82)	
t2	426	556	476	8.40 (0.45)	78.02 (13.77)	25.01 (9.98)	49.89 (9.52)
t3	387	452	414	9.46 (0.44)	81.52 (15.10)	34.9 (12.98)	49.06 (10.51)
t4	383	469	413	10.46 (0.45)	84.22 (10.89)	46.05 (14.53)	49.39 (13.35)

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 (20-item ELFE 1-6, Lenhard & Schneider, 2006).

Table 2

Attrition Rates and Mean Values for Word-recognition Accuracy and Speed, and Reading Comprehension per Measurement Point for Dropped-out Children

	Timepoint of Attrition		
	Grade 2	Grade 3	Grade 4
Attrition Rates (%)	5.66	9.32	7.21
Word-recognition accuracy (*M* (*SD*))			
Grade 1	59.78 (18.65)	58.20 (16.48)	62.68 (16.76)
Grade 2	-	71.09 (17.43)	74.72 (15.36)
Grade 3	-	-	77.30 (17.07)
Word-recognition speed (*M* (*SD*))			
Grade 1	23.45 (26.52)	16.97 (18.48)	18.20 (15.78)
Grade 2	-	22.63 (9.57)	23.18 (9.35)
Grade 3	-	-	30.48 (10.90)
Reading comprehension (*M* (*SD*))			
Grade 2	-	28.55 (12.39)	35.40 (15.30)
Grade 3	-	-	47.81 (22.64)

Note. Word recognition = 16-item lexical decision task (subtest of ProDi-L, Richter et al., 2017). Reading comprehension = Percentage of correct responses in 20 text comprehension items (20-item ELFE 1-6, Lenhard & Schneider, 2006).

Table 3

Characteristics of the Samples and Mean Values for Word-recognition Accuracy and Speed, and Reading Comprehension by Group at the End of Each Grade

	Time point when threshold reached				
	By the end of Grade 1	By the end of Grade 2	By the end of Grade 3	By the end of Grade 4	Later than the end of Grade 4
Number of students (*n*)					
Grade 1	417	312	70	18	13
Grade 2	361	373	81	21	14
Grade 3	312	272	93	19	11
Grade 4	339	290	78	47	11
Word-recognition accuracy (*M* (*SD*))					
Grade 1	79.45 (5.27)	61.60 (8.16)	58.21 (9.84)	57.29 (10.34)	53.37 (8.31)
Grade 2	83.47 (9.82)	83.48 (6.68)	63.50 (6.05)	63.39 (6.93)	61.16 (5.01)
Grade 3	86.96 (9.38)	83.89 (10.75)	82.73 (7.22)	62.50 (5.51)	61.36 (6.74)
Grade 4	87.76 (7.64)	85.43 (7.50)	82.05 (8.57)	81.02 (5.61)	66.48 (4.21)
Word-recognition speed (*M*(*SD*))					
Grade 1	16.53 (6.51)	14.73 (6.18)	13.16 (6.30)	18.69 (17.50)	17.03 (13.26)
Grade 2	27.41 (9.38)	23.87 (8.17)	19.96 (7.66)	18.71 (6.54)	17.55 (7.56)
Grade 3	37.77 (12.39)	33.46 (10.68)	28.95 (11.24)	24.81 (8.64)	23.07 (8.74)
Grade 4	48.88 (14.05)	45.50 (12.94)	40.53 (14.85)	35.97 (12.59)	30.93 (11.86)
Reading comprehension (*M*(*SD*))					
Grade 2	44.41 (19.23)	35.85 (16.77)	29.32 (12.09)	23.42 (9.73)	27.65 (9.37)
Grade 3	63.89 (22.69)	54.34 (22.32)	40.86 (16.79)	33.57 (13.32)	32.37 (10.85)
Grade 4	76.30 (20.79)	68.30 (23.20)	51.52 (18.80)	45.21 (16.18)	41.76 (18.20)

Note. Word-recognition accuracy and speed based on 16-items from the ProDi-L lexical decision task (ProDi-L, Richter et al., 2017). Reading comprehension = Percentage of correct responses of 20 items from the ELFE text comprehension subtest (ELFE 1-6, Lenhard & Schneider, 2006).

Table 4

Fixed Effects and Variance Components for the Multilevel Non-linear Growth Curve Analyses with Word-recognition Speed and Reading Comprehension as Dependent Variables.

	Word-recognition speed	Reading comprehension
Parameter	*B (SE)*	*B (SE)*
Fixed effects		
Intercept	32.68 (0.69)**	61.03 (1.48)**
Time	10.59 (0.24)**	15.50 (0.57)**
Squared time	-0.11 (0.19)	-3.69 (0.80)**
Gender	0.94 (0.48)*	-4.33 (1.14)**
First language	-0.76 (0.52)	4.60 (1.22)**
Group 2	-3.91 (0.78)**	-7.83 (1.54)**
Group 3	-7.73 (1.17)**	-17.83 (2.22)**
Group 4	-9.54 (1.87)**	-23.92 (3.46)**
Group 5	-9.22 (2.62)**	-28.47 (4.66)**
Time by Group 2	-0.32 (0.36)	-0.10 (0.84)
Time by Group 3	-1.04 (0.53)*	-4.24 (1.22)**
Time by Group 4	-3.95 (0.77)**	-4.72 (1.94)*
Time by Group 5	-4.92 (1.30)**	-11.05 (2.65)**
Squared time by Group 2	0.66 (0.28)*	0.96 (1.17)
Squared time by Group 3	1.33 (0.42)*	1.97 (1.65)
Squared time by Group 4	1.99 (0.67)*	5.13 (2.82)*
Squared time by Group 5	2.36 (1.08)*	5.30 (3.58)
Variance components		
Time	12.20 (3.49)	40.60 (6.37)
Squared time	1.83 (1.35)	–
Children	82.10 (9.06)	257.30 (14.68)

Note. The four measurement points were coded as -1.5 = t1, -0.5 = t2, 0.5 = t3, 1.5 = t4. For reading comprehension as dependent variable, only three measurement points were coded (-1 = t2, 0 = t3, 1 = t4). Gender was coded as 0 = female, 1 = male. First language was coded as 0 = other language than German, 1 = German. Group 1 consisting of children who reached the threshold at the end of Grade 1 (reference group) were coded with 0, Group 2 (1 = threshold reached by the end of Grade 2), Group 3 (1 = threshold reached by the end of Grade 3), Group 4 (1 = threshold reached by the end of Grade 4), and Group 5 (1 = threshold not reached by the end of Grade 4).
* $p < 0.05$; ** $p < 0.01$

Table 5

Fixed Effects (Simple Slopes) and Variance Components for Multilevel Non-linear Growth Curve Analyses for each Group with Word-recognition Speed and Reading Comprehension as Dependent Variables.

	Word recognition speed					Reading comprehension				
	By the end of Grade 1	By the end of Grade 2	By the end of Grade 3	By the end of Grade 4	Later than the Grade 4	By the end of Grade 1	By the end of Grade 2	By the end of Grade 3	By the end of Grade 4	Later than the Grade 4
Parameter	*B (SE)*	*B (SE)*	*B (SE)*	*B (SE)*	*B (SE)*	*B (SE)*	*B (SE)*	*B (SE)*	*B (SE)*	*B (SE)*
Fixed Effects										
Intercept	31.53 (0.92)*	29.47 (0.73)*	26.52 (1.34)*	22.56 (3.11)*	18.29 (4.53)*	58.48 (2.11)*	54.06 (1.92)*	44.12 (2.46)*	42.14 (3.73)*	34.62 (4.00)*
Time	10.57 (0.23)*	10.24 (0.25)*	9.60 (0.58)*	6.59 (0.88)*	6.07 (2.35)*	15.56 (0.56)*	15.34 (0.65)*	11.24 (1.03)*	10.87 (1.78)*	4.61 (2.32)*
Squared Time	-0.12 (0.18)	0.52 (0.19)*	1.16 (0.32)*	1.98 (0.80)*	0.67 (0.77)	-3.65 (0.82)*	-2.80 (0.86)*	-1.81 (1.35)	2.28 (2.68)	1.28 (2.88)
Gender	2.68 (0.77)*	-0.91 (0.66)	0.09 (1.35)	-1.71 (3.27)	9.99 (5.06)*	-2.01 (1.86)	-7.15 (1.85)*	-1.79 (2.48)	-3.82 (3.77)	-4.65 (3.62)
First language	-0.37 (0.87)	-0.39 (0.70)	-3.02 (1.35)*	3.00 (3.21)	-0.06 (5.00)	6.52 (2.08)*	5.43 (1.95)*	0.19 (2.48)	-6.29 (3.75)*	0.85 (3.55)
Variance Components										
Time	9.28	9.59	24.99	13.23	91.95	33.01	49.06	41.43	34.68	47.05
Children	75.92	47.93	64.82	74.33	117.61	296.58	264.18	162.87	77.28	45.26
d	2.24	2.64	2.27	0.93	1.44	0.82	0.93	0.76	1.56	0.74

Note. The four measurement points were coded as -1.5 = t1, -0.5 = t2, 0.5 = t3, 1.5 = t4. For reading comprehension as dependent variable, only three measurement points were coded (-1 = t2, 0 = t3, 1 = t4). Gender was coded as 0 = female, 1 = male. First language was coded as 0 = other language than German, 1 = German. Group 1 consisting of children who reached the threshold at the end of Grade 1 (reference group) were coded with 0, Group 2 (1 =

threshold reached by the end of Grade 2), Group 3 (1 = threshold reached by the end of Grade 3), Group 4 (1 = threshold reached by the end of Grade 4), and Group 5 (1 = threshold not reached by the end of Grade 4). Cohen's *d* values reflect the standardized growth (from Grade 1 to Grade 4) for each of the five groups. For the estimation of the time-varying effect size *d* the coefficients of linear and quadratic slopes were multiplied by the span from baseline to the end of the study and then divided by the pooled standard deviation (Feingold, 2019).

* $p < 0.05$

Table 6

Adjusted Mean Differences of Multiple Pairwise Comparisons with Bonferroni Correction for Word-recognition Speed (Below Main Diagonal) and Reading Comprehension (Above Main Diagonal) at the End of Grade 4. Parentheses show Cohen's d values.

	Group 1	Group 2	Group 3	Group 4	Group 5
Group 1	–	6.98* (0.39)	20.10* (1.13)	23.51* (1.33)	34.23* (1.93)
Group 2	2.91* (0.24)	–	13.13* (0.74)	16.53* (0.93)	27.25* (1.54)
Group 3	6.29* (0.50)	3.38 (0.27)	–	3.40 (0.19)	14.13 (0.80)
Group 4	10.99* (0.89)	8.09* (0.65)	4.70 (0.38)	–	10.72 (0.61)
Group 5	11.29* (0.91)	8.38 (0.68)	4.99 (0.40)	0.29 (0.02)	–

Note. Group 1 = threshold reached by the end of Grade 1; Group 2 = threshold reached by the end of Grade 2; Group 3 = threshold reached by the end of Grade 3; Group 4 = threshold reached by the end of Grade 4; Group 5 = threshold not reached by the end of Grade 4. For Cohen's *d* values mean difference was divided by the pooled standard deviation at baseline.

* $p < 0.05$

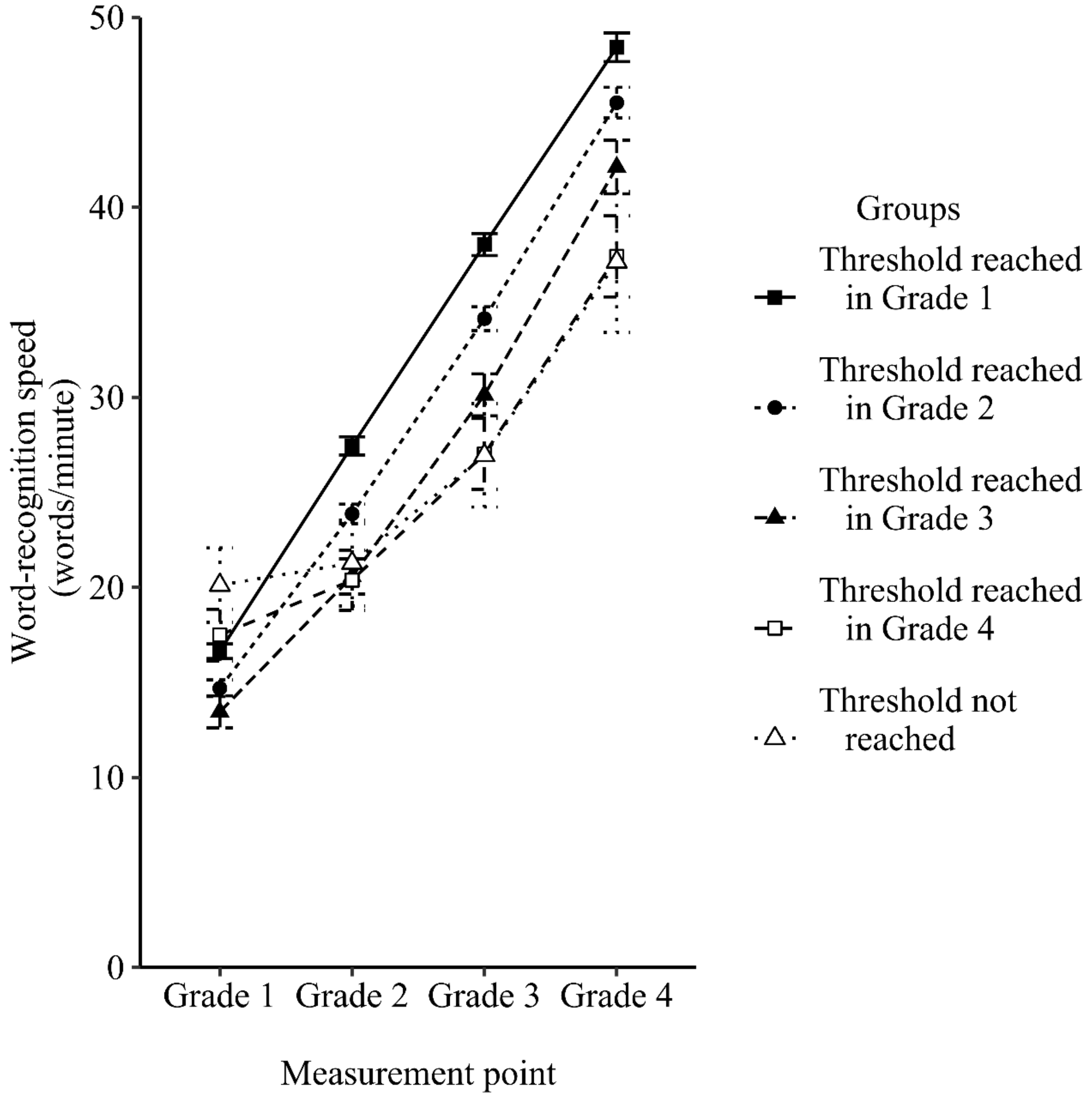

Figure 1. Average adjusted growth trajectories of word-recognition speed with standard errors depending on the time point when the basic accuracy threshold was reached.

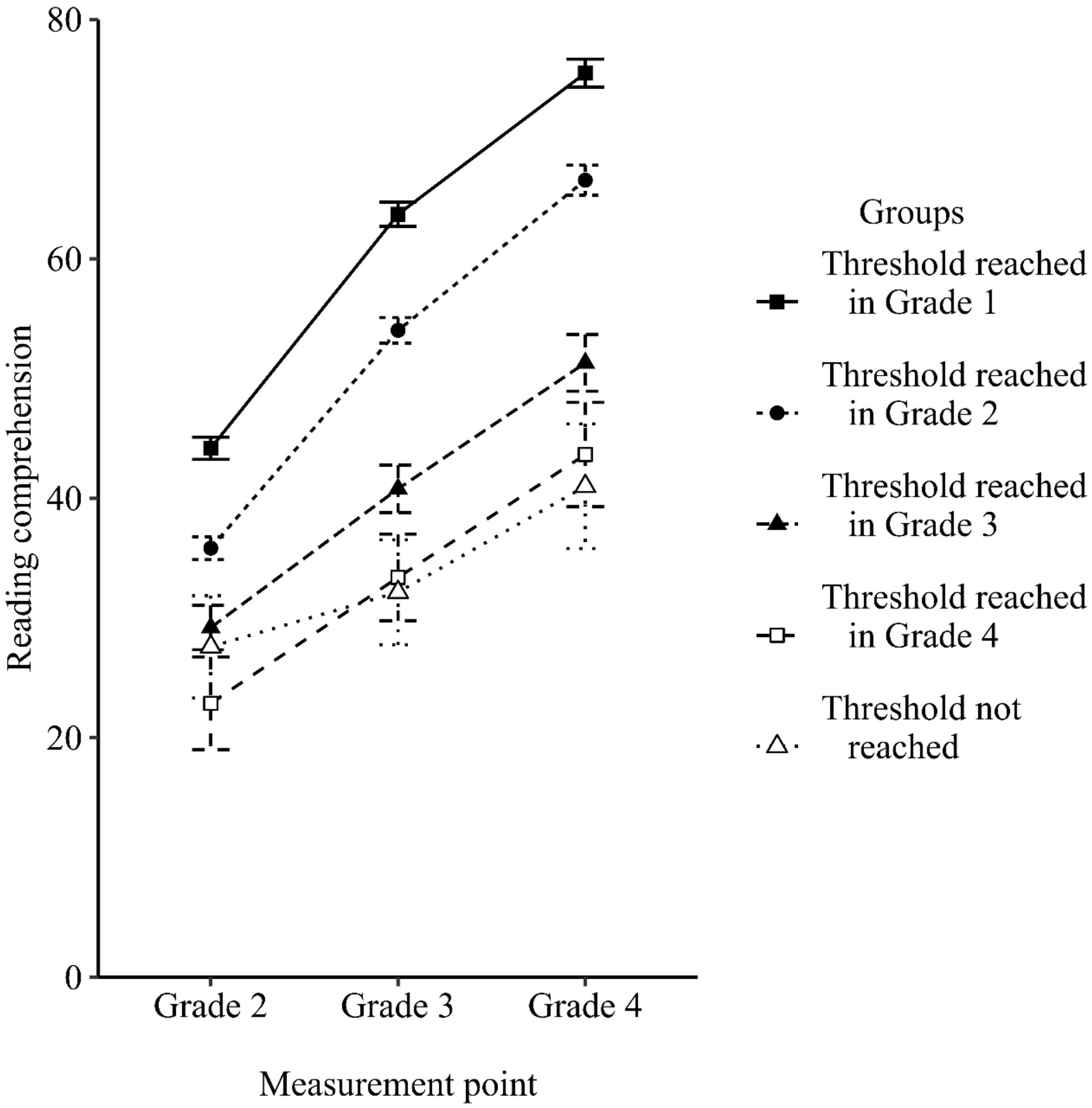

Figure 2. Average adjusted growth trajectories of reading comprehension with standard errors depending on the time point when the basic accuracy threshold was reached.