Investigating Reading Fluency in German Primary School Children: Interplay of Word Reading Accuracy, Speed, and Prosody

Inaugural-Dissertation

zur Erlangung der Doktorwürde der

Fakultät für Humanwissenschaften

der

Julius-Maximilians-Universität Würzburg

Vorgelegt von

Panagiotis Karageorgos

aus Kastoria (Griechenland)

Würzburg

2022

Betreuer: Prof. Dr. Tobias Richter

Erstgutachter: Prof. Dr. Wolfgang Lenhard

Zweitgutachter: PD Dr. Peter Marx

Tag der Disputation: 11. November 2022

Zusammenfassung

Lesefähigkeiten zählen zu den wichtigsten Grundfertigkeiten in der heutigen Gesellschaft. Jedoch gelingt es nicht allen Lesern und Leserinnen Texte angemessen zu verstehen oder einzelne Wörter zu dekodieren. Erkenntnisse der Internationale Grundschul-Lese-Untersuchung (PIRLS; Deutsch: IGLU) belegen, dass etwa ein Fünftel der Viertklässler und Viertklässlerinnen Kohärenz nur auf lokaler Ebene herstellen können und in manchen Fällen nur über ein rudimentäres Verständnis des gelesenen Textes verfügen (Bremerich-Vos et al., 2017). Außerdem bleiben diese Lesedefizite bestehen und haben einen negativen Einfluss auf den schulischen und beruflichen Erfolg (Jimerson, 1999). Deshalb ist es wichtig die Ursachen dieser Defizite zu identifizieren und frühzeitig Möglichkeiten für Interventionen zu schaffen.

Ziel der vorliegenden Dissertation war es, den Zusammenhang zwischen den Teilaspekten der Leseflüssigkeit und deren Einfluss auf das Leseverständnis näher zu untersuchen. Trotz der zunehmenden wissenschaftlichen Beschäftigung mit der Leseflüssigkeit in den letzten Jahren, besteht hier eine Forschungslücke sowohl bezüglich des Zusammenhangs zwischen Worterkennungsgenauigkeit und -geschwindigkeit als auch bezüglich der Relevanz prosodischer Muster für das Leseverständnis.

In Studie 1 wurde untersucht, ob deutsche Viertklässler und Viertklässlerinnen (N = 826) eine bestimmte Worterkennungsgenauigkeitsschwelle erreichen müssen, bevor sich ihre Worterkennungsgeschwindigkeit verbessert. Darüber hinaus wurde in einer Teilstichprobe (n = 170) mit einem Prä-/Posttest Design untersucht, inwiefern die bestehende Worterkennungsgenauigkeit die Effekte einer silbenbasierten Leseintervention auf die Worterkennungsgenauigkeit und Worterkennungsgeschwindigkeit beeinflussen kann. Die Ergebnisse zeigten, dass die Worterkennungsgeschwindigkeit sich verbesserte, nachdem die Kinder eine Worterkennungsgenauigkeit von 71% erreichten. Zudem zeigten sich ein

positiver Interventionseffekt auf die Worterkennungsgenauigkeit für die Kinder, die vor der Intervention unter der 71% Schwelle lagen. Auf die Worterkennungsgeschwindigkeit ergab sich für alle Kinder ein positiver Interventionseffekt. Auf das Leseverständnis hingegen zeigte sich nur für die Kinder, die vor der Intervention über der 71% Schwelle lagen, ein positiver Effekt.

Studie 2 untersuchte den in der ersten Studie aufgezeigten Zusammenhang zwischen Worterkennungsgenauigkeitsschwelle und Worterkennungsgeschwindigkeit in einem längsschnittlichen Design mit deutschen Schülerinnen und Schüler (N = 1,095). Die Worterkennungsgenauigkeit und -geschwindigkeit wurden von Ende der Jahrgangsstufe 1 bis Jahrgangsstufe 4 erfasst, während das Leseverständnis von Ende der Jahrgangsstufe 2 bis Jahrgangsstufe 4 erfasst wurde. Die Ergebnisse zeigten, dass die Entwicklungsverläufe der Worterkennungsgeschwindigkeit und des Leseverständnisses bei Kindern, die die Worterkennungsgenauigkeitsschwelle bis Ende der Jahrgangsstufe 1 erreichten, steiler waren als bei Kindern, die diese Schwelle erst später oder gar nicht erreichten.

In Studie 3 wurden mit Hilfe der Rekurrenzanalyse (RQA) prosodische Muster aus der Leseaufnahmen leseschwacher und -starker Kinder der Jahrgangsstufe 2 (n = 67) und der Jahrgangsstufe 4 (n = 69) extrahiert und für die Klassifikation in leseschwach und lesestark verwendet. Darüber hinaus wurde die Klassifikation anhand der prosodischen Muster aus der Rekurrenzanalyse mit der Klassifikation prosodischer Merkmale aus der manuellen Transkription der Leseaufnahmen verglichen. Die Ergebnisse zeigten, dass leseschwache Kinder der Jahrgangsstufe 2 längere Pausen innerhalb oder zwischen einzelnen Wörtern machen und im Durchschnitt mehr Zeit zwischen Pausen verstreichen lassen, während leseschwache Kinder der Jahrgangsstufe 4 mehr Zeit zwischen wiederkehrende Betonungen verstreichen lassen, viele Muster bei der Tonhöhe aufweisen und häufiger wiederkehrende Betonungen zeigen. Obwohl die Rekurrenzanalyse eine gute Anpassungsgüte hatte und zusätzliche Informationen zu dem Zusammenhang der Prosodie mit dem Leseverständnis lieferte, ergab sich für das Model mit den prosodischen Merkmalen aus der Transkription eine bessere Anpassungsgüte.

Zusammenfassend liefern die drei Studien der vorliegenden Dissertation vier bedeutsame Erkenntnisse bezüglich der Leseflüssigkeit im Deutschen. Erstens: Es gibt eine gewisse Schwelle bei der Worterkennungsgenauigkeit, die erreicht werden muss, bevor sich die Worterkennungsgeschwindigkeit verbessert. Zweitens: Je früher diese Genauigkeitsschwelle erreicht wird, desto stärker ist der Zuwachs in der Worterkennungsgeschwindigkeit und im Leseverständnis. Drittens: Die Interventionseffekte einer Leseintervention in die Grundschule werden von der Genauigkeitsschwelle beeinflusst. Viertens: Während inkorrekte Pausen innerhalb oder zwischen einzelnen Wörtern eine wichtige Rolle für die Identifikation und Beschreibung von leseschwachen Kindern in die Jahrgangsstufe 2 spielen, nimmt die Bedeutung der prosodischen Muster in der Jahrgangsstufe 4 zu.

Abstract

Reading skills are among the most important basic skills in society. However, not all readers are able to adequately understand texts or decode individual words. Findings from the Progress in International Reading Literacy Study (PIRLS; German: IGLU) show that about one fifth of fourth graders can only establish coherence at the local level, and in some cases they only have a rudimentary understanding of the text they read (Bremerich-Vos et al., 2017). In addition, these reading deficits persist and have a negative impact on academic and professional success (Jimerson, 1999). Therefore, identifying the causes of these deficits and creating opportunities for interventions at an early stage is an important research objective.

The aim of this dissertation was to examine the relationship between the aspects of reading fluency and their influence on reading comprehension. Despite the increasing scientific interest in reading fluency in recent years, a research gap still exists in the relationship between word recognition accuracy and both speed and the relevance of prosodic patterns for reading comprehension.

Study 1 investigated whether German fourth graders (N = 826) were required to reach a certain word-recognition accuracy threshold before their word-recognition speed improved. In addition, a sub-sample (n = 170) with a pre-/posttest design was examined to assess the extent that the existing word-recognition accuracy can influence the effects of a syllable-based reading intervention on word-recognition accuracy and word-recognition speed. Results showed that word-recognition speed improved after children achieved a word-recognition accuracy of 71%. A positive intervention effect was also found on word-recognition accuracy for children who were below the 71% threshold before the intervention, whereas the intervention effect on word-recognition speed was positive for all children. However, a positive effect on reading comprehension was only found for children who were above the 71% threshold before the intervention.

Study 2 investigated the relationship between word-recognition accuracy threshold and word-recognition speed shown in the first study in a longitudinal design with German students (*N* = 1,095). Word-recognition accuracy and speed were assessed from the end of Grade 1 to 4, whereas reading comprehension was assessed from the end of Grade 2 to 4. The results showed that the developmental trajectories of word recognition speed and reading comprehension were steeper in children who reached the word-recognition accuracy threshold by the end of the first grade than in children who later reached or had not reached this threshold.

In Study 3, recurrence analysis (RQA) was used to extract prosodic patterns from reading recordings of struggling and skilled readers in the second (*n* = 67) and fourth grade (*n* = 69) and was used for the classification into struggling and skilled readers. In addition, the classification based on the prosodic patterns from the recurrence quantification analysis was compared with the classification of prosodic features from the manual transcription of the reading recordings. The results showed that second-grade struggling readers have lengthier pauses within or between words and take more time between pauses on average, whereas fourth-grade struggling readers spend more time between recurring stresses and have multiple diverse patterns in pitch and more recurring accents. Although the recurrence analysis had a good goodness of fit and provided additional information about the relationship of prosody with reading comprehension, the model using prosodic features from transcription had a better fit.

In summary, the three studies in this dissertation provide four important insights into reading fluency in German. First, a threshold in word-recognition accuracy must be achieved before word-recognition speed improves. Second, the earlier this accuracy level is reached, the greater the gain in word-recognition speed and reading comprehension. Third, the intervention effects of a primary school reading intervention are influenced by the accuracy level. Fourth, although incorrect pauses within or between words play an important role in identifying and describing struggling readers in second grade, the importance of prosodic patterns increases in fourth grade.

Table of Contents

1 Introduction

Reading is a vital aspect of daily living. The Organisation for Economic Co-operation and Development (OECD, 2019) defines reading literacy as “an individual’s capacity to understand, use, evaluate, reflect on, and engage with texts in order to achieve one’s goals, develop one’s knowledge and potential, and participate in society” (p. 14). People are confronted daily with situations in which they must read information, comprehend it, and sometimes make quick decisions based on this information. However, the amount of time and effort that readers must devote to comprehending a text varies depending on their reading ability and the context. For example, reading, comprehending, and memorizing a large amount of complex information, such as the current dissertation, normally requires more time and effort than reading the menu at a restaurant or sending a short text message on a smartphone.

Learning to read is undeniably a continuous process that begins at a young age and develops with time and effort over many years. Readers must first be able to recognize words efficiently, comprehend sentences, and construct coherence between sentences before they can comprehend a text (e.g., Castles et al., 2018; Kintsch, 1998; Perfetti & Hart, 2002; Vellutino et al., 2004). If word recognition is inefficient, that is, slow and inaccurate, readers will most likely struggle, reading the text word for word, with long pauses between each word (Godde et al., 2022). This inefficiency puts a strain on readers’ limited cognitive capacity and consequently leaves them with less cognitive resources for higher-order cognitive processes, such as reading comprehension (e.g., Castles et al., 2018; LaBerge & Samuels, 1974; Nouwens et al., 2021; Perfetti, 1985; Perfetti & Hart, 2002). Hence, fluent and efficient word recognition are important prerequisites for successful reading comprehension (e.g., Adlof et al., 2006; Hoover & Gough, 1990; Ober et al., 2020; Pikulski & Chard, 2005).

According to the Progress in International Literacy Study (PIRLS; German: IGLU), which documents and provides internationally comparative data on reading competence of fourth graders every five years, in 2016, 18.9% of fourth graders learning to read in German failed to grasp elementary reading comprehension tasks (Bremerich-Vos et al., 2017). This statistic is rather concerning compared to 15.4% in 2011 and 13.3% in 2006 (Bremerich-Vos et al., 2017). In addition, only every third child of this subsample took part in an intervention program, which indicates a large discrepancy between the need for interventions and the support that these children receive (Bos et al., 2017).

A similar tendency can be found when examining data from adult German readers. According to the 2018 Level-One (LEO) survey, which documents reading and writing skills among adults, 12.1% of German-speaking adults between the ages of 18 to 64 years can only read and write up to the level of simple sentences (Grotlüschen et al., 2020). Despite a 2.4% decline from the 2010 LEO survey, this proportion still translates to 6.2 million adults in the German population (Grotlüschen et al., 2020). Surprisingly, 76% of the low-literate adults had some type of school-leaving qualification, with 40.6% even having a lower secondary school certificate or an equivalent degree (Grotlüschen et al., 2020). However, 12.9% of this subsample was unemployed, although the unemployment rate for the overall population of Germany in 2018 was 5.2% (Bundesagentur für Arbeit, n.d.;Grotlüschen et al., 2020). Hence, the literacy issues appear to remain stable over time and have a detrimental influence not just on academic accomplishment but also on professional careers (e.g., Jimerson, 1999). This underscores the need for identifying the reasons behind literacy issues as early as possible, with a focus on whether students can read fluently and, if possible, on interventions to improve fundamental reading abilities (Rasinski et al., 2011; Tarelli et al., 2012).

Reading fluency is a complex, multifaceted construct described by word-recognition accuracy, automaticity, and prosody (e.g., Hudson et al., 2008; Kuhn et al., 2010; Rasinski, 2004; Samuels, 2007). Even though the first two components, word-recognition accuracy and automaticity, are related to word recognition, they are distinct concepts (Yildiz et al., 2014). Word-recognition accuracy refers to correct word identification with minimal errors (Hudson et al., 2005; Rasinski, 2004), whereas automaticity reflects effortless word recognition (LaBerge & Samuels, 1974; Paige et al., 2014). Hence, word-recognition accuracy refers to a reader's ability to accurately recognize words regardless of the amount of cognitive resources necessary, whereas automaticity refers to a reader's ability to recognize words efficiently with the least amount of cognitive resources required (Rasinski, 2004; Yildiz et al., 2014). Nonetheless, automaticity is largely dependent on adequate word-recognition accuracy (Altani et al., 2020; Juul et al., 2014). According to a study of Danish children by Juul et al. (2014; see also Altani et al., 2020), children must first achieve a fundamental level of word-recognition accuracy (i.e., the basic accuracy level) before their word-recognition speed can improve. Furthermore, the findings of Juul et al. also indicate that the age at which children reach this milestone in their reading development (i.e., the basic accuracy achievement time) is a strong predictor of their growth of word-recognition speed.

The third component of reading fluency, prosody, refers to the appropriate expression during oral reading (Rasinski & Hoffman, 2003; Yildiz et al., 2014). Numerous studies have reported a relationship between reading fluency and reading comprehension (e.g., Aaron et al., 1999; Benjamin & Schwanenflugel, 2010; Rasinski et al., 2009; Shinn et al., 1992). Nevertheless, most studies focus on word-recognition accuracy and automaticity while neglecting prosody (e.g., Adlof et al., 2006; Fuchs et al., 2001; Kara et al., 2020; Silverman et al., 2013; Wolters et al., 2022). This neglect might be due to the more complex and time-consuming nature of collecting and analyzing prosody data compared to measuring and analyzing data on word reading accuracy and word reading speed (Dowhower, 1991).

The primary aim of this dissertation was to advance the understanding of the components reading-fluency components in the German language. In particular, this dissertation focused on investigating the basic accuracy level required for the development of word-recognition speed (Study 1), the impact of the basic accuracy achievement time on the development of word-recognition speed (Study 2), and prosodic patterns that may differentiate between struggling and skilled readers (Study 3).

The first study is a published paper that investigated whether the basic accuracy level required for the development of word-recognition speed proposed by Juul et al. (2014), which had to date only been established in the Danish language, can be found in German fourth graders. Furthermore, it was examined if a syllable-based reading intervention had different effects depending on whether participants had reached this level prior to the intervention.

Study 2 is also a published paper that investigated whether the basic accuracy achievement time reported in the study of Juul et al. (2014), that is, the time when children achieve the basic accuracy level established in Study 1, affects word-recognition speed and reading comprehension in German. For that reason, longitudinal data of German primary school children from Grades 1 to 4 were examined.

Study 3 refers to a currently submitted manuscript that explores prosodic patterns of skilled and struggling German readers in Grades 2 and 4. The primary goal of this study was to investigate separately for each grade if variations in the prosodic patterns between these two groups could be found. Furthermore, another purpose of this study was to compare the effectiveness of classifying struggling and skilled readers using prosodic patterns obtained with RQA to the effectiveness using prosodic features obtained by prosodic transcription.

In the current thesis, Chapter 2 provides a review of the theoretical and empirical foundations of reading fluency and its components, as well as an outline of reading interventions. The research questions for the three studies are illustrated in Chapter 3, and the empirical studies conducted to address these research questions are presented in Chapters 4, 5, and 6. Finally, in Chapter 7, the three studies' findings, limitations, and implications are reviewed, and future research directions are discussed.

2 Theoretical and Empirical Background

This chapter summarizes the theoretical and empirical background of the three studies included in this doctoral dissertation. The first section focuses on reading fluency and on its relationship with reading comprehension and provides an overview of the three components, that is, word reading accuracy, automaticity, and prosody. Furthermore, the first section briefly summarizes current prosody research methods. The second section focuses on reading interventions and their effects on reading comprehension.

2.1 Reading Fluency

Although the concept of reading fluency has been around for years (e.g., Clay, 1979; Clay & Imlach, 1971; Resnick, 1970), and researchers such as Allington (1983) have attempted to shift focus towards the concept, it had not received considerable attention until after the Report of the National Reading Panel (2000) was published. The National Reading Panel defined fluency as "the ability to read a text quickly, accurately, and with proper expression" (p. 3-5). To make sense of what is read, readers must be able to recognize words accurately and effortlessly and read them with the appropriate prosody (Rasinski, 2006). This statement implies that although accurate and efficient word recognition is necessary for reading fluency, these requirements are not sufficient.

Reading prosody is also an important aspect of reading fluency that reflects the segmentation of a text into meaningful units and is even considered by some researchers as the link between reading fluency and reading comprehension (Dowhower, 1991; Kuhn et al., 2010; Kuhn & Stahl, 2003). Nevertheless, reading fluency is frequently used interchangeably with word reading automaticity while overlooking the third component, prosody (e.g., Adlof et al., 2006; Fuchs et al., 2001; Kim, 2015; Rasinski et al., 2009; Silverman et al., 2013). One possible explanation for this omission could be that assessing word reading accuracy and speed is more straightforward than collecting and evaluating prosody data, which requires more effort (Dowhower, 1991; Schwanenflugel et al., 2004). Word-recognition accuracy and speed can be examined in larger settings during silent reading, whereas prosody must be evaluated individually during oral reading. In this dissertation, the term "reading fluency" will be used to refer to all three components.

2.1.1 Development of Visual Word Recognition

As emphasized in the previous sections, efficient and reliable word recognition is fundamental to reading fluency and reading comprehension (Snowling & Hulme, 2005). Reading research provides us with a plethora of models that attempt to describe the processes and the development of visual word recognition (e.g., Castles et al., 2018; Eysenck & Keane, 2020; Norris, 2013). In this context, the dual route cascaded model (DRC) established by Coltheart et al. (2001) is one of the most widely known and used examples. The DRC model supports the idea that word recognition may be accomplished via two routes: the non-lexical route, which is accomplished through phonological recoding, and the lexical route, which is accomplished through orthographical decoding (see also Eysenck & Keane, 2020; Martens & de Jong, 2006; Perfetti & Stafura, 2014; Zoccolotti et al., 2005).

During phonological recoding, readers utilize grapheme-phoneme correspondence rules to successively convert each letter into a phoneme and subsequently compose a word (Coltheart et al., 2001). Needless to say, understanding the alphabetic principle, that is, the concept that letters represent sounds of spoken language, is crucial for phonological recoding

(Van Orden & Kloos, 2005). However, using phonological recoding to recognize words has limitations. Reading a word letter-by-letter may be taxing and time-consuming, requiring additional cognitive resources to compensate for the lack of automaticity, which eventually impairs reading fluency and reading comprehension (Perfetti, 1985; Pikulski & Chard, 2005). Nevertheless, readers might rely on this route to recognize new or unfamiliar words, and then after several encounters, these words become stored in their orthographic lexicon (Doctor & Coltheart, 1980; Ehri, 2014; Share, 1999, 2004; Ziegler & Goswami, 2005). Meanwhile, the degree of error during phonological recoding may also depend on the transparency of the language (Seymour et al., 2003). Although recognizing words via phonological recoding is error prone in deep orthographies (i.e., languages with irregular grapheme-phoneme mappings), novice readers in languages with transparent orthography (i.e., languages with consistent grapheme-phoneme mappings), such as German, may recognize words correctly in their first year of formal reading (Landerl & Wimmer, 2008).

Unlike phonological recoding, orthographic decoding retrieves words directly from the orthographic lexicon rather than recoding them letter-by-letter (Landerl & Wimmer, 2008). However, this efficient processing is only feasible for familiar words that have already been encountered and stored in the orthographic lexicon (Zarić & Nagler, 2021). Recognizing words via orthographic decoding is typically faster and more efficient than phonological recoding (Acha & Perea, 2008; Thaler et al., 2004). As a result, readers have more cognitive resources available for higher-order comprehension processes at the sentence and text levels (Perfetti & Hart, 2002; Pikulski & Chard, 2005), which facilitates reading comprehension (e.g., Fuchs et al., 2001; Kim et al., 2015).

According to the DRC model, efficient word recognition involves a combination of phonological recoding and orthographical decoding processes (Coltheart et al., 2001; Zarić & Nagler, 2021). Even though beginning readers may rely solely on phonological reading, with time and effort, they learn to recognize familiar words through orthographical decoding and unfamiliar words through phonological recoding to become skilled readers (Castles et al., 2018). Hence, struggling readers might get stuck in the non-lexical, less efficient route, relying on phonological recoding to recognize words slowly but accurately, whereas skilled readers rely more on the lexical, more efficient route and employ both phonological recoding and orthographical decoding to recognize words accurately and quickly (Zarić & Nagler, 2021).

2.1.2 Relationship of Word Recognition Accuracy and Word Reading Speed

As previously indicated, in languages with a transparent orthography, such as German, recognizing words utilizing phonological recoding may be already accurate by the end of one year of formal teaching (Landerl & Wimmer, 2008). In contrast, readers in languages with a deep orthography, such as English or Danish, can take twice as long as their German counterparts to correctly recognize words via phonological recoding (Seymour et al., 2003). Nevertheless, recognizing words via phonological recoding is still slower and less efficient than orthographical decoding (Castles et al., 2018). This explanation is also consistent with the findings of an 8-year longitudinal research conducted by Landerl and Wimmer (2008) who found that even German poor readers were able to recognize 72% of the words and 61% of the pseudowords correctly by the end of Grade 1 via phonological recoding, but disparities in word reading speed between poor and good readers remained large and stable until the end of Grade 8 (see also Bast & Reitsma, 1998; Gangl et al., 2018). Moreover, 70% of poor readers whose word reading speed was one standard deviation below the group mean at the end of first grade were still among poor readers by the end of the study. These results suggest that when comparing skilled and struggling German readers, word reading speed should be prioritized above word reading accuracy.

Nevertheless, the findings of Juul et al. (2014) suggest that focusing only on word reading speed might be a misguided prioritization. According to Juul et al., before readers can recognize words quickly, they must first achieve a particular degree of word reading accuracy. In their study, Juul et al. monitored the development of 172 Danish beginning students' word reading skills from the end of kindergarten (Grade 0) to the end of Grade 2. At the end of Grade 0, participants were tested on phoneme awareness, letter name knowledge, rapid automatized naming, and word reading accuracy. Additionally, tests of oral word-reading accuracy and speed were administered every second month throughout Grades 1 and 2. Scatterplots of accuracy and speed scores revealed a curvilinear distribution, with word reading speed increasing once individuals achieved a 70% accuracy. Furthermore, correlations between word reading accuracy and speed for children with less than 70% accuracy scores were not statistically significant at any of the measurement points. Apparently, the gains in word reading speed were larger after readers achieved an accuracy score of 70%. This accuracy score was established as the basic accuracy level.

Basic accuracy level is not the only feature that might affect the development of word reading speed. The basic accuracy achievement time, that is, the time it takes to achieve a score of 70% accuracy, was also found to be an important predictor of word reading speed at the end of Grade 2. The average reader required around 4 months of formal reading instruction to attain the basic accuracy level, although some children needed twice as much time. Even among children who achieved the basic accuracy level, the words-per-minute score at the end of Grade 2 ranged from 28 to 181, indicating that some participants either just entered the speed dimension, or they had specific developmental problems (Juul et al., 2014). Hence, provided that no developmental issues are present, the longer a reader has achieved the basic accuracy level, the more time for improving word reading speed.

2.1.3 Prosody

Many researchers support that prosody, which “refers to reading with expression” (Rasinski et al., 2009 p. 351), is important for reading fluency (e.g., Dowhower, 1991; Hudson et al., 2008) and reading comprehension (e.g., Benjamin & Schwanenflugel, 2010; Rasinski et al., 2009). However, prosody is often neglected in favor of word reading accuracy and word reading automaticity because of its complexity.

The quality of prosodic reading can be assessed by evaluating variations in a) pitch, b) duration, c) stress, and d) pausing (Couper-Kuhlen, 1986; Kuhn et al., 2010). *Pitch* refers to the vibration of a speaker’s vocal folds and indicates the rises and falls of the voice while reading orally or speaking (Dowhower, 1991). The variation in pitch, which is captured by changes in fundamental frequency, may differ substantially between languages. For example, English speakers start a phrase with a high pitch and show more pitch variations than German speakers (Mennen et al., 2012). Moreover, pitch could also depend on gender, with male speakers older than 15 years of age having lower fundamental frequencies than female speakers (Lee et al., 1999). Finally, skilled readers demonstrate a larger pitch range and more appropriate pitch rises and falls during oral reading than struggling readers (Álvarez-Cañizo et al., 2015; Benjamin et al., 2013).

Duration refers to the pronunciation length of vowels and syllables, which can depend on the position of a vowel in a word or a phrase (Dowhower, 1991). Vowels in stressed words or at the final position of a phrase have a longer pronunciation than the same vowel in unstressed words (Temperley, 2009). Even though duration might depend on the individual speaking rate, stressing words at the final position of a phrase has been found to be a strong indicator of chunking the reading material, which in turn facilitates reading comprehension (Dowhower, 1991).

Stress corresponds to the loudness used to read a vowel or a word and is quantified in decibels (Ktori et al., 2018). In English, stressing a word is of outmost importance, because it indicates whether a word is a noun or a verb in some cases (Ktori et al., 2018). However, rules for stressing a syllable differs between languages. Some languages make use of special diacritics to indicate which syllable to stress, and other languages make use of other lexical features such as morphemes (Protopapas, 2006; Rastle & Coltheart, 2000). Furthermore, research in English has shown that struggling readers insert a stress more often compared to skilled readers (Clay & Imlach, 1971).

Finally, *pausing* refers to moments of silence within and between sentences while reading (Kuhn et al., 2010) and can occur for several reasons. For example, readers might need time to take a breath after a long sentence, comprehend a sentence, or decode a word (Godde et al., 2020). Depending on the language or reading skill, the duration, frequency, and appropriateness of pauses differ (Godde et al., 2022). For example, English, German, and French readers take longer pauses than Italian readers and shorter pauses than Spanish readers (Campione & Veronis, 2002). Furthermore, several studies have shown that skilled readers take more appropriate, shorter, and fewer pauses than struggling readers (Álvarez-Cañizo et al., 2015; Miller & Schwanenflugel, 2006).

2.1.4 Present Methods to Prosody Assessment

As explained above, capturing, and analyzing reading prosody is not an easy task. The most common methods to date would be prosodic transcription, reading prosody scales, and spectrographic analysis (Wolters et al., 2022). These methods can either be employed individually or in combination to capture different prosodic aspects. During prosodic transcription, transcribers listen to an audio file and annotate all prosodic cues with the help of a transcription system, which for the sake of comparability usually follows common rules.

By following these rules, transcribers can divide a signal into segments and mark changes in pitch, accentuation, and lengthening (e.g., Bressem, 2013). However, transcribing a signal can be rigorous and requires the transcriber to be familiar with the transcription rules. One of the most popular transcription systems in American English is Tones and Break Indices (ToBi; Silverman et al., 1992), whereas in German language, the Gesprächsanalytisches Transkriptionssystem (conversation-analytic transcription system, GAT, Selting et al., 1998) has gained popularity over the years.

The second method, reading prosody rating scales, involves rating different facets of reading competence with the help of standardized rating scales. These facets mostly include expressiveness, pace, smoothness, and phrasing (Wolters et al., 2022). Rating can be either achieved in real time by listening to and rating a reader's oral reading or post hoc by listening to a recording. The main advantage of prosodic rating scales is that they can be employed in the field, which makes them ideal for school settings (Miller & Schwanenflugel, 2006). However, raters must be familiar with the respective rating scales before they are administered.

In contrast the third most common method, spectrographic analysis, cannot be employed in real time. Nevertheless, spectrographic analysis allows for the extraction and quantification of prosodic features, such as pitch and stress by visualizing the signal in a spectrogram (Schwanenflugel et al., 2004). However, a certain degree of expertise is required to conduct this analysis and comprehend the spectrogram. Despite its importance in prosody research, identifying prosodic patterns in the signal manually can be a laborious and error-prone process.

The non-linear analysis tool RQA, which has been effectively applied in another field of reading research in recent years, may be able to overcome the challenge of the spectrographic analysis (see Wallot, 2017, for an extensive tutorial). RQA can quantify the complexity of a time-series and depict the patterns of a signal semiautomatically. Hence, even though this method also cannot be used in real time and is as complex as spectrographic analysis, it might prove to be a valuable tool for objectively and reliably detecting prosodic patterns in a large sample without the need to scrutinize patterns manually.

2.2 Reading Interventions

Reading interventions is comprised of two components, the instructional content and the organization. The content of the instruction addresses the content taught to improve a specific reading outcome, whereas the organization refers to the structure of the intervention, such as group size, training duration, and session length. The most widely used reading interventions could be classified into *phonemic awareness instruction*, *phonics instruction*, *reading fluency interventions*, and *reading comprehension interventions* (National Reading Panel, 2000; Suggate, 2016).

Phonemic awareness instruction teaches participants how to identify and manipulate phonemes in words (Galuschka et al., 2014; Küspert & Schneider, 2018). These interventions include blending phonemes into a word by listening to a sequence of spoken sounds and combining them into a word, segmenting a word (i.e., counting or breaking a word into sounds), omitting a phoneme (i.e., identifying the new word when one phoneme is deleted), and adding a phoneme to a word and pronouncing the new word (Galuschka et al., 2014; National Reading Panel, 2000). In phonemic awareness instruction, the phoneme manipulation is primarily conducted orally without the use of letters (National Reading Panel, 2000).

In contrast, *phonics instruction* teaches participants the grapheme-phoneme correspondence rules of a given language and how they can be used to identify words. They mainly focus on converting single letters into sounds and blending them into words, transforming sounds into letters to write words, or using grapheme-phoneme correspondence rules to identify unfamiliar words, but they might also include dividing a word into larger segments such as syllables or morphemes (National Reading Panel, 2000).

Reading fluency interventions, in contrast, focus on improving word recognition abilities through repeated or guided oral reading. During this method, participants must read a meaningful short passage repeatedly until a satisfying level of word reading accuracy and reading speed is achieved before beginning to read a new passage (Samuels, 1979). Furthermore, according to Samuels (1979), presenting the progress to the reader (e.g., in the form a graph) could encourage them. This method's goal is to help readers achieve automatic word recognition and a consequent improvement in reading comprehension.

Finally, in *reading comprehension interventions*, participants learn to extract textual content, synthesize it, and relate it to existing knowledge (Galuschka et al., 2014). This can be accomplished by employing cognitive strategies, for example, generating questions about the text as it is read to process the text more actively (National Reading Panel, 2000). However, comprehension interventions are most suitable for readers who have already acquired fundamental reading skills and can efficiently decode words (Suggate, 2010, 2016).

Despite the ample studies on various reading interventions and their efficacy (e.g., Galuschka et al., 2014, 2020; Gersten et al., 2020; Neitzel et al., 2021), deciding on the best suited reading intervention or a combination of reading interventions for each grade becomes complicated. For example, according to the meta-analysis of Galuschka et al. (2014), phonics instruction was the most thoroughly investigated intervention and the only one that had a significant effect on reading and spelling performance of individuals with reading disabilities. However, even though Galuschka et al. suggested that neither phonemic awareness nor reading fluency interventions alone are sufficient to achieve substantial improvements in individuals with reading difficulties, this result should be interpreted with caution. The effect sizes of these three approaches were very similar. Hence, phonemic awareness and reading fluency interventions could have failed to reach significance because of the low number of studies included in this meta-analysis. Similarly, given the low number of studies that had employed comprehension interventions, the effect of this intervention type could not be evaluated. Lastly, within the scope of this meta-analysis, establishing which intervention is suitable for a particular age or grade level was not possible.

The relationship between intervention type and grade was investigated in the meta-analysis of Suggate (2010). Suggate's findings suggest that even though phonics interventions had a larger short-term overall effect from kindergarten to Grade 1, comprehension and mixed interventions (i.e., interventions that included a phonemic awareness or phonics instruction and a comprehension component) were associated with greater short-term overall effects after Grade 1. Furthermore, a new meta-analysis conducted a few years later by Suggate (2016) revealed that to achieve the optimal long-term effects, phonemic awareness should be implemented in preschool and kindergarten, fluency and mixed trainings in Grades 1 and 2, and reading comprehension interventions should be implemented starting from Grade 3. Hence, despite the variety of results, phonemic awareness and phonics instruction apparently play an important role before the children start learning to read, whereas fluency and reading comprehension interventions work best when children can already decode words efficiently.

3 The Present Research

The purpose of this dissertation was to shed light on two major research topics concerning word reading fluency in the German language: the extent that basic accuracy level and basic accuracy achievement time affect word-recognition speed and the relationship between prosodic patterns and reading comprehension. According to the theory of automaticity (LaBerge & Samuels, 1974), even though beginning readers can recognize words accurately to some extent, they are still slow because of inefficient word recognition skills. The findings of Juul et al. (2014) in Danish also give support to this theory by showing that beginning readers first need to achieve a certain degree of word reading accuracy before their word reading speed starts increasing. Furthermore, the earlier that readers achieve a sufficient word recognition level, the more time they have at their disposal to develop their word reading speed (Juul et al., 2014).

Consequently, the first aim of this thesis was to investigate whether such a word reading accuracy level also exists in the German language and whether the timepoint at which this level is achieved also has an effect on word reading speed. Moreover, this thesis extends the findings of Juul et al. (2014) by investigating whether a syllable reading training has different effects on participants that had attained this level before partaking in the training compared to participants who had not reached this level. In Study 1, a sample of German fourth graders was chosen because word recognition differences in this grade should be more apparent than in other grades. In this study, one objective was to establish a word-recognition accuracy level for the German language by examining whether participants responded differently to a syllable reading intervention, depending on the extent of their word-recognition accuracy before our training. The study is presented in detail in Chapter 4. Study 2 extended the results of the first study by investigating the development of word recognition in a longitudinal sample of German primary school children. We were particularly interested in the effect of the basic accuracy achievement time on the development of children's word-recognition speed and reading comprehension throughout primary school. The study is presented in detail in Chapter 5.

The second aim of this thesis was to investigate prosody, the third component of reading fluency. In Study 3, it was investigated whether prosodic patterns can differentiate between skilled and struggling German readers in primary school by employing a non-linear analysis tool, *Recurrence Quantification Analysis* (RQA; see Wallot, 2017, for an extensive tutorial), to capture prosodic patterns that might emerge in the signals of second and fourth graders' oral readings. Furthermore, prosodic patterns obtained with RQA were analyzed to determine whether they provided additional information to prosodic features obtained with standard methods. To my knowledge, no study to date has quantified prosodic patterns while reading in German. The study is presented in detail in Chapter 6.

4 Study 1: Modelling the Relationship of Accurate and Fluent Word Recognition in Primary School

A version of this chapter was published in:

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

https://doi.org/10.1016/j.lindif.2019.101779

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, 1996; Perfetti et al., 2005; Vellutino et al., 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 et al. (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 et al., 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; Pikulski & Chard, 2005), which in turn fosters text comprehension (e.g., Fuchs et al., 2001; Kim et al., 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 et al., 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 et al., 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 et al., 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 et al. (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 suggest 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 et al., 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 et al., 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 et al., 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 et al., 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 et al., 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 et al., 2013; Huemer et al., 2010; Dutch: Berends, & Reitsma, 2006; Wentink et al., 1997; French: Ecalle et al., 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 et al., 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 et al., 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 et al., 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 post-test 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 pre-test to read words more fluently after the intervention at post-test compared to untrained children in the control group who scored above the basic accuracy level at pre-test (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 pre-test compared to same-skilled children in the control condition who were above the basic accuracy level at pre-test, 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 et al., 2017) for word recognition processes and ELFE 1-6 (Lenhard & Schneider, 2006) for reading comprehension in Study 1 and the revised ELFE II (Lenhard et al., 2017) in Study 2. Children with pre-test 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 et al., 2016).

The groups in the treatment condition received the intervention after the pre-test, and the groups in the control condition (wait-list) received the same intervention after the post-test. 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 et al., 2017). In the previous study, we examined the effects of the intervention by comparing the integrated post-test scores of orthographical decoding (computed as the ratio of accuracy and response time) and reading comprehension post-test 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 post-test. 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 pre-test 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 pre-test to post-test we conducted ANOVAs with the difference scores of accuracy and speed of orthographical decoding and reading comprehension at post-test minus their corresponding pre-test scores as outcome variables (cf. Judd et al., 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.

Table 1

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

	Screening sample	Treatment group	Wait-list control group
*N t*otal 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).

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 post-test. 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 et al. (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 post-test 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). According to the rules of thumb provided by Cicchetti (1994), the estimated test-rest reliability was acceptable for the accuracy percentage score, $\rho_I = .62$, 95% CI [.54, .68], and good for the words-per-minute score, $\rho_I = .77$, 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 post-test 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 et al., 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 et al., 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 et al., 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 et al., 2015) and lmerTest (Kuznetsova et al., 2019). The low ICC ($\rho < .05$) for the words-per-minute score at pre-test 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 et al., 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 pre-test, which would be an indication for anomalous responding. None of the participants matched the second outlier identification rule.

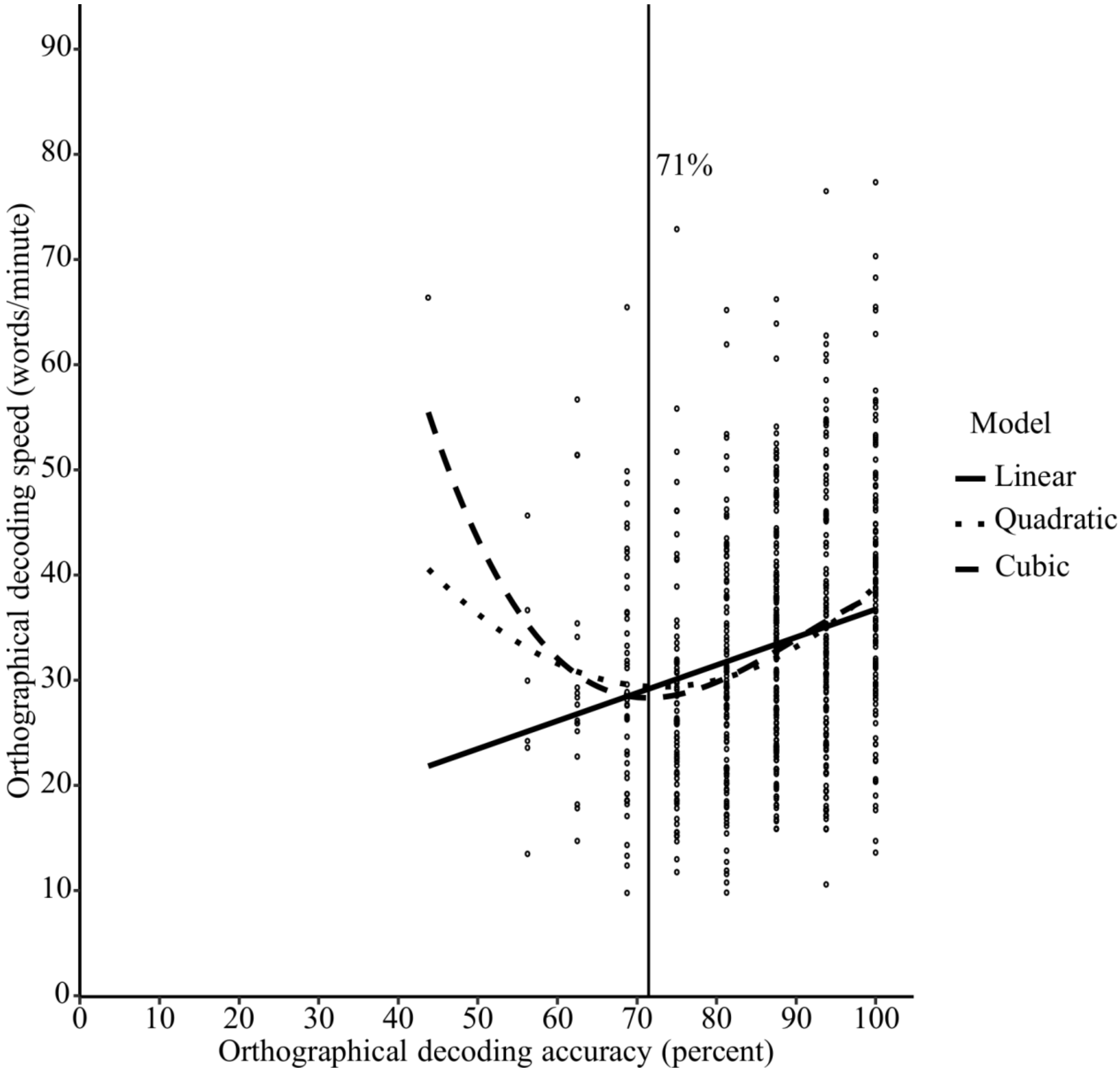
Accuracy-before-Speed Pattern

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 pre-test was plotted against the accuracy percentage score at pre-test 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 pre-test was included as the outcome variable, and the accuracy percentage score at pre-test 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.

Figure 1

Association Between Orthographical Decoding Accuracy and Orthographical Decoding Speed



Note. 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).

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.

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

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):

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

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

[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.

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 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.

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 Post-test

	Orthographical decoding accuracy					Orthographical decoding speed					Reading comprehension				
	Pre-test		Post-test		Cohen's *d*	Pre-test		Post-test		Cohen's *d*	Pre-test		Post-test		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.

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 post-test (Hypothesis 2). The percentage score of the accuracy of orthographic decoding at post-test 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 pre-test 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, $B = -11.25$, $SE = 4.13$, $p = .007$, $\Delta R^2 = .04$, 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).

Table 4

Parameter Estimates for Multiple Regression Analysis with the Accuracy of Word Recognition as Outcome Variables at the Post-test, Treatment Condition as Predictor, and Pre-test 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
Pre-test scores	0.32*	0.13	.03
Goodness of fit	R^2 = .17, $F(4,149) = 7.64$, $p < .001$		

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

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

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 post-test 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 post-test 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 pre-test 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 post-test 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 post-test 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 pre-test 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 pre-test and then received the treatment scored higher in reading comprehension at post-test than the other three groups.

Table 5

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

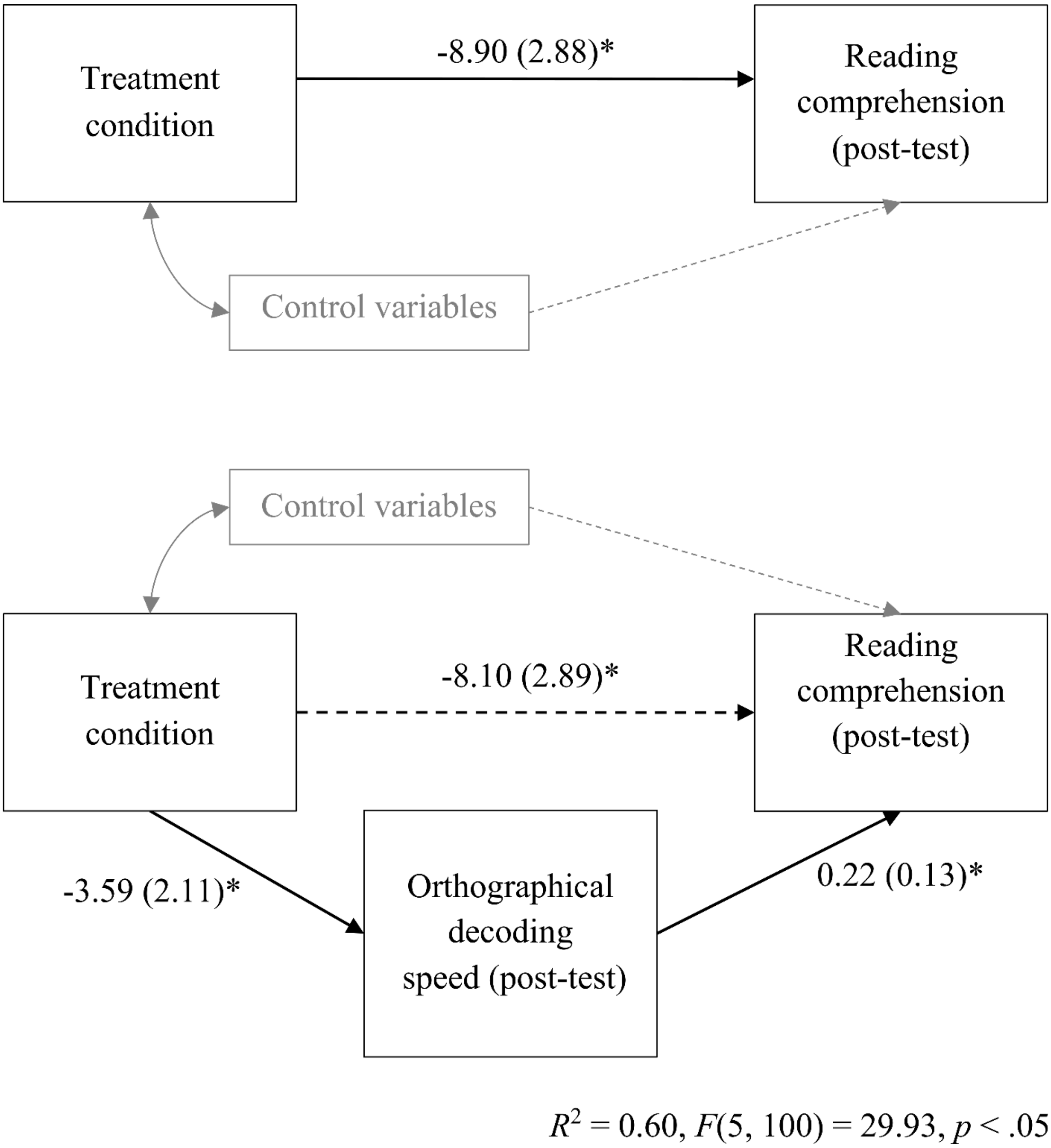
	Orthographical decoding speed			Reading comprehension		
	B	*SE*	ΔR^2	*B*	*SE*	ΔR^2
Intercept	34.89**	1.27		51.77**	2.14	
Trained above (reference group) vs. trained below	-0.86	2.63	.00	-10.26*	3.50	.03
Trained above (reference group) vs. untrained above	-4.21*	2.02	.02	-8.01*	2.80	.03
Trained above (reference group) vs. untrained below	-1.81	3.23	.00	-13.04*	4.15	.03
Pre-test 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 pre-test 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 2

Mediation Model Predicting Reading Comprehension



Note. 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 post-test with orthographical decoding speed at post-test as mediator. Unstandardized regression weights with associated standard errors in parentheses.

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

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 post-test was included as an outcome variable and the words-per-minute score at the post-test 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 pre-test 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 et al., 2006), they still lag behind their native-speaking peers in reading comprehension (e.g., Proctor et al., 2006; Lesaux et al., 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). Akademiai Kiado.

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

Berends, I. E., & Reitsma, P. (2006). Remediation of fluency: Word specific or generalised training effects? *Reading and Writing, 19*(2), 221–234. https://doi.org/10.1007/s11145-005-5259-3

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

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest, 19*(1), 5–51. https://doi.org/10.1177/1529100618772271

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Lawrence Erlbaum Associates.

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*(4), 507–532. https://doi.org/10.1017/S0142716499004038

Coltheart, M. (2005). Modeling reading: The dual-route approach. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 6–23). Blackwell Publishing. https://doi.org/10.1002/9780470757642

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*(1), 204–256. https://doi.org/10.1037/0033-295x.108.1.204

Daneman, M. (1996). Individual differences in reading skills. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 512–538). Lawrence Erlbaum Associates. (Original work published 1991)

Dudenredaktion (2017). *Duden: Die deutsche Rechtschreibung* [Duden: The German spelling]. 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*(3), 554–561. https://doi.org/10.1016/j.compedu.2008.10.010

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, *9*(2), 167–188. https://doi.org/10.1207/s1532799xssr0902_4

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*(3), 393–447. https://doi.org/10.3102/00346543071003393

Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science, 18*(3), 233–239. https://doi.org/10.1111/j.1467-9280.2007.01882.x

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*(3), 239–256. https://doi.org/10.1207/S1532799XSSR0503_3

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*(2), Article e89900. https://doi.org/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*(3), 473–487. https://doi.org/10.1007/s11618-015-0650-6

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2019). *irr: Various Coefficients of Interrater Reliability and Agreement.* R package (Version 0.84.1) [Computer software]. 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*(1), 24–40. https://doi.org/10.1080/10888438.2017.1339709

Grainger, J., Lété, B., Bertand, D., Dufau, S., & Ziegler, J. C. (2012). Evidence for multiple routes in learning to read. *Cognition*, *123*(2), 280–292. https://doi.org/10.1016/j.cognition.2012.01.003

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. 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*(6), 398–414. https://doi.org/10.1080/10888438.2012.753452

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*(4), 317–340. https://doi.org/10.1080/10888430903150659

Judd, C. M., Kenny, D. A., & McClelland, G. H. (2001). Estimating and testing mediation and moderation in within-subject design. *Psychological Methods, 6*(2), 115–134. https://doi.org/10.1037/1082-989x.6.2.115

Juul, H., Poulsen, M., & Elbro, C. (2014). Separating speed from accuracy in beginning reading development. *Journal of Educational Psychology*, *106*(4), 1096–1106. https://doi.org/10.1037/a0037100

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*(6), 1259–1282. https://doi.org/10.1007/s11145-011-9316-9

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*(1), 131–150. https://doi.org/10.1007/s11145-013-9455-2

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractc Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

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). 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). 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]. 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*(3), 447–471. https://doi.org/10.1007/s11618-014-0560-z

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2019). *lmerTest: Tests in linear mixed effects models.* R package (Version 3.1-0) [Computer software] 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*(2), 293–323. https://doi.org/10.1016/0010-0285(74)90015-2

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*(1), 150–161. https://doi.org/10.1037/0022-0663.100.1.150

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]. 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]. 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*(6), 475–483. https://doi.org/10.1016/j.appdev.2010.09.004

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). Lawrence Erlbaum Associates.

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*(2), 140–149. https://doi.org/10.1016/j.bandl.2006.04.003

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*(2), 116–133. https://doi.org/10.1006/jecp.1999.2525

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*(4), 391–407. https://doi.org/10.1177/0022219413504996

McArthur, G., Sheehan, Y., Badcock, N. A., Francis, D. A., Wang, H.-C., Kohnen, S., Banales, E., Anandakumar, T., Marinus, E., & Castles, A. (2018). Phonics training for English-speaking poor readers. *Cochrane Database of Systematic Reviews, 11*(11), CD009115. https://doi.org/10.1002/14651858.CD009115.pub3

McGraw, K. O., & Wong S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30–46. https://doi.org/10.1037/1082-989X.1.1.30

Meyer, M. S., & Felton, R. H. (1999). Repeated reading to enhance fluency: Old approaches and new directions. *Annals of Dyslexia, 49*, 283–306. https://doi.org/10.1007/s11881-999-0027-8

Muggeo, V. M. R. (2017). *segmented: Regression models with break-points / Change-points estimation.* R package (Version 0.5-3.0) [Computer software]. 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*(1), 52–57. https://doi.org/10.2378/peu2018.art03d

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* (pp. 29–49). 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*, Article 1635. https://doi.org/10.3389/fpsyg.2017.01635

Perfetti, C. A. (1985). *Reading ability*. 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). John Benjamins.

Perfetti, C. A., Landi, N., & Oakhill J. (2005). In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 227–247). Blackwell Publishing. https://doi.org/10.1002/9780470757642

Pikulski, J., & Chard, D.J. (2005). Fluency: Bridge between decoding and reading comprehension. *The Reading Teacher, 58*(6), 510–519. https://doi.org/10.1598/RT.58.6.2

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*(1), 159–169. https://doi.org/10.1037/0022-0663.98.1.159

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, 11*(3), 165–197. https://doi.org/10.1080/10888430701344322

Price, R. B., Kuckertz, J. M., Siegle, G. J., Ladouceur, C. D., Silk, J. S., Ryan, N. D., Dahl, R. E., & Amir, N. (2015). Empirical recommendations for improving the stability of the dot-probe task in clinical research. *Psychological Assessment, 27*(2), 365–376. https://doi.org/10.1037/pas0000036

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

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*(4), 313–331. https://doi.org/10.1024/1010-0652/a000079

Richter, T., Isberner, M.-B., Naumann, J. & Kutzner, Y. (2013). Lexical quality and reading comprehension in primary school children. *Scientific Studies of Reading, 17*(6), 415–434. https://doi.org/10.1080/10888438.2013.764879

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]. Hogrefe.

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

Scheerer-Neumann, G. (1981). The utilization of intraword structure in poor readers: Experimental evidence and a training program. *Psychological Research, 43*(2), 155–178. https://doi.org/10.1007/BF00309827

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*(4), 1085–1094. https://doi.org/10.3758/s13428-014-0528-1

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology, 94*(2), 143–174. https://doi.org/10.1348/000712603321661859

Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology*, *72*(2), 95–129. https://doi.org/10.1006/jecp.1998.2481

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*(4), 267–298. https://doi.org/10.1016/j.jecp.2004.01.001

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7*(4), 422–445. https://doi.org/10.1037/1082-989X.7.4.422

Snowling, M. J., & Hulme, C. (2005). *The science of reading: A handbook*. Blackwell Publishing. https://doi.org/10.1002/9780470757642

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*(5), 576–590. https://doi.org/10.1177/0022219416638028

Suggate, S.P. (2010). Why what we teach depends on when: Grade and reading intervention modality moderate effect size. *Developmental Psychology, 46*(6), 1556–1579. https://doi.org/10.1037/a0020612

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*(1), 77–96. https://doi.org/10.1177/0022219414528540

Therrien, W. J. (2004). Fluency and comprehension gains as a result of repeated reading: A meta-analysis. *Remedial and Special Education, 25*(4), 252–261. https://doi.org/10.1177/07419325040250040801

Trochim, W. M., Donnelly, J. P., & Arora, K. (2016). *Research methods: The essential knowledge base* (2nd ed.). Cengage Learning.

Urbina, S. (2014) *Essentials of psychological testing* (2nd ed.). 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*(1), 2–40. https://doi.org/10.1046/j.0021-9630.2003.00305.x

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*(3), 163–192. https://doi.org/10.1023/A:1007921805360

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*(1), 53–79. https://doi.org/10.1037/a0024177

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*(3), 169–193. https://doi.org/10.1016/S0022-0965(03)00139-5

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*(3), 369–373. https://doi.org/10.1016/j.bandl.2004.10.010

5 Study 2: The Role of Word-Recognition Accuracy in the Development of Word-Recognition Speed and Reading Comprehension in Primary School: A Longitudinal Examination

A version of this chapter was published in:

Karageorgos, P., Richter, T., Haffmans, M.-B., Schindler, J. & Naumann, J. (2020). The role of word-recognition accuracy in the development of word-recognition speed and reading comprehension in primary school: A longitudinal examination. *Cognitive Development, 56*.

https://doi.org/10.1016/j.cogdev.2020.100949

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

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 et al., 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 et al. (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 et al. (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 et al. (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 et al., 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 et al., 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 et al., 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 et al., 1999; Ziegler et al., 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 et al., 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 et al., 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 et al. (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 et al., 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.

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.

Table 1

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

Measurement point	$n_{\text{cohort 1}}$	$n_{\text{cohort 2}}$	n_{females}	Age *M* (*SD*)	Word recognition		Reading comprehension *M* (*SD*)
					Accuracy *M* (*SD*)	Speed *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).

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 et al., 2017; see also Richter et al., 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 et al., 1995; Baayen et al., 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 et al., 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].

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).

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 et al., 2019) and *lmerTest* with Satterthwaite adjustments to denominator degrees of freedom (Kuznetsova et al., 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.

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).

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).

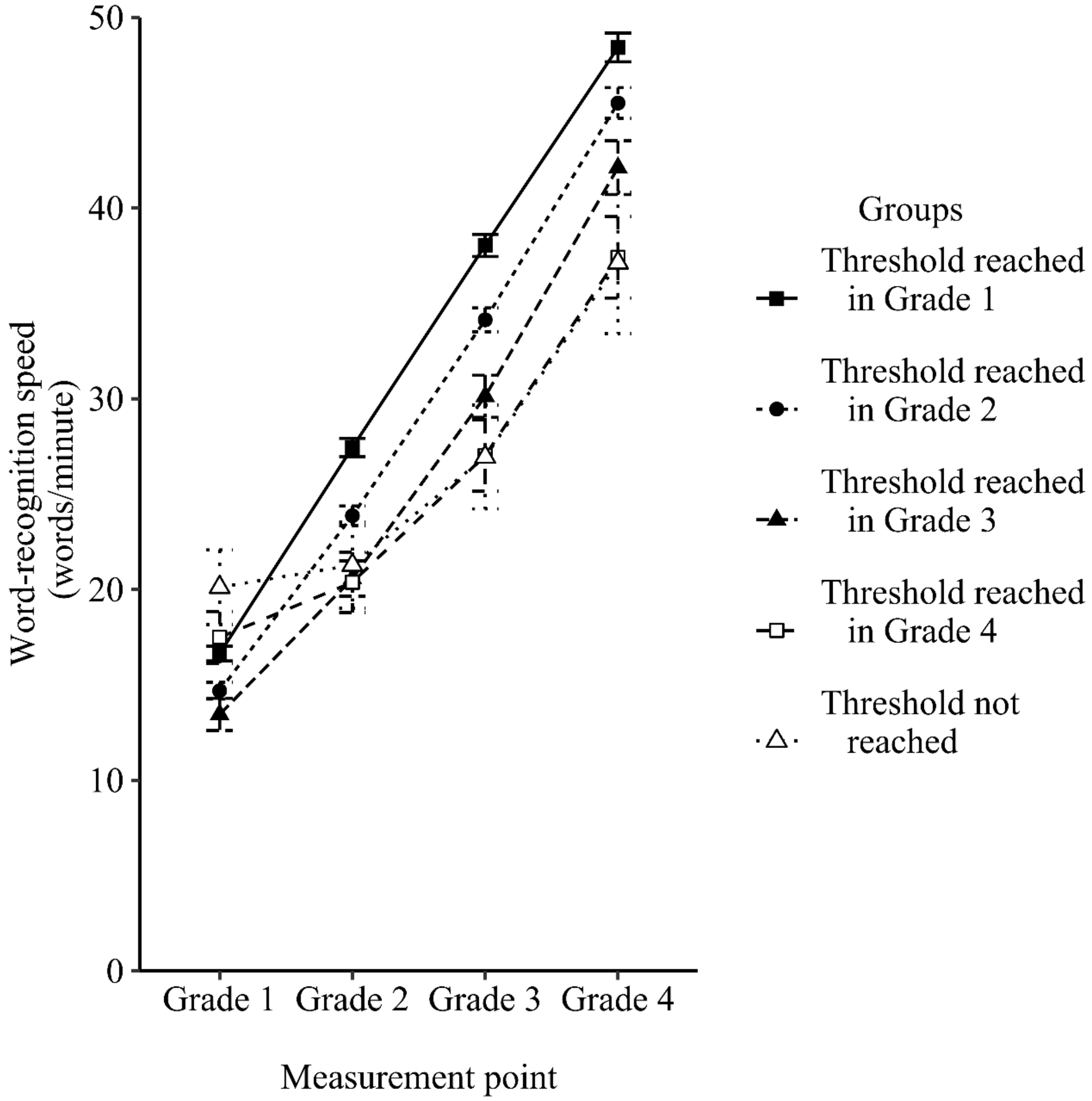
Effects on Word-recognition Speed

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 et al., 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.

Figure 1

Development of Word-recognition Speed Throughout Primary School



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

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.

Parameter	Word-recognition speed	Reading comprehension
	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).

* $p < 0.05$. ** $p < 0.01$.

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.

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.

Parameter	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
	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	29.47	26.52	22.56	18.29	58.48	54.06	44.12	42.14	34.62
	(0.92)*	(0.73)*	(1.34)*	(3.11)*	(4.53)*	(2.11)*	(1.92)*	(2.46)*	(3.73)*	(4.00)*
Time	10.57	10.24	9.60	6.59	6.07	15.56	15.34	11.24	10.87	4.61
	(0.23)*	(0.25)*	(0.58)*	(0.88)*	(2.35)*	(0.56)*	(0.65)*	(1.03)*	(1.78)*	(2.32)*
Squared time	-0.12	0.52	1.16	1.98	0.67	-3.65	-2.80	-1.81	2.28	1.28
	(0.18)	(0.19)*	(0.32)*	(0.80)*	(0.77)	(0.82)*	(0.86)*	(1.35)	(2.68)	(2.88)
Gender	2.68	-0.91	0.09	-1.71	9.99	-2.01	-7.15	-1.79	-3.82	-4.65
	(0.77)*	(0.66)	(1.35)	(3.27)	(5.06)*	(1.86)	(1.85)*	(2.48)	(3.77)	(3.62)
First language	-0.37	-0.39	-3.02	3.00	-0.06	6.52	5.43	0.19	-6.29	0.85
	(0.87)	(0.70)	(1.35)*	(3.21)	(5.00)	(2.08)*	(1.95)*	(2.48)	(3.75)*	(3.55)

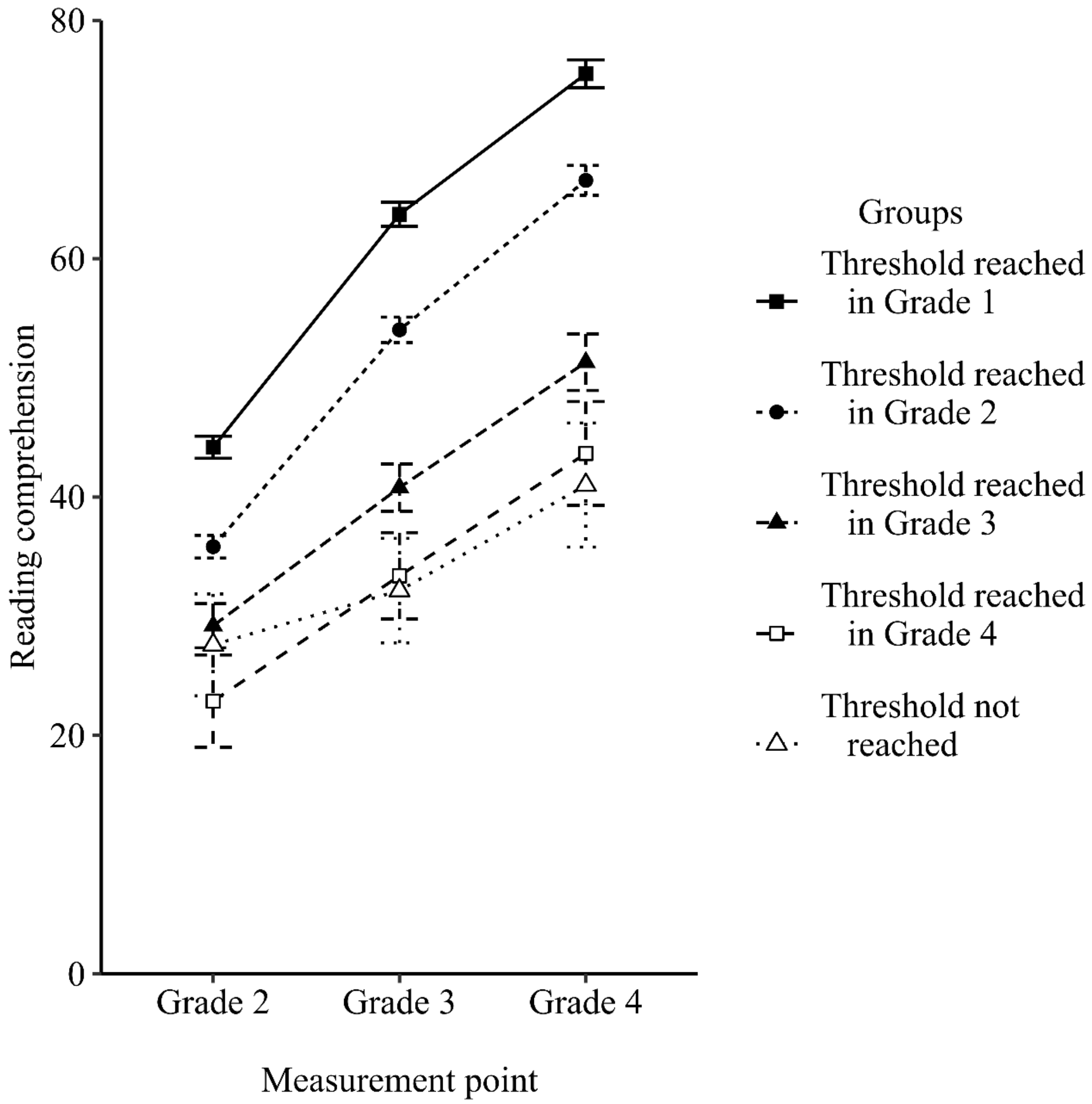
Parameter	Word recognition speed					Reading comprehension				
	By the end of Grade 1	By the end of Grade 2	By the end of Grade 3	By the end of Grade 4	Later than the Grade 4	By the end of Grade 1	By the end of Grade 2	By the end of Grade 3	By the end of Grade 4	Later than the Grade 4
	B (SE)	*B (SE)*	*B (SE)*	*B (SE)*	*B (SE)*	*B (SE)*	*B (SE)*	*B (SE)*	*B (SE)*	*B (SE)*
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).

* $p < 0.05$

Figure 2

Development of Reading Comprehension Throughout Primary School



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

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.
* $p < 0.05$.

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 is the 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.

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*(2), 245–264. https://doi.org/10.1348/000712607X224478

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

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

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database* [CD-ROM]. 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) [Computer software]. 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*(1), 3–19. https://doi.org/10.1037/0012-1649.35.1.3

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). Blackwell Publishing. https://doi.org/10.1002/9780470757642

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*(3), 203–219. https://doi.org/10.1016/j.jecp.2005.06.005

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest, 19*(1), 5–51. https://doi.org/10.1177/1529100618772271

Ciccheti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

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*(1), 204–256. https://doi.org/10.1037/0033-295x.108.1.204

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, *9*(2), 167–188. https://doi.org/10.1207/s1532799xssr0902_4

Feingold, A. (2019). Time-varying effect sizes for quadratic growth models in multilevel and latent growth modelling. *Structural Equation Modeling, 26*(3), 418–429. https://doi.org/10.1080/10705511.2018.1547110

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). Lawrence Erlbaum Associates.

Frith, U. (1986). A developmental framework for developmental dyslexia. *Annals of Dyslexia*, *36*(1), 69–81. https://doi.org/10.1007/BF02648022

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2019). *irr: Various coefficients of interrater reliability and agreement.* R package (Version 0.84.1) [Computer software]. 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*(1), 24–40. https://doi.org/10.1080/10888438.2017.1339709

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*(1), 6–10. https://doi.org/10.1177/074193258600700104

Grainger, J., Lété, B., Bertand, D., Dufau, S., & Ziegler, J. C. (2012). Evidence for multiple routes in learning to read. *Cognition*, *123*(2), 280–292. https://doi.org/10.1016/j.cognition.2012.01.003

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, *2*(2), 127–160. https://doi.org/10.1007/BF00401799

Juul, H., Poulsen, M., & Elbro, C. (2014). Separating speed from accuracy in beginning reading development. *Journal of Educational Psychology*, *106*(4), 1096–1106. https://doi.org/10.1037/a0037100

Karageorgos, P., Müller, B., & Richter, T. (2019). Modelling the relationship of accurate and fluent word recognition in primary school. *Learning and Individual Differences, 76*. https://doi.org/10.1016/j.lindif.2019.101779

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*(1), 10–16. https://doi.org/10.1111/ldrp.12025

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. 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*(2), 310–321. https://doi.org/10.1037/0022-0663.100.2.310

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*(3), 447–471. https://doi.org/10.1007/s11618-014-0560-z

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2019). *lmerTest: Tests in linear mixed effects models.* R package (Version 3.1-0) [Computer software]. 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*(2), 166–185. https://doi.org/10.1016/j.jecp.2005.09.003

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*(2), 293–323. https://doi.org/10.1016/0010-0285(74)90015-2

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*(1), 150–161. https://doi.org/10.1037/0022-0663.100.1.150

Lange-Küttner, C. (2005). Word structure effects in German and British reading beginners. *Zeitschrift für Pädagogische Psychologie, 19*(4), 207–218. https://doi.org/10.1024/1010-0652.19.4.207

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*(1), 19–37. https://doi.org/10.1024/1010-0652/a000027

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) [Computer software]. 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]. 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*(2), 116–133. https://doi.org/10.1006/jecp.1999.2525

McGraw, K. O., & Wong S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30–46. https://doi.org/10.1037/1082-989X.1.1.30

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* (pp. 29–49). Budrich.

Perfetti, C. A. (1985). *Reading ability*. 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). 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). Blackwell Publishing. https://doi.org/10.1002/9780470757642

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

Pikulski, J., & Chard, D.J. (2005). Fluency: Bridge between decoding and reading comprehension. *The Reading Teacher, 58*(6), 510–519. https://doi.org/10.1598/RT.58.6.2

Price, R. B., Kuckertz, J. M., Siegle, G. J., Ladouceur, C. D., Silk, J. S., Ryan, N. D., Dahl, R. E., & Amir, N. (2015). Empirical recommendations for improving the stability of the dot-probe task in clinical research. *Psychological Assessment, 27*(2), 365–376. https://doi.org/10.1037/pas0000036

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*(4), 313–331. https://doi.org/10.1024/1010-0652/a000079

Richter, T., Isberner, M.-B., Naumann, J. & Kutzner, Y. (2013). Lexical quality and reading comprehension in primary school children. *Scientific Studies of Reading, 17*(6), 415–434. https://doi.org/10.1080/10888438.2013.764879

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]. 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). De Gruyter. https://doi.org/10.1515/9781614514909

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*(2), 143–174. https://doi.org/10.1348/000712603321661859

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition, 55*(2), 151–218. https://doi.org/10.1016/0010-0277(94)00645-2

Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology*, *72*(2), 95–129. https://doi.org/10.1006/jecp.1998.2481

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*(4), 267–298. https://doi.org/10.1016/j.jecp.2004.01.001

Snowling, M. J., & Hulme, C. (2005). *The science of reading: A handbook.* Blackwell Publishing. https://doi.org/10.1002/9780470757642

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), 1–28. https://doi.org/10.1348/000712603762842075

Torgesen, J. K. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research & Practice*, *15*(1), 55–64. https://doi.org/10.1207/SLDRP1501_6

Verhoeven, L., & Perfetti, C. (2008). Advances in text comprehension: Model, process and development. *Applied Cognitive Psychology*, *22*(3), 293–301. https://doi.org/10.1002/acp.1417

Wimmer, H. (1996). The nonword reading deficit in developmental dyslexia: Evidence from children learning to read German. *Journal of Experimental Child Psychology*, *61*(1), 80–90. https://doi.org/10.1006/JECP.1996.0004

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*(4), 1026–1036. https://doi.org/10.1037/a0035187

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, *131*(1), 3–29. https://doi.org/10.1037/0033-2909.131.1.3

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*(3), 169–193. https://doi.org/10.1016/S0022-0965(03)00139-5

6 Study 3: Distinguishing Between Struggling and Skilled Readers Based on Their Prosodic Speech Patterns in Oral Reading: An Exploratory Study in Grades 2 and 4

A version of this chapter was submitted as:

Karageorgos, P., Wallot, S., Müller, B., Schindler, J., & Richter, T. (2022). *Distinguishing between struggling and skilled readers based on their prosodic speech patterns in oral reading: An exploratory study in Grades 2 and 4* [Manuscript submitted for publication]. Department of Psychology IV, University of Würzburg.

Abstract

The purpose of this study was to examine if prosodic patterns in oral reading derived from Recurrence Quantification Analysis (RQA) could distinguish between struggling and skilled German readers in Grades 2 and 4. Furthermore, we investigated whether models estimated with RQA measures outperformed models estimated with prosodic features derived from prosodic transcription. During an intervention study sixty-seven second graders and sixty-nine fourth graders, who were divided at the beginning of the study into struggling and skilled readers, read an age-appropriate text. Audio signals of the oral readings were analyzed using recurrence quantification analysis and transcribed using the transcription system GAT. According to the findings, struggling second graders make longer pauses with longer intervals between them, whereas struggling fourth graders have longer intervals between recurring stresses, a complex and unpredictable pitch pattern, and more recurrent stresses, indicating that they remain in a state for a longer period of time. However, even though models with prosodic patterns provided additional information, they did not outperform models with prosodic characteristics. These findings suggest that in addition to pausing as an important indicator for reading comprehension, the importance of other prosodic patterns for reading comprehension may increase over time.

Thinking back to our school days, we all remember classmates who, in comparison to their peers, struggled while reading a passage aloud. These students read passages word for word, syllable by syllable, or even letter by letter, with long pauses in between, and they took longer than most of their peers to complete reading the passage. This observable behavior has been described as a lack of reading fluency in reading research and by literacy educators (Allington, 1983; Dowhower, 1991).

Reading fluency has received considerable attention since the Report of the National Reading Panel (2000). However, providing a widely accepted definition appears rather challenging (Godde et al., 2020). A consensus exists that reading fluency is a complex, multifaceted construct, which mainly relies on word recognition accuracy, automaticity, and prosody (e.g., Hudson et al., 2009; Kuhn et al., 2010). Nonetheless, the contribution of each factor to reading fluency remains disputed (Kuhn et al., 2010). Despite the ample research on reading fluency, it tends to focus on accuracy and automaticity (e.g., Fuchs et al., 2001; Kara et al., 2020), whereas prosody is frequently overlooked (Silverman et al., 2013; Wolters et al., 2022). According to Dowhower (1991), one possible explanation for the neglect of prosody could be that measuring word reading accuracy and word reading speed is easier and less time consuming than measuring prosody. In the present study, we employed *Recurrence Quantification Analysis* (RQA – Marwan et al., 2007; Wallot et al., 2014) and *k-fold cross-validation analysis* (Browne, 2000) to explore the differences in the prosodic features of skilled and struggling German readers in second and fourth grade. We were particularly interested in identifying prosodic patterns that can distinguish between skilled and struggling German readers. Furthermore, we investigated whether these prosodic patterns could distinguish between struggling and skilled German readers better than prosodic features, such as appropriate and inappropriate pause occurrences acquired from prosodic transcription. In the following sections, we briefly describe word-reading automaticity, word-reading prosody, and the presently available methods to assess prosody, which provides the groundwork for our study.

Word Reading Automaticity and Prosody

Many researchers regard fluency as more or less synonymous with word reading automaticity (Dowhower, 1991). Word reading automaticity reflects the ability to recognize words accurately and with minimal cognitive effort (Paige et al., 2014). The number of words correctly read per minute, which incorporates both word reading accuracy and word reading speed, is a common metric of word reading automaticity (Benjamin & Schwanenflugel, 2010). Nevertheless, achieving a word recognition accuracy threshold is an important precondition before word reading speed starts improving (Altani et al., 2020; Juul et al., 2014; Karageorgos et al., 2019, 2020). For readers who lack a high level of word reading accuracy, word recognition will most likely be inefficient, resulting in a high demand on cognitive resources and slow reading speed (Castles et al., 2018; LaBerge & Samuels, 1974). Hence, before readers recognize words efficiently, they must first learn the grapheme-phoneme correspondence rules and start making use of syllables (Coltheart et al., 2001).

Even though word reading automaticity is necessary for reading fluency and reading comprehension in early primary school (e.g., Ehri, 2005; Miller & Schwanenflugel, 2008), it is not sufficient (Kuhn et al., 2010). Prosody, which "refers to reading with expression" (Rasinski et al., 2009, p. 351), is also considered by many reading researchers to be a key component of reading fluency (e.g., Dowhower, 1991; Hudson et al., 2009). Several studies have found a direct positive relationship between prosody and reading comprehension (e.g., Benjamin & Schwanenflugel, 2010; Rasinski et al., 2009). As a result, reading prosody may deserve closer attention than it currently receives.

According to Kuhn et al. (2010; see also Couper-Kuhlen, 1986), four well-established prosodic features describe the quality of prosody as variations in a) *pitch*, b) *duration*, c) *stress*, and d) *pausing*. *Pitch*, which is also referred to as intonation or fundamental frequency, could be described as the rate of vibration of a speaker's vocal folds (Dowhower, 1991). Pitch appears to depend on a variety of factors such as the language or even the dialect of the same language (e.g., Mennen et al., 2014). In a cross-language comparison, Mennen et al. (2012) found significant differences in pitch span between English and German female monolingual speakers. English female speakers had a higher pitch at the start of a phrase and showed more pitch variations than German female speakers (see also Scharff-Rethfeldt et al., 2008). Given that no apparent organic or physiological differences existed between the two groups, pitch changes could be traced back to the spoken language (Mennen et al., 2012, 2014).

Pitch may also vary depending on a speaker's gender, with men having lower fundamental frequencies than women (e.g., Hillenbrand & Clark, 2009; Lee et al., 1999; Titze, 1989). Lee et al. (1999) found in their study with children speaking American English that these differences begin to manifest around the age of 11 and are fully established by the age of 15. Finally, comparisons between skilled and struggling readers in primary school have shown that skilled readers have a larger pitch range and more appropriate pitch falls and rises than struggling readers (e.g., English: Benjamin et al., 2013; Spanish: Álvarez-Cañizo et al., 2015).

The second prosodic feature, *duration,* refers to the pronunciation length of vowels and syllables and is usually measured in milliseconds (Kuhn et al., 2010). Duration may depend on a variety of factors. For example, vowels in stressed words have longer pronunciations than vowels in unstressed words (Temperley, 2009). A vowel's pronunciation can even be longer during the phrase-final lengthening, which is the lengthening of the vowel at the final position of a phrase (e.g., Dowhower, 1987, 1991; Edwards et al., 1991; Turk & Shattuck-Hufnagel, 2007). Research has shown that appropriate phrase-final lengthening is a good indicator of readers chunking the reading material (Cooper & Cooper, 1980; Dowhower, 1991), which enhances reading comprehension (Stevens, 1981). Nevertheless, when comparing pronunciations between readers, their individual speaking rate should also be assessed. Faster readers have a faster speaking rate and therefore shorter segment durations than slower readers (Kuhn et al., 2010).

The third prosodic feature, *stress,* refers to the phonetical accentuation with which a vowel or a word is read and is usually measured in dB (Ktori et al., 2018). Stressing a syllable is particularly important because it aids in the distinction of words at the word and sentence level (Ktori et al., 2018). For example, stressing the first or second syllable in English could indicate whether the word is a noun or a verb (e.g., the noun *per*mit versus the verb per*mit*; Himmelmann & Ladd, 2008). Rhythmic patterns and rules may differ between languages (Kuhn et al., 2010). In some languages, such as Greek and French, special diacritics on written words denote the syllables that should be stressed when reading a polysyllabic word (Protopapas, 2006), but most written systems lack such diacritics and stress might be determined by other rules such as morphemes (Rastle & Coltheart, 2000; see also Ktori et al., 2018). Finally, comparisons between English skilled and struggling readers in primary school showed differences in the frequency of stresses. Skilled readers insert a stress every 4.7 words, whereas struggling readers insert a stress in almost every word (Clay & Imlach, 1971; Dowhower, 1991).

The last important prosodic feature, *pausing*, refers to the intra-sentential and inter-sentential pauses that occur during reading and is indicated by a silence in the signal (Kuhn et al., 2010). According to Lalain et al. (2016, as cited in Godde et al., 2020), pauses can be distinguished into three types: breath pauses, syntactic pauses, and hesitation pauses. Breath pauses are indispensable for air intake and may also be used as discourse markers (Bailly & Gouvernayre, 2012), whereas syntactic pauses support sentence parsing and reading comprehension, and hesitation pauses indicate cognitive activity and are associated with decoding (Godde et al., 2020). Most studies focus on the duration, frequency, and appropriateness of the pauses (e.g., English: Schwanenflugel et al., 2004; French: Godde et al., 2022), which are determined to a large extent by the grammatical structure of the text (Bailly & Gouvernayre, 2012; Miller & Schwanenflugel, 2008). In addition, the duration of pauses may also vary depending on the spoken language. For example, Italian has a shorter average pause duration than English, German, and French, whereas Spanish has a longer average pause duration than these three languages (Campione & Veronis, 2002). Several findings with primary school children also suggest that skilled readers take fewer and shorter pauses (e.g., Miller & Schwanenflugel, 2006; Schwanenflugel et al., 2004), whereas less skilled readers make more inappropriate pauses (e.g., Álvarez-Cañizo et al., 2015; Miller & Schwanenflugel, 2008).

In sum, despite differences between the languages in the contribution of the prosodic features to reading, some similarities can be observed when comparing skilled readers to struggling readers. Skilled readers in primary school show larger pitch ranges and shorter pause durations, more appropriate pitch variations, lengthening and stresses, and fewer pauses and stresses compared to struggling readers.

Present Methods for Assessing Prosody

Reading prosody rating scales, spectrographic analyses, and prosodic transcription are presently the most common methods used for measuring and analysing prosodic qualities (e.g., Kuhn et al., 2010; Wolters et al., 2022). However, as it is the case with every research method, they come with advantages and disadvantages. In reading prosody rating scales, human raters make subjective judgements on how readers perform according to certain criteria rubrics (Morrison & Wilcox, 2020). Several reading prosody rating scales exist, each one focusing on different aspects (e.g., Benjamin et al., 2013; Pinnell et al., 1995; Rasinski, 2004). According to Wolters et al. (2022), the most widely used reading prosody scales are the NAEP oral reading fluency scale (Daane et al., 2005; Pinnell et al., 1995) and the Multidimensional Fluency Scale (Rasinksi, 2004). The NAEP is a 4-point holistic rubric that measures phrasing, adherence to the author's syntax, and expressiveness (Daane et al., 2005; Pinnell et al., 1995). The Multidimensional Fluency Scale is a 16-point analytic rubric, with four dimensions and four criteria per dimension. Readers are rated in the dimensions of expression and volume, phrasing, smoothness, and pace. The main advantage of these rating scales is that they are convenient tools for teachers and can be used in the classroom (Miller & Schwanenflugel, 2006). However, some of these scales require the raters to be trained before they can use them and rating a single reading session can be time consuming.

In spectrographic analysis, sound waves are transformed into a visual representation which is known as a spectrogram (Schwanenflugel et al., 2004). Many features, such as stress, pitch, and pause lengths, can be extracted using this representation (Molholt, 1990; Schwanenflugel et al., 2004). Advances in technology and new software has made spectrographic analysis easier, and it allows for a more precise assessment of prosodic features (e.g., Praat by Boersma & Weenink, 2021), but it is more complex to use than reading rating scales. The required technical skills often exceed those of teachers and reading specialists (Kuhn & Schwanenflugel, 2010). Furthermore, even though spectrograms are a powerful tool, researchers must still examine spectrograms to discern patterns in the signal (Molholt, 1990).

In prosodic transcriptions, transcribers identify and annotate all prosodic cues by employing a transcription system while listening to an audio file. There are several well-known transcription systems for prosody, for example, Tones and Break Indices (ToBi, Silverman et al. 1992) for American English, the Gesprächsanalytisches Transkriptionssystem (conversation-analytic transcription system, GAT, Selting et al., 1998) for German, and the International Transcription System for Intonation (INTSINT, Hirst, 1991) for cross-linguistic comparisons (Bressem, 2013). These transcription systems divide utterances into prosodically marked segments and represent prosodic aspects such as changes in pitch, accentuation, and lengthening (Bressem, 2013). However, prosodic annotation can be rather challenging. Transcription of a signal not only requires much time, but it also requires the transcribers to be familiar with the corresponding transcription system to avoid errors.

The Present Study

The primary aim of the present study was to investigate potential differences in prosodic features and prosodic patterns between skilled and struggling German readers in primary school by applying the non-linear analysis tool RQA (see Wallot, 2017, for an extensive tutorial). Although conventional methods, such as prosodic transcription and spectrograms, allow for the investigation of this research question, these methods have limitations. The advantage of RQA over these methods is the possibility to quantify the complexity of a time-series such as the extent that a signal is repetitive, noisy, or stationary. This quantitative method produces various metrics that accurately and objectively represent a signal's patterns (Wallot, 2017) without the need to manually search for them in a spectrogram. Thus, with this robust semi-automated process, identifying patterns in a signal is faster and less error prone. We expected that signals of skilled German readers would be more structured and stable and would show a less complex and chaotic pattern in all prosodic features with less and shorter pauses and stresses than the signals of struggling German readers, which is analogous to findings from reading research in which the more structured word reading times and eye movements are associated with higher reading skills and comprehension (Tschense & Wallot, 2022; Wallot et al., 2014).

A secondary aim was to investigate whether the metrics obtained with RQA, such as signal patterns and complexity, could better distinguish between skilled and struggling readers than metrics derived from prosodic transcription. By prosodic transcription, transcribers must listen to each signal carefully to be able to identify and annotate each prosodic cue. Thus, the method yields data about the frequency of prosodic cue occurrences but no information on prosodic patterns. RQA, however, identifies and quantifies prosodic patterns that cannot be easily perceived with the human eye, but whether prosodic patterns can distinguish between skilled and struggling readers better than the prosodic cues obtained with prosodic transcription is still unclear. RQA is strongly data-driven and allows for a semi-automated pattern detection, whereas prosodic transcription is more theory-driven and requires manual coding of each piece of information. Thus, the importance of RQA metrics regarding the classification to skilled or struggling reader compared to prosodic cues obtained with prosodic transcription should also be investigated.

Method

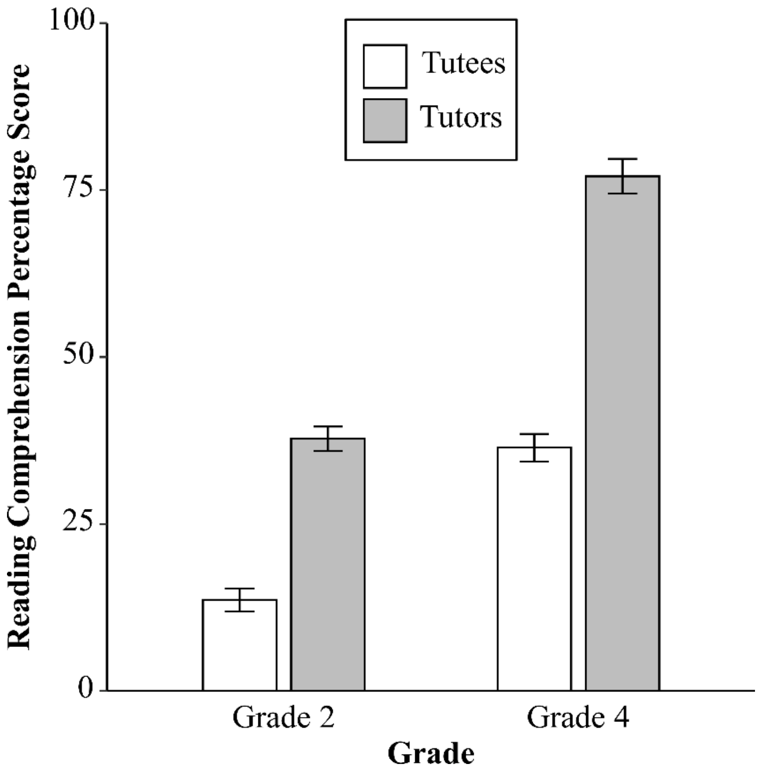
Participants, Procedure and Design

Oral reading data were collected in a subsample of a longitudinal study that investigated differential effects of three types of reading trainings in peer-tutored learning settings (Müller et al., 2015). A subsample of 67 children in Grade 2 from seven schools (32 girls and 35 boys, aged from 7.35 to 9.19 years, *M* = 8.04 years, *SD* = 0.34) and 69 children in Grade 4 from eight schools (38 girls and 31 boys, aged from 9.20 to 12.02 years, *M* = 10.14 years, *SD* = 0.57) were selected to participate in the present study. Half of the children per grade received a reading fluency training, the other half was in the control condition and received a training of visuospatial working memory. All parents gave a written informed consent for the participation of their children. For Grade 2, parents of 49 children reported that their children’s first language was German, parents of 12 children reported that the first language of their children was another language than German and parents of six children did not share any information about the first language. For Grade 4, parents of 43 children reported that their children's first language was German, parents of 14 children reported that the first language of their children is other than German and parents of 12 children did not share any information regarding the first language.

The study was based on an experimental pre-/post-test design with randomization at the class level and was conducted in the urban area of Kassel, Germany. Participants were first screened with ELFE 1–6 (Lenhard & Schneider, 2006) for reading comprehension (see Figure 1). Five children with the lowest scores (hereafter referred to as tutees) and five children with the highest scores (hereafter referred to as tutors) in each class were chosen for the training. To hold skill differences between tutees and tutors constant, the allocation of the children in pairs was based on their score ranking. The highest-scoring tutee was paired with the highest-scoring tutor, the second-best tutee with the second-best tutor, and so forth. In Grade 2, this procedure resulted in 33 tutees and 34 tutors, and in Grade 4, it resulted in 35 tutees and 34 tutors. The proportion of girls and boys in the tutees and tutors groups was approximately equal for Grade 2, $\chi^2(1) = 0.02$, $p = .90$, and Grade 4, $\chi^2(1) = 0.74$, $p = .39$.

Figure 1

Reading Comprehension Percentage Scores for Tutees and Tutors by Grade



Note. Average reading comprehension percentage scores of struggling readers (tutees) and skilled readers (tutors) are shown for each grade. Reading comprehension percentage scores of the sum of correct answers in ELFE 1-6 (Lenhard & Schneider, 2006). Error bars show standard errors.

Measures

Reading Comprehension

Reading comprehension skill was assessed with the computer version of the subtest text comprehension of ELFE 1–6 (Lenhard & Schneider, 2006). The test is a standardized reading comprehension test and is widely used in Germany. It consists of 13 short narrative and expository texts (two to five sentences) with one to three single-choice questions each consisting of four items. The test score is the sum of correct responses achieved within 7 min.

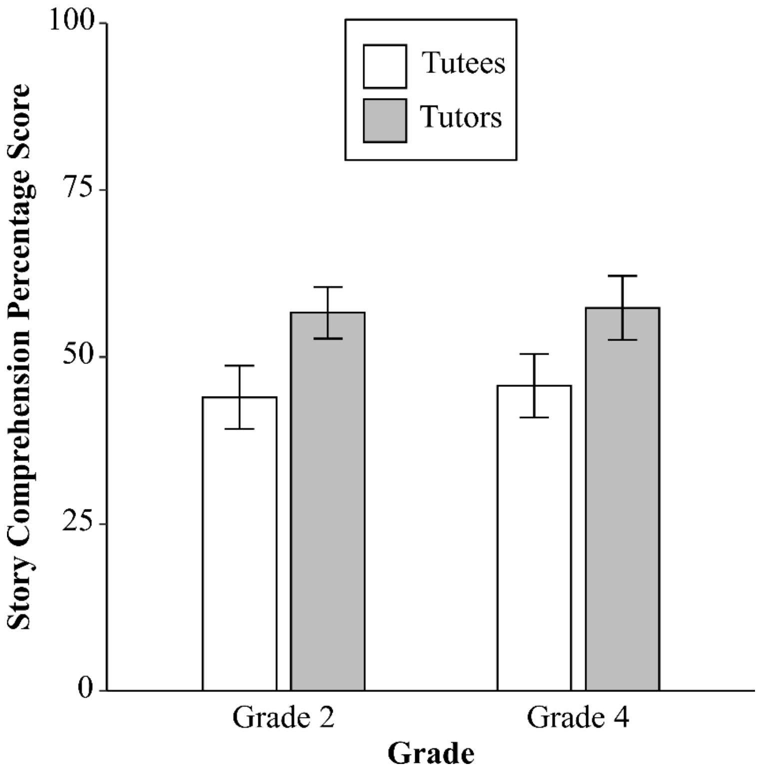
Prosody Measures

Data Collection. The training consisted of 25 sessions, each lasting 45 min and occurring in addition to regular school classes twice a week. At the 23rd training session children read a story aloud individually while they were being recorded (this issue will be addressed in the Discussion). At the end of the session, children were required to answer two open-ended questions on the story they had just read (see Figure 2). The story for the second graders was 66 words long (9 sentences) and the story for the fourth graders was 108 words long (12 sentences). The LIX readability measure of the passage for the second graders was 20.97, and of the passage for the fourth graders was 22.89. LIX is a measure of text difficulty (Bamberger & Rabin, 1984; Björnsson, 1968). The measure is computed by summing the average length of the sentences in a text and the percentage of long words with more than six letters. In German, a score of less than 25 indicates that the text has a low complexity and is appropriate for first graders (Bamberger & Rabin, 1984).

Each school had a quiet room where the children read the story, with only the student assistant and the child present. The story was written on a DIN A4 sheet of paper with the font style Calibri, font size 18, and 1.15 line spacing. During the session, children wore a headset that was connected to a laptop. The audio was digitally recorded using the Audacity software (Version 1.2.6; Audacity Team, 2006).

Figure 2

Story Comprehension Percentage Scores for Tutees and Tutors by Grade



Note. Average story comprehension percentage scores of struggling readers (tutees) and skilled readers (tutors) are shown for each grade. Story comprehension percentage scores of the sum of correct answers in the story open-ended questions. Error bars show standard errors.

Data Processing for RQA. All signals were processed via a custom script in the program MATLAB (Version R2017a; The Math Works, Inc., 2017). First, in an attempt to reduce the noise in each signal, a linear smoothing with equal weight in windows of a 50th of 44100 Hz was conducted. Afterwards, boundaries of pronunciation onsets and offsets using an amplitude criterion (see MATLAB script in online supplement) were detected and a cut-off criterion of 50 ms was used to identify and remove all pronunciation periods with a shorter duration. Hence, sounds such as lip smacking were removed from the signal. Finally, all remaining pronunciation periods were concatenated to have a continuous signal. This signal was partitioned into the pauses and the pronunciation periods to investigate them separately. Pitch was extracted with an algorithm based on the Voice Analysis Toolbox in MATLAB of Tsanas (2012). All time series were *z*-standardized, and the phase space was normalized with the help of the Euclidean distance.

Transcription and Coding. The lengthening within words, final pitch movements of intonation at the end of each phrase, pauses between and within words of a phrase, pauses between phrases, and accentuation were transcribed with the transcription editing software EXMARaLDA Partitur Editor (Version 1.4.4.; Schmidt & Wörner, 2009) while applying the GAT transcription system (Selting et al., 1998). The GAT transcription system is based on a variety of principles and conventions from various disciplines and provides guidelines for the notation of wording and prosody of spoken interaction (Selting et al., 1998). Based on the GAT transcription, we estimated the number of occurrences for lengthening, the inappropriate pauses within and between words of a phrase, the appropriate pauses between phrases, the accentuations (i.e., syllables that carried a focus accent), and the appropriate final pitch movements of intonation (i.e., falling intonations at the end of a sentence). All prosodic features were log transformed to reduce the skewness and then z-standardized.

Results

Data Analysis

Intercorrelations for the most important variables within the groups of tutors and tutees are summarized in Table 1.

Table 1

Intercorrelations for all Important Variables

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. Group (Tutees vs. tutors)	–	.08	-.10	-.40**	-.12	-.14	-.41**	-.39**	-.46**	-.43**	-.37**	-.26*	-.39**	-.48**	-.21	.06	-.16	-.66**
Prosodic patterns obtained with RQA																		
2. Determinism of amplitude	-.09	–	.14	.00	.46**	.17	-.17	.14	.03	.01	.08	.07	.03	-.02	.07	.12	.22	.14
3. Entropy of amplitude	-.22	.48**	–	.01	.74**	.13	-.07	.09	.13	.16	.05	-.12	-.01	.06	-.08	.05	.14	.06
4. Laminarity of amplitude	-.20	.46**	.01	–	.12	.25*	.03	.17	.18	.20	.23	-.01	.15	.20	-.05	-.08	.22	.35**
5. Maximal line length of amplitude	-.38**	.57**	.81**	.18	–	.09	-.02	.14	.13	.21	.09	-.09	.15	.09	.03	.14	.26*	.25*
6. Recurrence rate of amplitude	-.07	.60**	.23	.55**	.29*	–	-.58**	.01	-.04	-.17	.08	.03	.02	.06	.01	-.02	.16	.21
7. Recurrence interval of amplitude	-.28*	-.28*	.09	-.42**	.14	-.62**	–	.16	.18	.30*	.24	.13	.35**	.29*	.20	-.08	-.05	.24*
8. Determinism of pitch	-.26*	.04	-.01	.05	.15	-.07	.13	–	.46**	.75**	.73**	-.04	.29*	.16	.13	.12	.33**	.29*
9. Entropy of pitch	-.27*	-.18	-.03	.01	.05	-.07	.20	.50**	–	.69**	.35**	.08	.15	.31*	-.13	.12	.09	.29*
10. Maximal line length of pitch	-.48**	.07	.10	.16	.34**	.06	.26*	.67**	.74**	–	.65**	-.02	.32**	.23	-.08	.15	.26*	.33**
11. Recurrence rate of pitch	-.31*	-.03	-.09	.04	.02	-.03	.04	.77**	.46**	.63**	–	.00	.37**	.31*	.19	.07	.40**	.30*
12. Average pause time	-.66**	.10	.04	.25*	.24	-.03	.36**	.39**	.44**	.52**	.40**	–	-.07	.22	.19	-.11	-.18	.40**
13. Recurrence interval of pauses	-.59**	-.00	.01	.00	.23	-.12	.38**	.47**	.34**	.53**	.50**	.49**	–	.39**	.25*	-.04	.17	.46**
Prosodic features obtained with GAT																		
14. Lengthening	-.67**	-.03	.20	.06	.43**	.02	.39**	.28*	.28*	.54**	.26*	.51**	.49**	–	.15	-.22	.06	.48**
15. Final pitch	-.55**	-.17	-.02	.08	.14	-.08	.29*	.38**	.31*	.43**	.42**	.43**	.52**	.46**	–	.02	.19	.25*
16. Accentuation	.42**	.22	-.04	.01	-.11	.07	-.29*	-.48**	-.42**	-.52**	-.50**	-.40**	-.49**	-.36**	-.51**	–	.31**	-.07
17. Appropriate pauses	-.06	.24	-.08	.16	.03	.20	-.07	.08	-.01	.12	.10	.22	.05	.06	-.12	.11	–	.03
18. Inappropriate pauses	-.66**	.06	.16	.20	.39**	.06	.39**	.37**	.38**	.55**	.33**	.66**	.54**	.54**	.57**	-.47**	.12	–

Note. The results for Grade 4 are shown above the diagonal. The results for Grade 2 are shown below the diagonal.

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

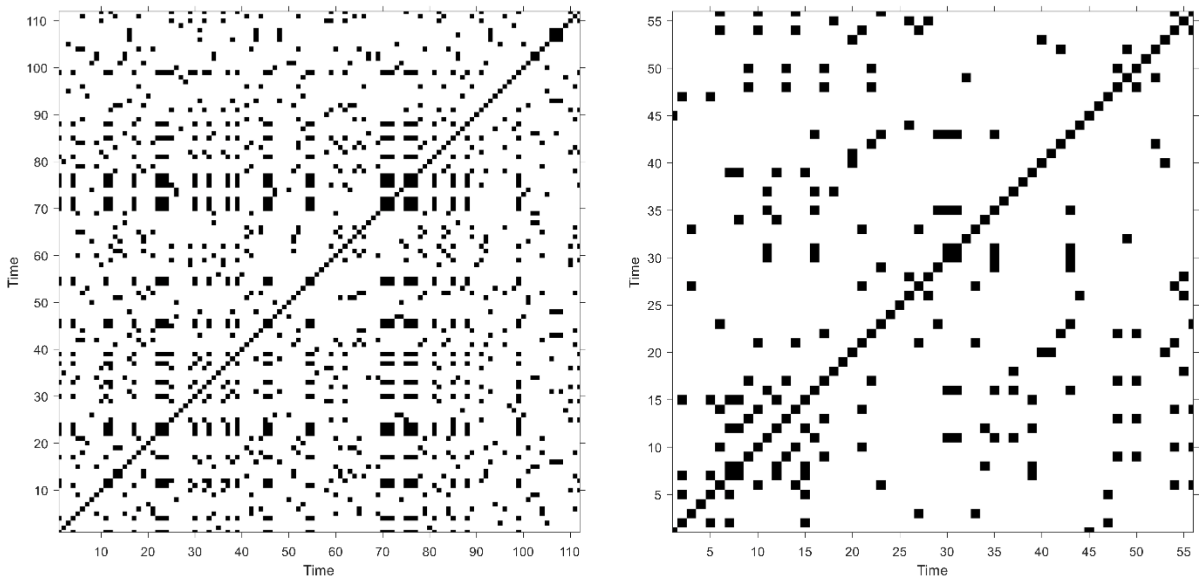
Recurrence Quantification Analysis

To capture the complexity and regularity of reading over time in each signal, we used *Recurrence Quantification Analysis* (RQA; Marwan et al., 2007; Wallot, 2017; Wallot et al., 2014) estimated with the CRP toolbox (Version 5.24 [R34]; Marwan, 2022) in the program MATLAB (Version R2017a; The Math Works, Inc., 2017). RQA is a general nonlinear time-series analysis technique (see also Wallot, 2017). The advantage of RQA compared to other methods is the possibility to quantify the complexity of a time series such as the extent that it is repetitive, noisy, or stationary. This quantification is accomplished through the recurrence plot (Eckmann et al., 1987), which is the visualisation of the phase space trajectory of a dynamical system in a square matrix (see Figure 3). The recurrence plot allows the extraction of a comprehensive set of measures. Four different recurrence plots for each child were generated: One for the amplitude of the signal, one for the duration of signal segments with amplitude above zero, one for the pitch of the signal, and one for the duration of all pauses in the signal. The generated plots resulted in the extraction of 48 different RQA measures (12 variables per plot). Means and medians for each prosodic feature were also estimated and included in the analyses. In the following text, we briefly describe the most important measures of RQA (Marwan & Webber, 2015; Webber & Zbilut, 2005). A complete list of the measures can be found in the online supplement.

Figure 3

Example of Recurrence Plots of Amplitudes of (A) a Struggling Second Grader and (B) a Skilled Fourth Grader

A B



Recurrence rate = 10.62 % | Recurrence rate = 5.71 %

Determinism = 22.73 % | Determinism = 12.50 %

Maximal line length = 3 | Maximal line length = 3

Laminarity = 32.95 % | Laminarity = 10.23 %

Recurrence interval = 7.85 | Recurrence interval = 10.10

Note. Recurrence rate is estimated as the proportion of all recurrence points to all cells (without the main diagonal line); Determinism is estimated as the proportion of all diagonal recurrence points, parallel to the main diagonal, to all recurrence points (without the main diagonal line); Maximal line length is estimated as the longest non-main diagonal line; Laminarity is estimated as the proportion of all vertical recurrence points to all recurrence points; Recurrence interval is estimated by averaging the time needed until a recurrence is occurred.

One of the most common measures of RQA is the *recurrence rate* (Marwan & Webber, 2015). This measure is the proportion of recurrence points to the total number of possible points in the recurrence plot (Wallot, 2017; see also Figure 3). Therefore, it indicates the overall tendency of the signal to repeat itself within a specified radius (i.e., how often a state is revisited). This feature can range from 0% to 100%, with the lowest value indicating zero probability of recurrent points, whereas the highest value indicates that all points in the plot are recurrent.

Another important measure, *percent determinism*, refers to the exact repetition of a sequence over time and is estimated by dividing the recurrence points of the diagonal lines by the total number of recurrence points in the recurrence plot (Wallot, 2017; see also Figure 3). Similar to recurrence rate, the larger the value, the more exact repetitions in the signal. Put simply, percent determinism is a measure of the proportion of data points in the signal that are part of recurring patterns and indicates the predictability of the system (Marwan & Webber, 2015). Recurrence rate and determinism of reading times have been shown to be positively related to reading skill (O'Brien et al., 2014), and determinism was shown to be predictive of reading comprehension and reading speed (Wallot et al., 2014).

The third measure, *maximal line length*, refers to the length of the longest diagonal line in the recurrence plot apart from the main diagonal line (Wallot, 2017; see also Figure 3). The inverse of this feature indicates the divergence of the trajectory segments and the stability of the signal. The diagonal lines in the recurrence plot of a chaotic dynamical system are short, whereas diagonal lines in the recurrence plot of a periodic system are longer.

The fourth measure, *entropy*, is a measure of the complexity of the signal and is computed as the Shannon's entropy of the frequency distribution of the diagonal lines' lengths (Marwan & Webber, 2015; see also Figure 3). In short, entropy indicates the complexity of the patterns in the signal (Fusaroli & Tylén, 2016). The lower the entropy, the less complex and more predictable the system is (Leonardi, 2018), indicating the presence of fewer patterns (Fusaroli & Tylén, 2016).

The fifth measure, *laminarity*, indicates the degree to which a signal is 'trapped' into repeating the same patterns over time (Cox et al., 2016; see also Figure 3). Put it differently, laminarity quantifies the amount of smoothness (Konvalinka et al., 2011). Similar to determinism, laminarity is estimated by dividing the recurrence points of the vertical lines by the total number of recurrence points in the recurrence plot (Marwan et al., 2007). High laminarity indicates a smoother system that remains unchanged or changes slowly and the less vertical structures compared to single recurrence points, the lower the laminarity (Brick et al., 2018; Marwan et al., 2007).

Finally, *recurrence intervals* refer to the average time for the system to revisit a successive point in the signal (see Figure 3). In other words, they reflect the average duration between recurrences and indicate the expected frequency of recurrences. Recurrence intervals are estimated by dividing the entire signal time by the total number of a recurrences. The shorter the recurrence interval, the more frequently a recurrence is likely to occur.

Repeated k-fold Cross-Validation

Given the large number of variables obtained through RQA, we proceeded with a 100x repeated 5-fold cross-validation separately for each grade to identify the most important predictors for children's classification as skilled and struggling readers. To this end, we used the R caret package (Kuhn, 2008, 2020). In k-fold cross-validation, the data are partitioned in *k* subsets (folds), with *k*-1 folds used for the model construction and the one remaining fold used for the model validation (de Rooij & Weeda, 2020). In this study, the data were partitioned into five folds of approximately equal size. Four folds were randomly used to train the models and the one remaining fold was used to validate the final models.

The following models were estimated: linear discrimination analysis, lasso and elastic-net regularized generalized linear model, partial least squares regression, and random forest. Children's status as tutees or tutors (dummy-coded; 0 = tutee; 1 = tutor) was used as the outcome variable. Pairwise correlations were estimated for all variables that were obtained from the recurrence plots. For pairwise correlations above .75, the variable with the largest mean absolute correlation to all other variables was excluded from all further analyses to avoid multicollinearity and increase the interpretability of the model. The remaining variables were *z*-standardized and entered as predictors in the models. The estimated parameters of the five models were saved at the end of each calculation. These parameters were then used on the test data to obtain the prediction errors and the area under the curve. To increase the accuracy of the estimation and assess how many times a particular model had a smaller prediction error than the other models, this procedure was repeated 100 times. After identifying the model with the largest area under the curve and the lowest Brier score, the built-in evaluation function of the caret package (Kuhn, 2008, 2020) was used to generate a list with the predictors ranked according to their contribution to the model. The forward selection approach was used to import these predictors into a logistic regression with the children's status (dummy-coded; 0 = tutee; 1 = tutor) as dependent variable. In other words, the most important predictor obtained from the 100x repeated 5-fold cross-validation was entered first in the model. The other predictors on the list were added in the model according to their rank, until the model's BIC was not reduced further. Furthermore, the nested models were compared with a likelihood-ratio test to confirm that the selected model had the best fit.

Important Variables for Distinguishing Between Tutees and Tutors in Grade 2

Data of four children in Grade 2 were excluded from all analyses: Three children had an average pause time of 3 *SD* above the sample average, which caused a complete separation in the logistic model and one child had at least one missing value, which led to difficulties with the convergence of the models. For the data of Grade 2, the 100x repeated 5-fold cross-validation showed that the model random forest had the largest area under the curve and the lowest Brier score of all four models (see Table 2). The summarized confusion matrix for the random forest model validated on the test fold is shown in Table 3. The forward stepwise selection for the most significant predictors resulted in the selection of two variables (see Table 4). The average time of pauses during reading, $B = -2.22$, $SE = 0.62$, $p < .001$, $OR = 0.11$, 95% CI [0.03, 0.32], and the recurrence interval of pauses during reading, $B = -1.63$, $SE = 0.57$, $p = .004$, OR = 0.20, 95% CI [0.05, 0.53], were the most important predictors for children's classification into tutees and tutors in Grade 2. That is, for a one standard deviation increase in the average time of pauses the odds for being classified as tutors are 89% lower than for tutees, whereas for a one standard deviation increase in the recurrence interval of pauses the odds are 80% lower.

Table 2

Comparison of Fit Indices in Models Estimated with 100x Repeated 5-fold Cross-Validation and Validated on the Test Fold with a Probability Cut-Off of 0.5

Model	BS	AUC	Sensitivity	Specificity	Accuracy	F1 score
Grade 2						
Linear discrimination analysis	.28	.70	.67	.78	.72	.71
Lasso and elastic-net regularized generalized linear model	.19	.78	.89	.67	.78	.80
Partial least squares	.19	.79	.56	.78	.67	.63
Random forest	.19	.81	.67	.78	.72	.71
Grade 4						
Linear discrimination analysis	.51	.49	.33	.60	.47	.34
Lasso and elastic-net regularized generalized linear model	.17	.84	.44	1.00	.74	.62
Partial least squares regression	.17	.86	.44	1.00	.74	.62
Random forest	.18	.82	.67	.90	.79	.75

Note. BS = Brier score; AUC = area under the curve; Sensitivity = proportion of correct classified tutors; Specificity = proportion of correct classified tutees; Accuracy = accuracy of the model's prediction on the test fold; F1 score = mean between precision and recall.

Table 3

Confusion Matrix for Grade 2 of the Random Forest Model Validated on the Test Fold

Predicted values	Actual values	
	Tutors	Tutees
Tutors	6	2
Tutees	3	7

Table 4

Model Parameters and Goodness of Fit for Classification of Tutees and Tutors by Prosodic Features in Grade 2

Variable	Model 1			Model 2			Model 3		
	B	*SE*	*OR*	*B*	*SE*	*OR*	*B*	*SE*	*OR*
Constant	-0.06	0.36	0.94	0.14	0.40	1.16	0.19	0.42	1.21
Average time of pauses	-2.46***	0.60	0.09	-2.22***	0.62	0.11	-2.20***	0.62	0.11
Recurrence interval of pauses				-1.63**	0.57	0.20	-1.57**	0.57	0.21
Recurrence interval of speech rates							-0.20	0.37	0.82
Goodness of fit									
Deviance	51.57			40.14			39.85		
Likelihood ratio test	35.63***			11.43***			0.30		
Wald test	17.06***			8.29**			0.29		
BSS	.46			.62			.62		
Nagelkerkes R^2	.58			.70			.71		
Δdf				1			1		
BIC	59.85			52.57			56.42		

Note. *B* = unstandardized regression coefficients; *SE* = standard error; *OR* = odds ratio; BSS = Brier skill score against the null model; Nagelkerkes R^2 = multiple correlation squared; BIC = Bayesian information criterion.

$^{**}p < .01$. $^{***}p < .001$.

Important Variables for Distinguishing Between Tutees and Tutors in Grade 4

Four children in Grade 4 had at least one missing value, which led to difficulties with the convergence of the models and were thus excluded from all analyses. For Grade 4, the 100x repeated 5-fold cross-validation showed that the partial least squares model had the largest area under the curve and the lowest Brier score of all four models (see Table 2). The summarized confusion matrix for the partial least squares model validated by the test fold is shown in Table 5. The forward stepwise selection for the most significant predictors resulted in the selection of three variables (see Table 6). The recurrence interval of amplitude, $B = -1.01$, $SE = 0.39$, $p = .009$, $OR = 0.37$, 95% CI [0.16, 0.73], the entropy of pitch, $B = -1.04$, $SE = 0.39$, $p = .007$, $OR = 0.35$, 95% CI [0.14, 0.70], and the laminarity of amplitude, $B = -0.96$, $SE = 0.36$, $p = .008$, $OR = 0.38$, 95% CI [0.18, 0.74], were the most significant predictors for children's classification into tutees and tutors in Grade 4. For a one standard deviation increase in the recurrence interval of amplitude, entropy of pitch, and laminarity of amplitude the odds for being classified as tutors were 63%, 65%, and 62% lower than for tutees, respectively.

Table 5

Confusion Matrix for Grade 4 of the Partial Least Squares Regression Validated on the Test Fold

Predicted values	Actual values	
	Tutors	Tutees
Tutors	4	0
Tutees	5	10

GAT – Transcription system

For the variables with the coding schema GAT, logistic regressions with children's status (dummy-coded; 0 = tutee, 1 = tutor) as dependent variable were estimated for each grade (see Table 7). The lengthening occurrences, *B* = -2.20, *SE* = 0.86, *p* = .010, *OR* = 0.11, 95% CI [0.01, 0.44], and the sum of inappropriate pauses within and between words of a phrase, *B* = -2.54, *SE* = 0.94, *p* = .007, *OR* = 0.08, 95% CI [0.01, 0.38], were significant predictors for children's classification into tutees and tutors in Grade 2. For a one standard deviation increase in the lengthening occurrences the odds for being classified as tutors were 89% lower than for tutees, whereas for one standard deviation increase in the inappropriate pauses the odds were 92% lower.

In Grade 4, only the sum of inappropriate pauses within and between words of a phrase, *B* = -2.82, *SE* = 0.85, *p* < .001, *OR* = 0.06, 95% CI [0.01, 0.24], was a significant predictor for children's classification into tutees and tutors. For a one standard deviation increase in the inappropriate pauses, the odds for being classified as tutors was 94% lower than for tutees. Figure 4 for Grade 2 and Figure 5 for Grade 4 show the estimated probabilities for the models with GAT variables versus the estimated probabilities for the models with RQA variables.

Finally, for each grade, the most important predictors of both RQA and GAT were selected to estimate new models by entering them simultaneously (see Table 8). In Grade 2 only the lengthening occurrences, *B* = -1.50, *SE* = 0.77, *p* = .030, *OR* = 0.18, 95% CI [0.03, 0.65] was a significant predictor for children's classification into tutees and tutors, whereas in Grade 4 the entropy of pitch, *B* = -0.80, *SE* = 0.40, *p* = .048, *OR* = 0.45, 95% CI [0.18, 0.93] and the sum of inappropriate pauses within and between words of a phrase, *B* = -2.57, *SE* = 0.83, *p* = .002, *OR* = 0.08, 95% CI [0.01, 0.29] were the most significant predictors.

Table 6

Model Parameters and Goodness of Fit for Classification of Tutees and Tutors by Prosodic Features in Grade 4

Variable	Model 1			Model 2			Model 3			Model 4		
	B	*SE*	*OR*	*B*	*SE*	*OR*	*B*	*SE*	*OR*	*B*	*SE*	*OR*
Constant	-0.18	0.27	0.84	-0.30	0.32	0.74	-0.25	0.34	0.78	-0.23*	0.34	0.79
Recurrence interval of amplitude	-.90**	0.31	0.40	-0.86	0.34	0.43	-1.01**	0.39	0.37	-0.96**	0.39	0.38
Entropy of pitch				-1.17	0.41	0.31	-1.04**	0.39	0.35	-1.07**	0.40	0.34
Laminarity of amplitude							-0.96**	0.36	0.38	-0.97	0.37	0.38
Determinism of amplitude										0.24	0.36	1.27
					Goodness of fit							
Deviance	79.18			66.75			58.37			57.91		
Likelihood ratio test	10.55**			12.43***			8.38**			0.46		
Wald test	8.67**			8.14**			7.01**			0.45		
BSS	.16			.32			.41			.40		
Nagelkerkes R^2	.20			.40			.51			.52		
Δdf				1			1			1		
BIC	87.53			79.27			75.07			78.78		

Note. *B* = unstandardized regression coefficients; *SE* = standard error; *OR* = odds ratio; BSS = Brier skill score against the null model; Nagelkerkes R^2 = multiple correlation squared; BIC = Bayesian information criterion.

**p* < .05. ***p* < .01. ****p* < .001.

Table 7

Model Parameters and Goodness of Fit for Classification of Tutees and Tutors by Prosodic Features Coded with the GAT Transcription System in Grade 2 and Grade 4

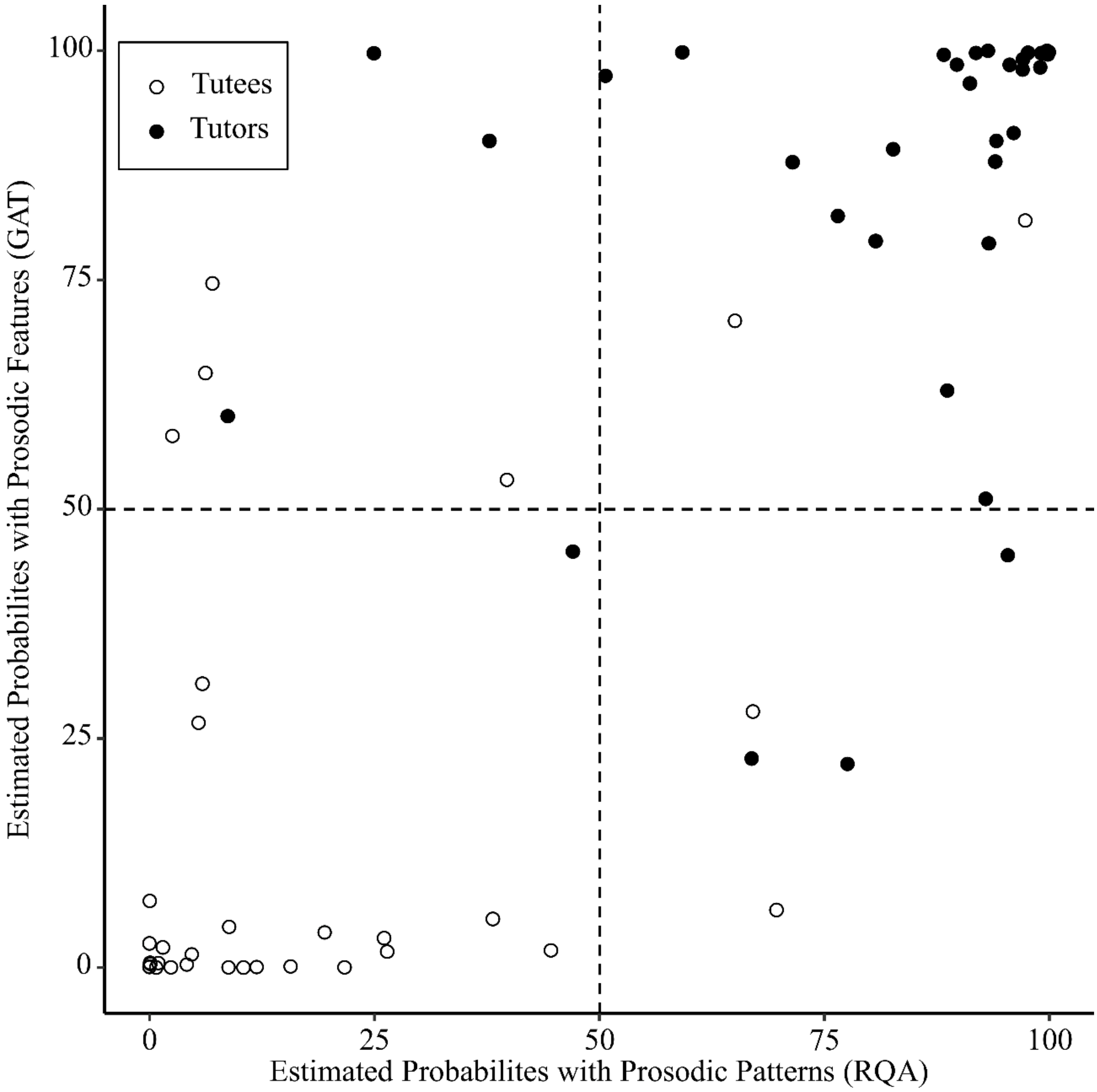
Variable	Grade 2			Grade 4		
	B	*SE*	*OR*	*B*	*SE*	*OR*
Constant	-0.22	0.49	0.81	0.17	0.41	1.18
Final pitch	-1.10	0.61	0.33	0.11	0.47	1.12
Lengthening	-2.20*	0.86	0.11	-0.93	0.51	0.39
Accentuation	-0.23	0.58	0.78	-0.10	0.42	0.91
Appropriate pauses	-0.12	0.42	0.89	-0.38	0.40	0.68
Inappropriate pauses	-2.54**	0.94	0.08	-2.82***	0.85	0.06
Goodness of fit						
Deviance	32.23			48.22		
Likelihood ratio test	60.63***			47.42***		
Wald test	12.66*			15.38**		
BSS	.67			.57		
Nagelkerkes R^2	.79			.66		
BIC	57.46			73.62		

Note. *B* = unstandardized regression coefficients; *SE* = standard error; *OR* = odds ratio; BSS = Brier skill score against the null model; Nagelkerkes R^2 = multiple correlation squared; BIC = Bayesian information criterion.

*p < .05. **p < .01. ***p < .001.

Figure 4

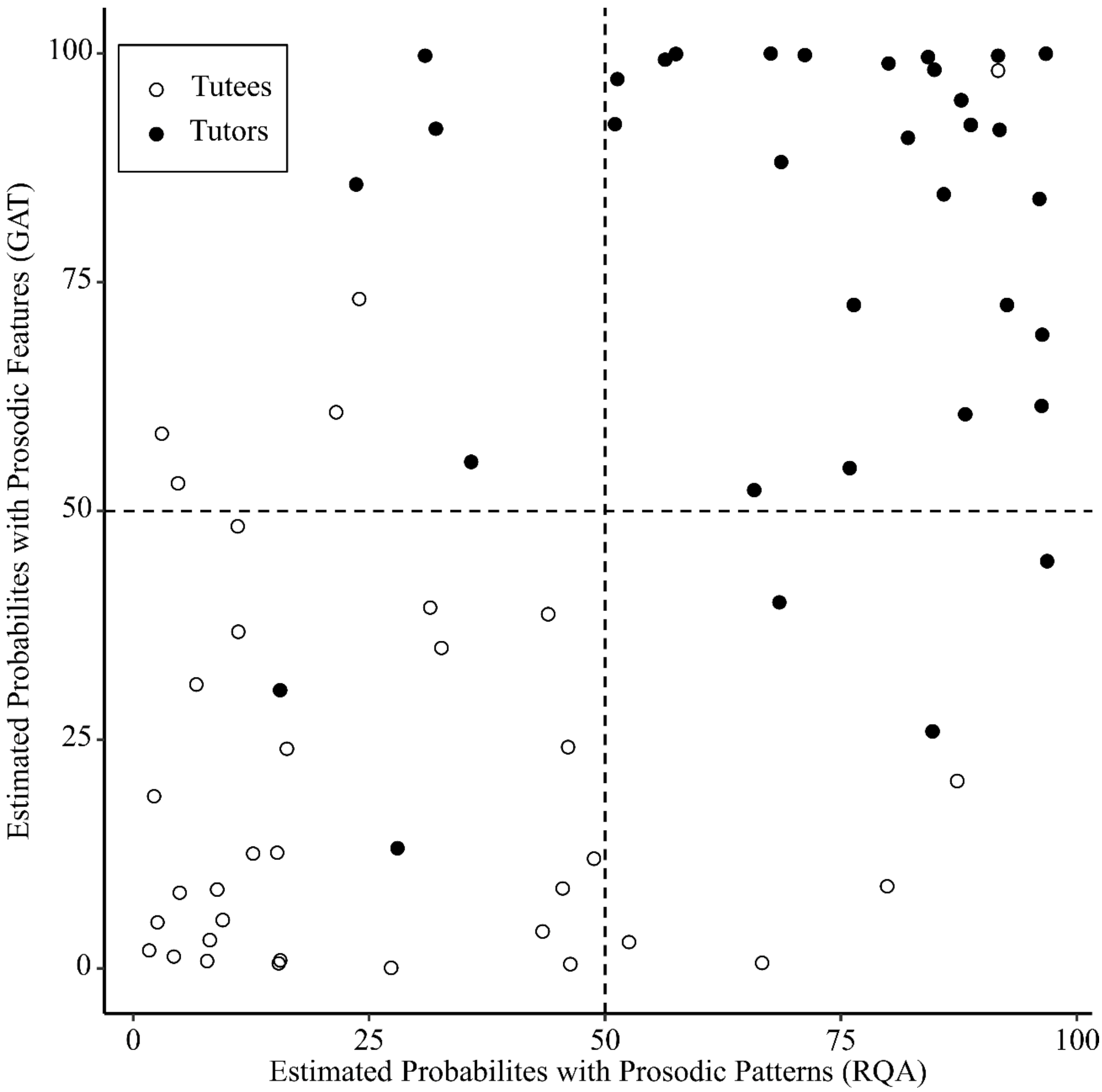
Scatter Plot of the Estimated Probabilities from the Model with Prosodic Features Versus the Estimated Probabilities for the Model with Prosodic Patterns in Grade 2



Note. For each child in each model, logits were estimated and then transformed into probabilities. Estimated probabilities are presented as percentages. Classified struggling readers (tutees) from both models are shown in the bottom-left quadrant, whereas classified skilled readers (tutors) from both models are shown in the top-right quadrant.

Figure 5

Scatter Plot of the Estimated Probabilities from the Model with Prosodic Features Versus the Estimated Probabilities for the Model with Prosodic Patterns in Grade 4



Note. For each child in each model, logits were estimated and then transformed into probabilities. Estimated probabilities are presented as percentages. Classified struggling readers (tutees) from both models are shown in the bottom-left quadrant, whereas classified skilled readers (tutors) from both models are shown in the top-right quadrant.

Table 8

Model Parameters and Goodness of Fit for Classification of Tutees and Tutors by Combination of Prosodic Patterns Obtained with RQA and Prosodic Features Coded with the GAT Transcription System in Grade 2 and Grade 4

Variable	Grade 2			Grade 4		
	B	*SE*	*OR*	*B*	*SE*	*OR*
Constant	0.14	0.50	1.15	0.19	0.43	1.21
Average time of pauses	-1.15	0.75	0.32	–	–	–
Recurrence interval of pauses	-1.50	0.77	0.22	–	–	–
Lengthening	-1.73*	0.80	0.18	–	–	–
Inappropriate pauses	-1.85	1.00	0.16	-2.57**	0.83	0.08
Recurrence interval of amplitude	–	–	–	-0.99	0.52	0.37
Entropy of pitch	–	–	–	-0.80*	0.40	0.45
Laminarity of amplitude	–	–	–	-0.88	0.51	0.41
Goodness of fit						
Deviance	27.42			37.92		
Likelihood ratio test	59.78***			57.72***		
Wald test	11.56*			17.49**		
BSS	0.77			0.71		
Nagelkerkes R^2	.82			.76		
BIC	48.13			59.09		

Note. *B* = unstandardized regression coefficients; *SE* = standard error; *OR* = odds ratio; BSS = Brier skill score against the null model; Nagelkerkes R^2 = multiple correlation squared; BIC = Bayesian information criterion.

**p* < .05. ***p* < .01. ****p* < .001.

Discussion

The primary goals of this study were to identify prosodic patterns in oral reading that differentiate struggling and skilled German readers using RQA measures, which have been successfully used to predict comprehension scores from reading time measures (Wallot et al., 2014), and to compare the discrimination of these patterns with prosodic features obtained using the transcription method GAT. For this purpose, oral readings of a passage by German second and fourth graders were recorded and analysed.

The results revealed that in Grade 2, the average pause time, and the recurrence interval between pauses (estimated using RQA) were the most important predictors. Pausing appears to have an essential function for German second graders, with longer pauses being a hallmark of struggling second graders. Thereby the predictor "pauses" includes intra- and inter-sentential pauses, that is, pauses within and between words independent of the syntactic phrases. Similar results have also been reported in the study of Miller and Schwanenflugel (2006; see also Godde et al., 2022 for French; Schwanenflugel et al., 2004) with American struggling readers. According to Miller and Schwanenflugel, longer pauses in younger readers might be related to word decoding issues. Accordingly, lengthier pauses may provide struggling readers with the necessary time they need to decode the next word in a phrase (Godde et al., 2022). One possible explanation for the lengthier intervals between recurring pauses could be that struggling readers stumble when reading a word or sentence, resulting in longer reading periods between pauses. This explanation, however, fails to account for the total number of pauses. According to the results of GAT, struggling readers still made 19 inappropriate pauses on average, compared to 5.71 inappropriate pauses on average for skilled readers.

In Grade 4, an apparent shift occurs in the impact of prosodic patterns on reading comprehension, with pausing patterns being less prevalent than in Grade 2 and less relevant for distinguishing skilled from struggling readers. The most important predictors were the recurrence interval of amplitudes, entropy of pitch, and laminarity of amplitude. Thus, longer intervals between recurring amplitudes, a more complex signal with more pitch patterns, and a slower rate of amplitude change appear to be characteristics of struggling fourth graders who learn to read in the German language. The longer intervals between recurring amplitudes indicates that when skilled readers read aloud, less time passes until recurring amplitudes of pronunciations emerge in their signal. This finding is not surprising considering that skilled readers read fluently and have shorter segment durations than struggling readers (Kuhn et al., 2010). Furthermore, the higher entropy of pitch suggests the presence of a diverse range of pitch patterns in German struggling readers, that is, more pitch repetitions that consist of multiple occurrences in a word or a whole phrase than in German skilled readers (Fusaroli & Tylén, 2016). German skilled readers tend to read with consistent pitch adjustments, resulting in a more predictable signal than German struggling readers' rather complex unpredictable signal with little structure. One possible explanation for this finding could be that efficient word recognition frees up cognitive resources for semantic processing of longer segments (phrases and sentences), allowing for more appropriate pitch variation in reading (Kim et al., 2021). Finally, amplitudes of struggling readers changed in a slower pace than those of skilled readers, indicating that they showed more repetitions of amplitudes and lingered in a state for a longer period. This is also consistent with prior research that revealed that skilled readers insert a stress less frequently than struggling readers (Clay & Imlach, 1971; Dowhower, 1991). According to Clay and Imlach (1971), good readers read in syntactic chunks, whereas poor readers read the text as if it was a list of words. Hence, it seems plausible that inserting a stress can result from inefficient word recognition. Readers who are unable to recognize words from memory, according to Torgesen and Hudson (2006), may stumble over several words and make numerous errors while trying to "sound out" the word before recognizing it and moving on to the next one.

The results of the prosodic features obtained with coding schema GAT revealed that the sum of inappropriate pauses could distinguish between skilled and struggling readers in

Grades 2 and 4. Struggling readers made hesitant pauses within words or in the middle of a phrase. Skilled readers making fewer and shorter ungrammatical pauses than struggling readers is also consistent with prior studies (Álvarez-Cañizo et al., 2015; Godde et al., 2020; Schwanenflugel et al., 2004). Inappropriate pauses appear to be the result of decoding issues or the misreading of a sentence's syntax (Godde et al., 2020; Miller & Schwanenflugel, 2008) and have been shown to negatively relate to other aspects of reading fluency and reading comprehension (Benjamin & Schwanenflugel, 2010; Dowhower, 1991). Furthermore, struggling readers in Grade 2 in the current study prolonged syllables within words, which lead to inappropriate word pronunciation. This finding is in line with Lyytinen et al. (1995), who showed that lengthening occurrences within words in struggling Finnish readers occurs more frequently than in skilled readers. One possible explanation for this improper use of vowel duration marking could be that children have difficulty distinguishing between long and short vowel phonemes (Landerl & Reitsma, 2005).

Contrary to our expectations, when models with RQA predictors were compared to models with GAT predictors, models with GAT predictors had a slightly higher Nagelkerke's R^2, higher Brier skill scores and lower BIC (except for Grade 2) than the RQA models. Even though the estimated models using prosodic patterns had a good fit, they still did not perform better than the models estimated with prosodic features. Nevertheless, as can be seen in the upper left and lower right quadrants of Figures 3 and 4, both models misclassified similar numbers of children as skilled or struggling readers, although the children misclassified by the two methods were usually not the same. This inconsistency is particularly interesting, especially for Grade 4, given that correlations between most significant prosodic patterns and prosodic features in this grade were small to medium.

It appears that despite tapping into distinct processes, both methods can distinguish between skilled and struggling readers. This assumption is also substantiated by the results of the models estimated with the combination of prosodic patterns and prosodic features. These models had an overall better fit than the RQA and GAT models, with the difference more pronounced in Grade 4, while misclassifying only six children in Grade 2 and three children in Grade 4. Hence, despite the overlap in certain aspects of these models, they nonetheless complement each other and explain unique variance.

In sum, our findings indicate that German skilled readers in Grade 2 make shorter pauses with shorter intervals, whereas German skilled readers in Grade 4 need less time between recurrent amplitudes, have consistent pitch adjustments, and express fewer matching amplitudes for some period of time. However, even though RQA models provided additional information, they were unable to outperform GAT models in the classification of skilled and struggling readers.

Limitations

Given the scarcity of RQA for prosody data, our findings are encouraging. They must, however, be interpreted with some limitations in mind. First, because of the large number of variables, we needed to conduct a 100x 5-fold cross-validation to deal with the RQA's curse of dimensionality. We reduced this number by employing model comparison approaches after identifying the most important predictors influencing model performance with a feature selection algorithm. Nonetheless, despite the exploratory nature of our analyses, this procedure may alleviate some concerns regarding overfitting. Therefore, we made every effort to document and disclose every stage of our analyses so that they may be utilized as a reference for future studies.

Another limitation of our study is that even though children were separated into tutees and tutors at the start of the study based on their pre-test reading comprehension scores, they had already participated in 23 training sessions by the time of the oral reading. Therefore, notwithstanding the disparities in reading comprehension between the two groups at the post-test, we cannot be certain that the results would have been the same if the children had not participated in the training. In addition, half of the children received a reading fluency training whereas the other half received a training of visuospatial working memory. However, investigating prosodic patterns separately for each training would require a larger sample. Sample sizes in this study were relatively small, which might have also impacted the parameter estimates and lead to a relatively low power.

Conclusion

Despite these limitations, the study demonstrates that using RQA for prosody analysis is a promising approach. If similar results were to be obtained in a longitudinal design identifying prosodic patterns that distinguish between struggling and skilled readers throughout primary school, RQA might prove to be a versatile tool in research on prosody. Confirming our findings would provide experts with another option for a precise semi-automatic assessment of prosodic difficulties, which is far more economic than hand-coding prosodic features with GAT. Finally, the RQA seems to provide additional information about prosody that augments established approaches, which might be useful for advancing the existing theories.

References

Allington, R. L. (1983). Fluency: The neglected reading goal. *The Reading Teacher*, *36*(6), 556–561.

Altani, A., Protopapas, A., Katopodi, K., & Georgiou, G. K. (2020). From individual word recognition to word list and text reading fluency. *Journal of Educational Psychology, 112*(1), 22–39. https://doi.org/10.1037/edu0000359

Álvarez-Cañizo, M., Suárez-Coalla P., & Cuetos F. (2015). The role of reading fluency in children's text comprehension. *Frontiers in Psychology*, *6*, Article 1810. https://doi.org/10.3389/fpsyg.2015.01810

Audacity Team (2006). *Audacity: Free audio editor and recorder* (Version 1.2.6) [Computer software]. Audacity Team. https://audacityteam.org

Bailly, G., & Gouvernayre, C. (2012). Pauses and respiratory markers of the structure of book reading. In *13th annual conference of the international speech communication association 2012 (INTERSPEECH 2012)* (pp. 2218–2221). Curran Associates. https://doi.org/10.21437/Interspeech.2012-591

Bamberger, R., & Rabin, A. T. (1984). New approaches to readability: Austrian research. *The Reading Teacher*, *37*(6), 512–519.

Benjamin, R. G., & Schwanenflugel, P. J. (2010). Text complexity and oral reading prosody in young readers. *Reading Research Quarterly*, *45*(4), 388–404. https://doi.org/10.1598/RRQ.45.4.2

Benjamin, R. G., Schwanenflugel, P. J., Meisinger, E. B., Groff, C., Kuhn, M. R., & Steiner, R. (2013). A spectrographically grounded scale for evaluating reading expressiveness. *Reading Research Quarterly*, *48*(2), 105–133. https://doi.org/10.1002/rrq.43

Björnsson, C. H. (1968). *Läsbarhet* [Readability]. Liber.

Boersma, P., & Weenink, D. (2021). *Praat: Doing phonetics by computer* [Computer program]. Phonetic Sciences. https://www.fon.hum.uva.nl/praat/

Bressem, J. (2013) Transcription systems for gestures, speech, prosody, postures, and gaze. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & S. Tessendorf (Eds.), *Body – language – communication: An international handbook on multimodality in human interaction.* (Vol. 1, pp. 1037–1059). De Gruyter Mouton. https://doi.org/10.1515/9783110261318.1037

Brick, T. R., Gray, A. L., & Staples, A. D. (2018). Recurrence quantification for the analysis of coupled processes in aging. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences*, *73*(1), 134–147. https://doi.org/10.1093/geronb/gbx018

Browne, M. W. (2000). Cross-validation methods. *Journal of Mathematical Psychology, 44*(1), 108–132. https://doi.org/10.1006/jmps.1999.1279

Campione, E., & Véronis, J. (2002). A large-scale multilingual study of silent pause duration. In B. Bel & I. Marlien (Eds.), *Speech prosody 2002: Proceedings: Aix-en-Provence, France, 11-13 April 2002* (pp. 199-202). Laboratoire Parole et Langage.

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest, 19*(1), 5–51. https://doi.org/10.1177/1529100618772271

Clay, M. M., & Imlach, R. H. (1971). Juncture, pitch, and stress as reading behavior variables. *Journal of Verbal Learning and Verbal Behavior, 10*(2), 133–139. https://doi.org/10.1016/S0022-5371(71)80004-X

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*(1), 204–256. https://doi.org/10.1037/0033-295X.108.1.204

Couper-Kuhlen, E. (1986). *An introduction to English prosody.* Edward Arnold.

Cox, R. F. A., van der Steen, S., Guevara Guerrero, M., Hoekstra, L., & van Dijk, M. (2016). Chromatic and anisotropic cross-recurrence quantification analysis of interpersonal behavior. In C. Webber, C. Ioana, & N. Marwan (Eds.), *Recurrence plots and their quantifications: Expanding horizons: Proceedings of the 6th international symposium on recurrence plots, Grenoble, France, 17-19 June 2015* (pp. 209-225). Springer. https://doi.org/10.1007/978-3-319-29922-8

Daane, M. C., Campbell, J. R., Grigg, W. S., Goodman, M. J., & Oranje, A. (2005). *Fourth-grade students reading aloud: NAEP 2002 special study of oral reading* (NCES 2006–469). National Center for Education Statistics. https://files.eric.ed.gov/fulltext/ED488962.pdf

de Rooij, M., & Weeda, W. (2020). Cross-Validation: A method every psychologist should know. *Advances in Methods and Practices in Psychological Science*, *3*(2), 248–263. https://doi.org/10.1177/2515245919898466

Dowhower, S. L. (1987). Effects of repeated reading on second-grade transitional readers' fluency and comprehension. *Reading Research Quarterly*, *22*(4), 389–406. https://doi.org/10.2307/747699

Dowhower, S. L. (1991). Speaking of prosody: Fluency's unattended bedfellow. *Theory Into Practice*, *30*(3), 165–175. https://doi.org/10.1080/00405849109543497

Eckmann, J. -P., Kamphorst, S. O., & Ruelle, D. (1987). Recurrence plots of dynamical systems. *Europhysics Letters*, *4*(9), 973–977. https://doi.org/10.1209/0295-5075/4/9/004

Edwards, J., Beckman, M. E., & Fletcher, J. (1991). The articulatory kinematics of final lengthening. *The Journal of the Acoustical Society of America*, *89*(1), 369–382. https://doi.org/10.1121/1.400674

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, *9*(2), 167–188. https://doi.org/10.1207/s1532799xssr0902_4

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*(3), 239–256. https://doi.org/10.1207/S1532799XSSR0503_3

Fusaroli R., & Tylén, K. (2016). Investigating conversational dynamics: Interactive alignment, interpersonal synergy, and collective task performance. *Cognitive Science, 40*(1), 145–171. https://doi.org/10.1111/cogs.12251

Godde, E., Baily, G., & Bosse, M.-L. (2022). Pausing and breathing while reading aloud: Development from 2nd to 7th grade in French speaking children. *Reading and Writing*. https://doi.org/10.1007/s11145-021-10168-z

Godde, E., Bosse, M.-L., & Bailly, G. (2020). A review of reading prosody acquisition and development. *Reading and Writing, 33*(2), 399–426. https://doi.org/10.1007/s11145-019-09968-1

Hillenbrand, J. M., & Clark, M. J. (2009). The role of f0 and formant frequencies in distinguishing the voices of men and women. *Attention, Perception, & Psychophysics, 71*(5), 1150–1166. https://doi.org/10.3758/APP.71.5.1150

Himmelmann, N. P., & Ladd, D. R. (2008). Prosodic description: An introduction for fieldworkers. *Language Documentation & Conservation*, *2*(2), 244–274.

Hirst, D. (1991). Intonation models: Towards a third generation. *Proceedings of the 12th International Congress of Phonétic Sciences*, *1,* 305–310.

Hudson, R. F., Pullen, P. C., Lane, H. B., & Torgesen, J. K. (2009). The complex nature of reading fluency: A multidimensional view. *Reading & Writing Quarterly*, *25*(1), 4–32. https://doi.org/10.1080/10573560802491208

Juul, H., Poulsen, M., & Elbro, C. (2014). Separating speed from accuracy in beginning reading development. *Journal of Educational Psychology, 106*(4), 1096–1106. https://doi.org/10.1037/a0037100

Kara, Y., Kamata, A., Potgieter, C., & Nese, J. F. T. (2020). Estimating model-based oral reading fluency: A Bayesian approach. *Educational and Psychological Measurement, 80*(5), 847–869. https://doi.org/10.1177/0013164419900208

Karageorgos, P, Müller, B., & Richter, T. (2019). Modelling the relationship of accurate and fluent word recognition in primary school. *Learning and Individual Differences, 76*. https://doi.org/10.1016/j.lindif.2019.101779

Karageorgos, P., Richter, T., Haffmans, M.-B., Schindler, J., & Naumann, J. (2020). The role of word-recognition accuracy in the development of word-recognition speed and reading comprehension in primary school: A longitudinal examination. *Cognitive Development, 56*. https://doi.org/10.1016/j.cogdev.2020.100949

Kim, Y.-S. G., Quinn, J. M., & Petscher, Y. (2021). Reading prosody unpacked: A longitudinal investigation of its dimensionality and relation with word reading and listening comprehension for children in primary grades. *Journal of Educational Psychology, 113*(3), 423–445. https://doi.org/10.1037/edu0000480

Konvalinka, I., Xygalatas, D., Bulbulia, J., Schjødt, U., Jegindø, E.-M., Wallot, S., Van Orden, G., & Roepstorff, A. (2011). Synchronized arousal between performers and related spectators in a fire-walking ritual. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 108*(20), 8514–8519. https://doi.org/10.1073/pnas.1016955108

Ktori, M., Mousikou, P., & Rastle, K. (2018). Cues to stress assignment in reading aloud. *Journal of Experimental Psychology: General, 147*(1), 36–61. http://dx.doi.org/10.1037/xge0000380

Kuhn, M. (2008). Building predicitve models in R using the caret package. *Journal of Statistical Software*, *28*(5), 1–26. https://doi.org/10.18637/jss.v028.i05

Kuhn, M. (2020). *caret: Classification and regression training*. R package (Version 6.0-86) [Computer software]. https://CRAN.R-project.org/package=caret.

Kuhn, M. R., Schwanenflugel, P. J., & Meisinger, E. B. (2010). Aligning theory and assessment of reading fluency: Automaticity, prosody, and definitions of fluency. *Reading Research Quarterly, 45*(2), 230–251. https://doi.org/10.1598/RRQ.45.2.4

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*(2), 293–323. https://doi.org/10.1016/0010-0285(74)90015-2

Landerl, K., & Reitsma, P. (2005). Phonological and morphological consistency in the acquisition of vowel duration spelling in Dutch and German. *Journal of Experimental Child Psychology, 92*(4), 322–344. https://doi.org/10.1016/j.jecp.2005.04.005

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children’s speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, *105*(3), 1455–1468. https://doi.org/10.1121/1.426686

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]. Hogrefe.

Leonardi, G. (2018). A method for the computation of entropy in the recurrence quantification analysis of categorical time series. *Physica A: Statistical Mechanics and its Applications, 512,* 824–836. https://doi.org/10.1016/j.physa.2018.08.058

Lyytinen, H., Leinonen, S., Nikula, M., Aro, M., & Leiwo, M. (1995). In search of the core features of dyslexia: Observations concerning dyslexia in the highly orthographically regular Finnish language. In V. Berninger (Ed.), *The varieties of orthographic* *knowledge II: Relationships to phonology, reading, and writing* (pp. 177–204). Kluwer.

Marwan, N. (2022). *Cross recurrence plot toolbox for MATLAB* (Version 5.24[R34]) [Computer software]. https://tocsy.pik-potsdam.de/CRPtoolbox/

Marwan, N., Romano, M. C., Thiel, M., & Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports, 438*(5), 237–329. https://doi.org/10.1016/j.physrep.2006.11.001

Marwan, N., & Webber, C. L., Jr. (2015) Mathematical and computational foundations of recurrence quantifications. In C. L. Jr. Webber & N. Marwan (Eds.), *Recurrence quantification analysis: Theory and best practices* (pp. 3–42). Springer. https://doi.org/10.1007/978-3-319-07155-8

Mennen, I., Schaeffler, F., & Dickie, C. (2014). Second language acquisition of pitch range in German learners of English. *Studies in Second Language Acquisition, 36*(2), 303–329. https://doi.org/10.1017/S0272263114000023

Mennen, I., Schaeffler, F., & Docherty, G. (2012). Cross-language differences in fundamental frequency range: A comparison of English and German. *The Journal of the Acoustical Society of America, 131*(3), 2249–2260. https://doi.org/10.1121/1.3681950

Miller, J., & Schwanenflugel, P. J. (2006). Prosody of syntactically complex sentences in the oral reading of young children. *Journal of Educational Psychology, 98*(4), 839–843. https://doi.org/10.1037/0022-0663.98.4.839

Miller, J., & Schwanenflugel, P. J. (2008). A longitudinal study of the development of reading prosody as a dimension of oral reading fluency in early elementary school children. *Reading Research Quarterly, 43*(4), 336–354. https://doi.org/10.1598/RRQ.43.4.2

Molholt, G. (1990). Spectrographic analysis and patterns in pronunciation. *Computers and the Humanities*, *24*(1), 81–92. https://doi.org/10.1007/BF00115030

Morrison, T. G., & Wilcox, B. (2020). Assessing expressive oral reading fluency. *Education Sciences*, *10*(3), 59. https://doi.org/10.3390/educsci10030059

Müller, B., Mayer, A., Richter, T., Križan, A., Hecht, T., & Ennemoser, M. (2015). Differential effects of reading trainings on reading processes: A comparison in Grade 2. *Zeitschrift für Erziehungswissenschaft*, *18*, 489–512. https://doi.org/10.1007/s11618-015-0648-0

National Reading Panel. (2000). Report of the National Reading Panel. *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). National Institute of Child Health and Human Development. https://files.eric.ed.gov/fulltext/ED444126.pdf

O'Brien, B. A., Wallot, S., Haussmann, A., & Kloos, H. (2014). Using complexity metrics to assess silent reading fluency: A cross-sectional study comparing oral and silent reading. *Scientific Studies of Reading*, *18*(4), 235–254. https://doi.org/10.1080/10888438.2013.862248

Paige, D. D., Rasinski, T., Magpuri-Lavell, T., & Smith, G. S. (2014). Interpreting the relationships among prosody, automaticity, accuracy, and silent reading comprehension in secondary students. *Journal of Literacy Research*, *46*(2), 123–156. https://doi.org/10.1177/1086296X14535170

Pinnell, G. S., Pikulski, J. J., Wixson, K. K., Campbell, J. R., Gough, P. B., & Beatty, A. S. (1995). *Listening to children read aloud: Data from NAEP's integrated reading performance record (IRPR) at grade 4* (NCES Publication No. 95-726). National Center for Education Statistics. https://files.eric.ed.gov/fulltext/ED378550.pdf

Protopapas, A. (2006). On the use and usefulness of stress diacritics in reading Greek. *Reading and Writing, 19*(2), 171–198. https://doi.org/10.1007/s11145-005-4023-z

Rasinski, T. (2004). Creating fluent readers. *Educational Leadership, 61*(6), 46–51. https://www.ascd.org/el/articles/creating-fluent-readers

Rasinski, T., Rikli, A., & Johnston, S. (2009). Reading fluency: More than automaticity? More than a concern for the primary grades? *Literacy Research and Instruction, 48*(4), 350–361. https://doi.org/10.1080/19388070802468715

Rastle, K., & Coltheart, M. (2000). Lexical and nonlexical print-to-sound translation of disyllabic words and nonwords. *Journal of Memory and Language, 42*(3), 342–364. http://dx.doi.org/10.1006/jmla.1999.2687

Scharff-Rethfeldt, W., Miller, N., & Mennen, I. (2008). Unterschiede in der mittleren Sprechtonhöhe bei Deutsch/Englisch bilingualen Sprechern [Speaking fundamental frequency differences in German-English bilinguals]. *Sprache Stimme Gehör, 32*(3), 123–128. https://doi.org/10.1055/s-0028-1083799

Schmidt, T., & Wörner, K. (2009). EXMARaLDA – Creating, analysing and sharing spoken language corpora for pragmatic research. *Pragmatics*, *19*(4), 565–582. https://doi.org/10.1075/prag.19.4.06sch

Schwanenflugel, P. J., Hamilton, A. M., Kuhn, M. R., Wisenbaker, J. M., & Stahl, S. A. (2004). Becoming a fluent reader: Reading skill and prosodic features in the oral reading of young readers. *Journal of Educational Psychology, 96*(1), 119–129. https://doi.org/10.1037/0022-0663.96.1.119

Selting, M., Auer, P., Barden, B., Bergmann, J., Couper-Kulhen, E., Günther, S., Quasthoff, U., Meier, C., Schlobinski, P., & Uhmann, S. (1998). Gesprächsanalytisches Transkriptionssystem (GAT) [Discourse and conversation-analytic transcription system]. *Linguistische Berichte, 173*, 91–122.

Silverman, K. E., Beckman, M. E., Pitrelli, J. F., Ostendorf, M., Wightman, C. W., Price, P. J., Pierrehumbert, J. B., & Hirschberg, J. (1992). TOBI: A standard for labeling English prosody. In *Proc. 2nd International Conference on Spoken Language Processing (ICSLP 1992)*, 867–870.

Silverman, R. D., Speece, D. L., Harring, J. R., & Ritchey, K. D. (2013). Fluency has a role in the simple view of reading. *Scientific Studies of Reading*, *17*(2), 108–133. https://doi.org/10.1080/10888438.2011.618153

Stevens, K. C. (1981). Chunking material as an aid to reading comprehension. *Journal of Reading*, *25*(2), 126–129.

Temperley, D. (2009). Distributional stress regularity: A corpus study. *Journal of Psycholinguistic Research*, *38*(1), 75–92. https://doi.org/10.1007/s10936-008-9084-0

The Math Works, Inc. (2017). *MATLAB* (Version R2017a) [Computer software]. MathWorks. https://www.mathworks.com/

Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *The Journal of the Acoustical Society of America*, *85*(4), 1699–1707. https://doi.org/10.1121/1.397959

Torgesen, J. K., & Hudson, R. F. (2006). Reading fluency: Critical issues for struggling readers. In S. J. Samuels & A. E. Farstrup (Eds.), *What research has to say about fluency instruction* (pp. 130–158). International Reading Association.

Tsanas, A. (2012). *Accurate telemonitoring of Parkinson's disease symptom severity using nonlinear speech signal processing and statistical machine learning* [Unpublished doctoral dissertation]. Oxford University.

Tschense, M. & Wallot, S. (2022). Using measures of reading time regularity (RTR) to quantify eye movement dynamics, and how they are shaped by linguistic information. *Journal of Vision, 22*(9), 1–21. https://doi.org/10.1167/jov.22.6.9

Turk, A. E., Shattuck-Hufnagel, S. (2007). Multiple targets of phrase-final lengthening in American English words. *Journal of Phonetics*, *35*(4), 445–472. https://doi.org/10.1016/j.wocn.2006.12.001

Wallot, S. (2017). Recurrence quantification analysis of processes and products of discourse: A tutorial in R. *Discourse Processes*, *54*(5-6), 382–405. https://doi.org/10.1080/0163853X.2017.1297921

Wallot, S., O'Brien, B. A., Haussmann, A., Kloos, H., & Lyby, M. S. (2014). The role of reading time complexity and reading speed in text comprehension. *Journal of experimental psychology. Learning, Memory, and Cognition*, *40*(6), 1745–1765. https://doi.org/10.1037/xlm0000030

Webber, C. L., & Zbilut, J. (2005). Recurrence quantification analysis of nonlinear dynamical systems. In M. A. Riley & G. C. Van Orden (Eds.), *Tutorials in contemporary nonlinear methods for the behavioral sciences* (pp. 26–94). https://www.nsf.gov/pubs/2005/nsf05057/nmbs/nmbs.pdf

Wolters, A. P., Kim, Y.-S. G., & Szura, J. W. (2022). Is reading prosody related to reading comprehension? A meta-analysis. *Scientific Studies of Reading*, *26*(1), 1–20. https://doi.org/10.1080/10888438.2020.1850733

7 General Discussion

The purpose of this thesis was to extend the literature on reading fluency in German. Even though ample research has investigated word reading processes and the effects of reading interventions in primary school, several aspects that have implications for interventions need to be clarified. This thesis pursued two major aims that should be addressed when tailoring future interventions, the result of which could act as a guideline for future research.

The first major aim addressed the existence of a basic accuracy level in German primary school children. Even though a basic accuracy level has been found in Danish (Juul et al., 2014) and this pattern was replicated in English (Altani et al., 2020), no study to date has investigated a basic accuracy level in the German language as well as its effects on a reading intervention and the development of German readers in primary school. Study 1 aimed to address this lack by identifying a basic accuracy level for the German language and investigating whether readers respond differently to a syllable reading intervention depending on the extent of their word reading accuracy before the intervention. Furthermore, Study 2 extended the results of Study 1 by investigating whether the achievement time of the basic accuracy level would have an impact on reading development of German children in primary school.

The second major aim investigated whether prosodic patterns in German primary school readers play an important role in distinguishing between skilled and struggling readers. In addition, the classification of struggling and skilled readers with prosodic patterns was compared to the classification with a standard method. Given that prosody is often neglected when assessing reading fluency, measured typically with only word reading accuracy and word reading speed (Dowhower, 1991), knowledge about prosody, especially prosodic patterns during oral reading, is still lacking. Therefore, Study 3 of this thesis focused on identifying prosodic patterns by employing RQA and comparing its effectiveness in classifying struggling from skilled readers with the effectiveness of prosodic features obtained with a transcription system.

In the following sections, the key findings of the three studies are briefly summarized and integrated into existing literature. Subsequently, limitations and directions for future research are discussed before drawing a conclusion and highlighting practical implications.

7.1 Establishing a Basic Accuracy Level in German

In Study 1, in a pre-post control-group design, a basic accuracy level was established in a sample of German fourth graders and its importance for the development of word reading speed was investigated. The results support the accuracy-before-speed pattern that was found by Juul et al. (2014) and Altani et al. (2020). Apparently, German fourth graders must achieve a specific word reading accuracy before their word reading speed increases. This finding is also in line with the model of automaticity proposed by LaBerge and Samuels (1974). According to LaBerge and Samuels, even though beginning readers may recognize words somewhat accurately, they still need a lot of time and attention. In contrast, skilled readers recognize words effortlessly as a whole and processing takes little time. Hence, exceeding an accuracy level could serve as an indication of starting to or even already achieving automaticity.

Moreover, the findings of Study 1 supported the hypothesis that the syllable reading intervention would have differential positive effects depending on the word reading accuracy that participants have before the training. Participants below the established accuracy level showed more gains on word-recognition accuracy than untrained children below the established accuracy level. This finding is also in accordance with results of a comparable intervention in the German language, which showed positive treatment effects on word-recognition accuracy and underscored the importance of syllables as a processing unit in reading (Ritter, 2005; Ritter & Scheerer-Neumann, 2009). The intervention used in Study 1 was a combination of phonics and reading-fluency approaches. In the first training sessions, participants were taught the principles of segmenting a word into syllables. In the later sessions, they put the segmenting rules they learned into practice by quickly reading regular and irregular words and then practicing recognizing these words via orthographical decoding to improve their word-recognition speed. Hence, participants had also the chance to store the new words they encountered into their orthographical lexicon (Zarić & Nagler, 2021).

Contrary to our expectation that only participants above the basic accuracy level would make gains on word-recognition speed, participants below the basic accuracy level also made significant gains and caught up to other more skilled readers. A similar result was also reported in the study of Heikkilä et al. (2013) with Finnish poor readers that attended the second and third grade. In their study, the slowest of poor readers made more gains in reading speed than the faster poor readers. According to Breznitz (2006), the development of reading speed appears to reach an asymptote at some point, which could explain why participants below the basic accuracy level were able to catch up to participants above the level. Furthermore, another plausible explanation, albeit speculative, could be that participants below the basic accuracy level exceeded the accuracy level at some time during the training and then also started to make gains on word-recognition speed.

Lastly, the results we obtained from Study 1 showed that participants above the established accuracy level make more gains on reading comprehension compared to all other groups. Participants above the basic accuracy threshold were already able to accurately recognize words at the beginning of the study. Thus, they had more time to develop their word-recognition speed than readers below the basic accuracy level, achieve automaticity at some point, and even outperform untrained children above the basic accuracy threshold because of the intervention. Recognizing words efficiently frees up cognitive resources that can be used for higher-order comprehension processes at the sentence and text levels (Perfetti & Hart, 2002), and facilitate reading comprehension (Kim et al., 2015). Contrary to our expectations, the indirect effect of word-recognition speed on reading comprehension was not statistically significant, whereas the direct treatment effect on reading comprehension remained significant. Thus, the findings preclude the conclusion that word-recognition speed is the cause of this difference. Moreover, whether the lack of an indirect effect was due to low power (given the small sample size) or whether our reading intervention had an indirect effect on reading comprehension via an unaccounted mechanism remains unclear.

Study 2 addressed some important limitations of Study 1, and it extended its results in a longitudinal design by measuring word recognition and reading comprehension of children at the end of each grade throughout German primary school. In Study 2, we included more measurement points to investigate the basic accuracy achievement time throughout German primary school. Furthermore, we also included participants' first language in our models to control for individual differences. As expected in the discussion of Study 1 and in accordance with Juul et al.'s (2014) results, the basic achievement accuracy time had an impact on word-recognition speed, with children who had achieved the basic accuracy level by the end of Grade 1 having a steeper word-recognition speed development compared to children who achieved this criterion only after Grade 3 or not at all. After achieving the basic accuracy level, the children might have had more time and cognitive resources for developing their word-recognition speed throughout primary school (see also Juul et al., 2014). However, counter to our expectations, this time advantage was not found in the comparison with children who achieved the basic accuracy level by the end of Grade 2. Word-recognition speed of children who achieved the basic accuracy level by the end of Grades 1 and 2 developed similarly. One possible explanation for this result could be that children who were assumed to have achieved the basic accuracy threshold by the end of Grade 2 had already achieved this level at the beginning of Grade 2. Hence, the basic accuracy achievement time of these two groups, could have been only a few months apart instead of the assumed one year. Nevertheless, the groups who achieved the basic accuracy threshold after Grade 2 or not at all were still unable to catch up with the children who achieved this criterion by the end of Grade 1, and the differences in word-recognition speed remained stable until the end of primary school.

The results of Study 2 also supported the second hypothesis which investigated the impact of the basic accuracy achievement time on the development of reading comprehension. Reading comprehension of children who achieved the accuracy level by the end of Grades 1 and 2 had a steeper development than children who achieved this accuracy level only after Grade 3 or not at all. When considering the development of word-recognition speed, these groups also showed a steeper word-recognition speed development than the other groups. The findings of the study are correlative and causal interpretations should be avoided. Nevertheless, one speculative explanation that could be investigated in future studies is that these groups were able to achieve automaticity. Automaticity freed cognitive resources for higher-order processes (Perfetti, 1985), resulting in better performance in reading comprehension tasks than for children who were still struggling to accurately recognize words. However, again, no developmental difference was found between children who achieved the basic accuracy threshold by the end of Grade 1 versus Grade 2. Given that reading comprehension was first measured at the end of Grade 2, children in both groups had by then already achieved the basic accuracy threshold and were making gains in word-recognition speed. Thus, this method limitation could have contributed to the null finding in developmental difference. Nevertheless, differences in reading comprehension between the two groups remained stable until the end of Grade 4.

The results of Study 1 and Study 2 together indicate a basic accuracy level in word recognition that German primary school children should achieve before their word-recognition speed starts improving. Furthermore, the earlier the children achieve this criterion, the more gains they will make in word-recognition speed and reading comprehension throughout primary school. Finally, a syllable reading intervention might have differential effects depending on the word-recognition accuracy of the participant.

7.2 Identifying Important Prosodic Patterns in German

For the second aim of this thesis, Study 3 focused on identifying prosodic patterns in oral readings that might differentiate between German struggling and skilled readers in second and fourth grade. Prosodic patterns were estimated with measures obtained with the RQA, which has been shown to predict reading comprehension from reading time measures (Wallot, 2017; Wallot et al., 2014). Even though Study 3 had an exploratory aim, we expected that the signals of skilled German readers would be more structured and stable and would show a less complex and chaotic pattern in all prosodic features with less and shorter pauses and stresses than the signals of struggling German readers. Furthermore, to ensure the additional value of this method, models estimated with prosodic patterns were also compared to models estimated with prosodic features that were obtained with the coding schema GAT.

The findings of the prosodic patterns obtained with RQA in Grade 2 indicated that pausing plays an important role in the distinction between struggling and skilled readers, with struggling readers making longer pauses within and between words. According to Miller and Schwanenflugel (2006,) longer pauses might be related to word decoding deficits. German struggling readers might recognize words somewhat accurately via phonological recoding, but they are still slow in comparison to skilled readers (Landerl & Wimmer, 2008). Hence, making longer pauses within and between words of a sentence might give them the time they need to decode a word (Godde et al., 2022).

Furthermore, struggling readers also had a longer average recurrence interval between pauses compared to skilled readers. Longer recurrence intervals indicate more time between recurring pauses in the text for struggling readers, which might be related to stumbling. This result is consistent with the assumption of Torgesen and Hudson (2006; see also Hudson et al., 2008) who suggested that children unable to recognize words efficiently might stumble over multiple words and make numerous errors before successfully reading it and going on to the next one. Nevertheless, this finding does not imply that struggling second graders paused less than skilled second graders. Pause occurrences within and between words estimated with GAT indicated that struggling readers made around 13 pauses more than skilled readers. Lastly, the findings with the prosodic features obtained with GAT measures in Grade 2 indicated that struggling readers paused more often within and between words of a phrase and prolonged syllables within a word more often than skilled readers. This result is also consistent with previous findings which found differences in the frequency of ungrammatical pauses between struggling and skilled pauses (e.g., Miller & Schwanenflugel, 2006). Apparently, struggling readers make more ungrammatical pauses because of decoding uncertainties (e.g., Benjamin & Schwanenflugel, 2010). In addition, syllable elongation is consistent with the findings of Lyytinen et al. (1995) with Finnish struggling and skilled adult readers. These struggling readers exhibited nearly 30.2 times more consonant lengthening occurrences and 6.6 times more often vowel lengthening occurrences than skilled readers of the equivalent age. These disparities, according to Lyytinen et al., might be attributable to a phonological processing deficit, which manifests as insensitivity to the duration of phonemic segments and difficulties manipulating phonemic segments.

In Grade 4, other prosodic patterns obtained with RQA were found to be more important than pausing when differentiating between struggling and skilled fourth graders. Skilled readers had, on average, shorter recurrence intervals between amplitudes, that is, less time passed until a recurrent pronunciation occurred, showed less pitch patterns, and a faster rate of amplitude change. Skilled readers can already efficiently recognize words and thus have more cognitive resources for the higher-order processes such as semantic processing and reading comprehension (LaBerge & Samuels, 1974; Perfetti, 1985). Thus, they can read fluently and consequently have shorter signal segments than struggling readers, which might explain the shorter intervals between recurring amplitudes. Furthermore, a recent study of Kim et al. (2021) also suggests that word reading is positively related to pitch variation (i.e., greater pitch change and intonation contour). Hence, efficient semantic processing might also allow skilled readers to read with a more appropriate and structured pitch variation than struggling readers. Lastly, the difference in amplitude change is in accordance with the results of the study of Clay and Imlach (1971) who found that struggling readers insert a stress in almost every word as if reading words in a list, whereas skilled readers insert stress every 4.7 words. This pattern is consistent with the current results in Study 2 in which struggling readers lingered in a state for a longer period and showed more amplitude repetitions. Lastly, the findings of prosodic features obtained with GAT indicated a similar outcome to that found in Grade 2, namely that struggling readers pause more often within and between words of a phrase than skilled readers.

Contrary to our expectations, models with RQA could not outperform models with GAT. However, further analyses with the most important variables from both methods revealed that these methods complement each other, and prosodic patterns obtained with RQA deliver additional information. The combined models for both grades exhibited a better overall fit and misclassified fewer children than the individual models in which RQA and

GAT were analyzed separately. Only syllable lengthening within and between words captured with GAT remained significant in Grade 2, whereas inappropriate pauses obtained with GAT and the complexity of pitch patterns obtained with RQA remained significant in Grade 4.

7.3 Limitations and Future Directions

The findings presented in this thesis contribute to understanding the underlying mechanisms of reading fluency and can serve as a steppingstone for future studies. However, there are some limitations that must be considered when interpreting the results.

First, the two studies in this dissertation differ from Juul et al. (2014) in fundamental ways. Even though both the German and Danish language have a complex syllabic structure, they are different in their orthographic depth. The German language has a relatively shallow orthography, whereas the Danish language has a deep orthography (Elbro, 2005; Seymour et al., 2003). Moreover, the teaching methods used in each countries differ. German children are taught to read through the synthetic phonics method, whereas Danish children are taught to read by a mixture of methods such as whole-word look-and-say, the use of contextual cues, some phonics, and easy book reading (Elbro, 2005). Consequently, in Studies 1 and 2, a silent word-recognition task (lexical decision task) was used instead of the read-aloud task (word naming task) used in Juul et al. Even though both tasks capture word recognition, they emphasize different processes.

Another discrepancy is that Juul et al. (2014) followed the development of Danish children from the end of kindergarten to the end of Grade 2 and had 11 measurement points, whereas only two measurement points were administered in Study 1 and four measurement points in Study 2. Thus, even though the results are similar, the exact time at which children achieved the basic accuracy level cannot be determined. Nonetheless, the goal of this dissertation was not to directly replicate the study of Juul et al. in a German sample but rather to use the fundamental idea of a basic accuracy level that can be applied to other alphabetic languages and to explore the potential role of this level in German reading development. Future studies could employ a longitudinal design with shorter measurement points to estimate a more accurate timepoint at which children achieve the basic accuracy level. Furthermore, studies should also investigate in a larger sample whether an indirect treatment effect occurs on reading comprehension through word-recognition speed by ensuring the temporal precedence of the mediator. Finally, assessing the co-development of word-recognition accuracy, speed, prosodic patterns, and reading comprehension with a crossed-lagged panel analysis to highlight the dependencies between the variables and their effects on potential exogenous variables would be informative.

Second, the basic accuracy level that was established in Study 1 should not be regarded as an absolute value. Even though we found a similar value in the study of Juul et al. (2014), the value could still vary depending on the task used to measure word reading accuracy. Nevertheless, the results of Study 1 and Study 2 supported the notion that word reading speed of German readers in primary school starts improving only after readers achieve a basic accuracy level, which for the ProDi-L instrument (Richter et al., 2017) is 71%. This basic accuracy level was also used in Study 2 to divide children into five groups. However, even though the same lexical decision task of ProDi-L was used in both studies to measure word-recognition accuracy and speed, the level was determined in a Grade 4 sample. Hence, no conclusion can be drawn on whether separate basic accuracy levels for each grade estimated with the new data would have been more appropriate in Study 2. Nevertheless, separate levels would have likely led to different criteria for each grade and would have impeded the group comparison. Future research could estimate the basic accuracy level using various instruments to determine the relationship between the accuracy level and the instrument employed. Furthermore, a variety of instruments could be employed to measure and estimate the basic accuracy level for different age groups and to investigate whether estimation of the basic accuracy level is age dependent.

Third, children in Study 3 were allocated to tutees and tutors after the pre-test and received either a reading fluency training or a visuospatial working memory training. Children read the story, however, almost at the end of the training. Thus, even though differences in reading comprehension were found between struggling and skilled readers at post-test, no conclusion can be drawn on whether the results would have been the same if children had not participated in the fluency training. Consequently, future studies could request from participants to orally read a text at the beginning and at the end of the intervention to compare prosodic patterns before and after the reading intervention. Future studies could also investigate whether prosodic patterns of children vary depending on whether they have achieved the basic accuracy level established in Studies 1 and 2. Word-recognition speed of children who have achieved the basic accuracy level can be assumed to improve faster than word-recognition speed of children who have not achieved this criterion. Hence, children who have achieved the basic accuracy level at some point will also likely show similar prosodic patterns as the skilled readers of Study 3.

7.4 Conclusion and Practical Implications

The goal of this thesis was to examine and enhance the understanding of reading fluency components in the German language. Despite several limitations, the results of the present dissertation complement and extend knowledge on reading fluency and can be considered an important contribution to applied research by providing useful information in the field of evidence-based education. Four interesting conclusions arose regarding the relationship between word-recognition accuracy and word-recognition speed, the effects of a syllable reading intervention, and prosodic patterns that distinguish between struggling and skilled readers.

First, the idea of an accuracy-before-speed pattern in German primary school children was established. German primary school children should first achieve an adequate word-recognition accuracy level before they start making gains in word-recognition speed. Second, the earlier primary school children achieve this accuracy level, the more gains they will make in word-recognition speed and reading comprehension. Third, a syllable reading intervention in fourth grade might have different results depending on the word-recognition accuracy of the participants. Finally, prosodic patterns can distinguish between skilled and struggling readers in second and fourth grade and might provide additional information beyond standard methods.

The findings provide insight into the components of reading fluency and have the potential to enhance knowledge about individual preconditions that are necessary to make specific reading programs. In Study 1, the existence of a basic accuracy level in German language was established, and in Study 2, achieving a basic accuracy level in the first 2 years of German primary school leads to different developmental trajectories. These findings have a theoretical significance for theories of reading development because they suggest that a general developmental pattern in which accurate reading is established before reading speed can develop through routinization. These findings are not trivial. In principle, word reading speed could begin to develop in parallel with word reading accuracy. The practical significance for reading instruction in children learning to read in German, and hopefully at a later point for other languages as well, is that word-reading interventions should be implemented no later than the first two primary school years to achieve the best results. Hence, in addition to existing approaches, measuring children’s word-recognition accuracy by the end of Grade 1 can be informative. Of course, the exact value for the basic accuracy level must be determined for any particular test before it can be utilized in practical applications, unless the ProDi-L instrument (Richter et al., 2017) is used, which was employed in the two studies and the value of the basic accuracy level has been established. The basic accuracy level might serve as a criterion to help in the assessment of whether children should participate in an intervention program.

Incorporating the basic accuracy level as a criterion for intervention into the regular school curriculum could compensate for the lack of educational promotion of struggling readers that was observed during PIRLS (Bos et al., 2017). According to the Staatsinstitut für Schulqualität und Bildungsforschung (n.d.), students should be able to read with motivation and train their reading fluency and accuracy by the end of Grade 2. As a result, rather than focusing on word reading speed, focusing on fostering word reading accuracy in children who have not achieved the basic accuracy level might be more appropriate. The same concept might also be applied when designing reading interventions for individual learners. Children below the basic accuracy level could receive phonemic awareness instruction before advancing to phonics instruction and fluency interventions Finally, prosodic patterns in oral reading might be utilized in conjunction with fluency interventions to assess children's growth and shortcomings.

References

Aaron, P. G., Joshi, R. M., Ayotollah, M., Ellsberry, A., Henderson, J., & Lindsey, K. (1999). Decoding and sight-word naming: Are they independent components of word recognition skill? *Reading and Writing*, *11*(2), 89–127. https://doi.org/10.1023/A:1008088618970

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*(2), 245–264. https://doi.org/10.1348/000712607X224478

Adlof, S. M., Catts, H. W., & Little, T. D. (2006). Should the simple view of reading include a fluency component? *Reading and writing*, *19*(9), 933–958. https://doi.org/10.1007/s11145-006-9024-z

Allington, R. L. (1983). Fluency: The neglected reading goal. *The Reading Teacher*, *36*(6), 556–561.

Altani, A., Protopapas, A., Katopodi, K., & Georgiou, G. K. (2020). From individual word recognition to word list and text reading fluency. *Journal of Educational Psychology*, *112*(1), 22–39. https://doi.org/10.1037/edu0000359

Álvarez-Cañizo, M., Suárez-Coalla, P., & Cuetos, F. (2015). The role of reading fluency in children’s text comprehension. *Frontiers in Psychology*, *6*, Article 1810. https://doi.org/10.3389/fpsyg.2015.01810

Bast, J., & Reitsma, P. (1998). Analyzing the development of individual differences in terms of Matthew effects in reading: Results from a Dutch longitudinal study. *Developmental Psychology*, *34*(6), 1373–1399. https://doi.org/10.1037/0012-1649.34.6.1373

Benjamin, R. G., & Schwanenflugel, P. J. (2010). Text complexity and oral reading prosody in young readers. *Reading Research Quarterly, 45*(4), 388–404. https://doi.org/10.1598/RRQ.45.4.2

Benjamin, R. G., Schwanenflugel, P. J., Meisinger, E. B., Groff, C., Kuhn, M. R., & Steiner, L. (2013). A spectrographically grounded scale for evaluating reading expressiveness. *Reading Research Quarterly, 48*(2), 105–133. https://doi.org/10.1002/rrq.43

Bos, W., Valtin, R., Hußmann, A., Wendt, H., & Goy, M. (2017). IGLU 2016: Wichtige Ergebnisse im Überblick [IGLU 2016: Important results at a glance]. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T. C. Stubbe, & R. Valtin (Eds.), *IGLU 2016: Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (pp. 13-28). https://doi.org/10.25656/01:15476

Bremerich-Vos, A., Wendt, H., & Bos, W. (2017). Lesekompetenzen im internationalen Vergleich: Testkonzeption und Ergebnisse [International comparison of reading skills: Test design and results]. In A. Hußmann, H. Wendt, W. Bos, A. Bremerich-Vos, D. Kasper, E.-M. Lankes, N. McElvany, T. C. Stubbe, & R. Valtin (Eds.), *IGLU 2016: Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (pp. 79–142). Waxmann Münster. https://doi.org/10.25656/01:15476

Bressem, J. (2013). Transcription systems for gestures, speech, prosody, postures, and gaze. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & S. Tessendorf (Eds.), *Body – language – communication: An international handbook on multimodality in human interaction.* (Vol. 1, pp. 1037–1059). De Gruyter Mouton. https://doi.org/doi:10.1515/9783110261318.1037

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

Bundesagentur für Arbeit. (n.d.). *Entwicklung Arbeitslosigkeit: Berichtsjahre* [Development of unemployment: Reporting years]. Retrieved June 3, 2022, from https://statistik.arbeitsagentur.de/DE/Navigation/Statistiken/Interaktive-Statistiken/Zeitreihen/Lange-Zeitreihen-Nav.html

Campione, E., & Véronis, J. (2002). A large-scale multilingual study of silent pause duration. In B. Bel & I. Marlien (Eds.), *Speech prosody 2002: Proceedings: Aix-en-Provence, France, 11-13 April 2002* (pp. 199-202). Laboratoire Parole et Langage.

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest*, *19*(1), 5–51. https://doi.org/10.1177/1529100618772271

Clay, M. M. (1979). *Reading: The patterning of complex behaviour*. Heinemann Educational Books.

Clay, M. M., & Imlach, R. H. (1971). Juncture, pitch, and stress as reading behavior variables. *Journal of Verbal Learning and Verbal Behavior*, *10*(2), 133–139. https://doi.org/10.1016/S0022-5371(71)80004-X

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*(1), 204–256. https://doi.org/10.1037/0033-295X.108.1.204

Couper-Kuhlen, E. (1986). *An introduction to English prosody*. Edward Arnold.

Doctor, E. A., & Coltheart, M. (1980). Children's use of phonological encoding when reading for meaning. *Memory & Cognition*, *8*(3), 195–209. https://doi.org/10.3758/BF03197607

Dowhower, S. L. (1991). Speaking of prosody: Fluency's unattended bedfellow. *Theory Into Practice*, *30*(3), 165–175. https://doi.org/10.1080/00405849109543497

Ehri, L. C. (2014). Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning. *Scientific Studies of Reading*, *18*(1), 5–21. https://doi.org/10.1080/10888438.2013.819356

Elbro, C. (2005). Literacy acquisition in Danish: A deep orthography in cross-linguistic light. In R. Malatesha Joshi & P. G. Aaron (Eds.), *Handbook of orthography and literacy* (pp. 31–45). Routledge. https://doi.org/10.4324/9780203824719

Eysenck, M. W., & Keane, M. T. (2020). *Cognitive psychology: A student's handbook* (8th ed.). Taylor & Francis. https://doi.org/10.4324/9781351058513

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*(3), 239–256. https://doi.org/10.1207/S1532799XSSR0503_3

Galuschka, K., Görgen, R., Kalmar, J., Haberstroh, S., Schmalz, X., & Schulte-Körne, G. (2020). Effectiveness of spelling interventions for learners with dyslexia: A meta-analysis and systematic review. *Educational Psychologist*, *55*(1), 1–20. https://doi.org/10.1080/00461520.2019.1659794

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*(2), Article e89900.

Gangl, M., Moll, K., Jones, M. W., Banfi, C., Schulte-Körne, G., & Landerl, K. (2018). Lexical reading in dysfluent readers of German. *Scientific Studies of Reading*, *22*(1), 24–40. https://doi.org/10.1080/10888438.2017.1339709

Gersten, R., Haymond, K., Newman-Gonchar, R., Dimino, J., & Jayanthi, M. (2020). Meta-analysis of the impact of reading interventions for students in the primary grades.

Journal of Research on Educational Effectiveness, *13*(2), 401–427. https://doi.org/10.1080/19345747.2019.1689591

Godde, E., Bailly, G., & Bosse, M.-L. (2022). Pausing and breathing while reading aloud: Development from 2nd to 7th grade in French speaking children. *Reading and Writing*, *35*(1), 1–27. https://doi.org/10.1007/s11145-021-10168-z

Godde, E., Bosse, M.-L., & Bailly, G. (2020). A review of reading prosody acquisition and development. *Reading and Writing*, *33*(2), 399–426. https://doi.org/10.1007/s11145-019-09968-1

Grotlüschen, A., Buddeberg, K., Dutz, G., Heilmann, L., & Stammer, C. (2020). Low literacy in Germany. Results from the second German literacy survey. *European Journal for Research on the Education and Learning of Adults*, *11*(1), 127–143. https://doi.org/10.3384/rela.2000-7426.rela9147

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*(6), 398–414. https://doi.org/10.1080/10888438.2012.753452

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, *2*(2), 127–160. https://doi.org/10.1007/BF00401799

Hudson, R. F., Lane, H. B., & Pullen, P. C. (2005). Reading fluency assessment and instruction: What, why, and how? *The Reading Teacher*, *58*(8), 702–714. https://doi.org/10.1598/RT.58.8.1

Hudson, R. F., Pullen, P. C., Lane, H. B., & Torgesen, J. K. (2008). The complex nature of reading fluency: A multidimensional view. *Reading & Writing Quarterly*, *25*(1), 4–32. https://doi.org/10.1080/10573560802491208

Jimerson, S. R. (1999). On the failure of failure: Examining the association between early grade retention and education and employment outcomes during late adolescence. *Journal of School Psychology*, *37*(3), 243–272. https://doi.org/10.1016/S0022-4405(99)00005-9

Juul, H., Poulsen, M., & Elbro, C. (2014). Separating speed from accuracy in beginning reading development. *Journal of Educational Psychology*, *106*(4), 1096–1106. https://doi.org/10.1037/a0037100

Kara, Y., Kamata, A., Potgieter, C., & Nese, J. F. T. (2020). Estimating model-based oral reading fluency: A Bayesian approach. *Educational and Psychological Measurement*, *80*(5), 847–869. https://doi.org/10.1177/0013164419900208

Kim, Y.-S. G. (2015). Developmental, component-based model of reading fluency: An investigation of predictors of word-reading fluency, text-reading fluency, and reading comprehension. *Reading Research Quarterly*, *50*(4), 459–481. https://doi.org/10.1002/rrq.107

Kim, Y.-S., Petscher, 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*(1), 131–150. https://doi.org/10.1007/s11145-013-9455-2

Kim, Y.-S. G., Quinn, J. M., & Petscher, Y. (2021). Reading prosody unpacked: A longitudinal investigation of its dimensionality and relation with word reading and listening comprehension for children in primary grades. *Journal of Educational Psychology*, *113*(3), 423–445. https://doi.org/10.1037/edu0000480

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge university press.

Ktori, M., Mousikou, P., & Rastle, K. (2018). Cues to stress assignment in reading aloud. *Journal of Experimental Psychology: General*, *147*(1), 36–61. https://doi.org/10.1037/xge0000380

Kuhn, M. R., Schwanenflugel, P. J., & Meisinger, E. B. (2010). Aligning theory and assessment of reading fluency: Automaticity, prosody, and definitions of fluency. *Reading Research Quarterly, 45*(2), 230–251. https://doi.org/10.1598/RRQ.45.2.4

Kuhn, M. R., & Stahl, S. A. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology*, *95*(1), 3–21. https://doi.org/10.1037/0022-0663.95.1.3

Küspert, P., & Schneider, W. (2018). *Hören, lauschen, lernen: Sprachspiele für Kinder im Vorschulalter. Würzburger Trainingsprogramm zur Vorbereitung auf den Erwerb der Schriftsprache* [Hear, listen, learn: Language games for preschoolers. Würzburg training program to prepare for the acquisition of written language] (7th ed.). Vandenhoeck & Ruprecht.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*(2), 293–323. https://doi.org/10.1016/0010-0285(74)90015-2

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*(1), 150–161. https://doi.org/10.1037/0022-0663.100.1.150

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, *105*(3), 1455–1468. https://doi.org/10.1121/1.426686

Lyytinen, H., Leinonen, S., Nikula, M., Aro, M., & Leiwo, M. (1995). In search of the core features of dyslexia: Observations concerning dyslexia in the highly orthographically regular Finnish language. In V. Berninger (Ed.), *The varieties of orthographic knowledge II: Relationships to phonology, reading, and writing* (pp. 177–204). Kluwer.

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*(2), 140–149. https://doi.org/10.1016/j.bandl.2006.04.003

Mennen, I., Schaeffler, F., & Docherty, G. (2012). Cross-language differences in fundamental frequency range: A comparison of English and German. *The Journal of the Acoustical Society of America*, *131*(3), 2249–2260. https://doi.org/10.1121/1.3681950

Miller, J., & Schwanenflugel, P. J. (2006). Prosody of syntactically complex sentences in the oral reading of young children. *Journal of Educational Psychology*, *98*(4), 839–853. https://doi.org/10.1037/0022-0663.98.4.839

National Reading Panel. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). National Institute of Child Health and Human Development. https://files.eric.ed.gov/fulltext/ED444126.pdf

Neitzel, A. J., Lake, C., Pellegrini, M., & Slavin, R. E. (2021). A synthesis of quantitative research on programs for struggling readers in elementary schools. *Reading Research Quarterly*, *57*(1), 149–179. https://doi.org/10.1002/rrq.379

Norris, D. (2013). Models of visual word recognition. *Trends in Cognitive Sciences*, *17*(10), 517–524. https://doi.org/10.1016/j.tics.2013.08.003

Nouwens, S., Groen, M. A., Kleemans, T., & Verhoeven, L. (2021). How executive functions contribute to reading comprehension. *British Journal of Educational Psychology*, *91*(1), 169–192. https://doi.org/10.1111/bjep.12355

Ober, T. M., Brooks, P. J., Homer, B. D., & Rindskopf, D. (2020). Executive functions and decoding in children and adolescents: A meta-analytic investigation. *Educational Psychology Review*, *32*(3), 735–763. https://doi.org/10.1007/s10648-020-09526-0

OECD. (2019). *PISA 2018 assessment and analytical framework*. OECD Publishing. https://doi.org/10.1787/b25efab8-en

Paige, D. D., Rasinski, T., Magpuri-Lavell, T., & Smith, G. S. (2014). Interpreting the relationships among prosody, automaticity, accuracy, and silent reading comprehension in secondary students. *Journal of Literacy Research*, *46*(2), 123–156. https://doi.org/10.1177/1086296x14535170

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

Perfetti, C. A. (1985). *Reading ability*. 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). John Benjamins Publishing Company. https://doi.org/10.1075/swll.11

Pikulski, J. J., & Chard, D. J. (2005). Fluency: Bridge between decoding and reading comprehension. *The Reading Teacher, 58*(6), 510–519. https://doi.org/10.1598/RT.58.6.2

Protopapas, A. (2006). On the use and usefulness of stress diacritics in reading Greek. *Reading and writing*, *19*(2), 171–198. https://doi.org/10.1007/s11145-005-4023-z

Rasinski, T. (2004). Creating fluent readers. *Educational leadership, 61*(6), 46–51. https://www.ascd.org/el/articles/creating-fluent-readers

Rasinski, T. (2006). Reading fluency instruction: Moving beyond accuracy, automaticity, and prosody. *The Reading Teacher*, *59*(7), 704–706. https://doi.org/10.1598/RT.59.7.10

Rasinski, T., Rikli, A., & Johnston, S. (2009). Reading fluency: More than automaticity? More than a concern for the primary grades? *Literacy Research and Instruction, 48*(4), 350–361. https://doi.org/10.1080/19388070802468715

Rasinski, T. V., & Hoffman, J. V. (2003). Oral reading in the school literacy curriculum. *Reading Research Quarterly*, *38*(4), 510–522. https://doi.org/10.1598/RRQ.38.4.5

Rasinski, T. V., Reutzel, D. R., Chard, D., & Linan-Thompson, S. (2011). Reading fluency. In M. L. Kamil, P. D. Pearson, E. Birr Moje, & P. P. Afflerbach (Eds.), *Handbook of reading research, Volume IV* (pp. 286–319). Routledge. https://doi.org/10.4324/9780203840412

Rastle, K., & Coltheart, M. (2000). Lexical and nonlexical print-to-sound translation of disyllabic words and nonwords. *Journal of Memory and Language*, *42*(3), 342–364. https://doi.org/10.1006/jmla.1999.2687

Resnick, L. B. (1970). Relations between perceptual and syntactic control in oral reading. *Journal of Educational Psychology*, *61*(5), 382–385. https://doi.org/10.1037/h0029814

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]. A computerized test. Hogrefe.

Ritter, C. (2006). *Entwicklung und empirische Überprüfung eines Lesetrainings auf Silbenbasis* [Development and empirical testing of a reading training based on syllables]. Unpublished doctoral dissertation. Universität Potsdam. https://nbn-resolving.org/urn:nbn:de:kobv:517-opus-10032

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

Samuels, S. J. (1979). The method of repeated readings. *The Reading Teacher*, *32*(4), 403–408.

Samuels, S. J. (2007). The DIBELS tests: Is speed of barking at print what we mean by reading fluency? *Reading Research Quarterly*, *42*(4), 563–566. https://doi.org/10.1598/RRQ.42.4.5

Schwanenflugel, P. J., Hamilton, A. M., Kuhn, M. R., Wisenbaker, J. M., & Stahl, S. A. (2004). Becoming a fluent reader: Reading skill and prosodic features in the oral reading of young readers. *Journal of Educational Psychology*, *96*(1), 119–129. https://doi.org/10.1037/0022-0663.96.1.119

Seymour, P. H., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*(2), 143–174. https://doi.org/10.1348/000712603321661859

Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology*, *72*(2), 95–129. https://doi.org/10.1006/jecp.1998.2481

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*(4), 267–298. https://doi.org/10.1016/j.jecp.2004.01.001

Shinn, M. R., Good, R. H., Knutson, N., Tilly, W. D., & et al. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review*, *21*(3), 459–479. https://doi.org/10.1080/02796015.1992.12085629

Selting, M., Auer, P., Barden, B., Bergmann, J., Couper-Kulhen, E., Günther, S., Quasthoff, U., Meier, C., Schlobinski, P., & Uhmann, S. (1998). Gesprächsanalytisches

Transkriptionssystem (GAT) [Discourse and conversation-analytic transcription system]. *Linguistische Berichte, 173*, 91–122.

Silverman, K. E., Beckman, M. E., Pitrelli, J. F., Ostendorf, M., Wightman, C. W., Price, P. J., Pierrehumbert, J. B., & Hirschberg, J. (1992). TOBI: A standard for labeling English prosody. In *Proc. 2nd International Conference on Spoken Language Processing (ICSLP 1992)*, 867–870.

Silverman, R. D., Speece, D. L., Harring, J. R., & Ritchey, K. D. (2013). Fluency has a role in the simple view of reading. *Scientific Studies of Reading*, *17*(2), 108–133. https://doi.org/10.1080/10888438.2011.618153

Snowling, M. J., & Hulme, C. (Eds.). (2005). *The science of reading: A handbook*. Blackwell Publishing. https://doi.org/10.1002/9780470757642

Staatsinstitut für Schulqualität und Bildungsforschung. (n.d.). Grundlegende Kompetenzen zum Ende der Jahrgangsstufe 2 [Basic skills at the end of grade 2]. Retrieved June 3, 2022, from https://www.lehrplanplus.bayern.de/jahrgangsstufenprofil/grundschule/2

Suggate, S. P. (2010). Why what we teach depends on when: Grade and reading intervention modality moderate effect size. *Developmental Psychology*, *46*(6), 1556–1579. https://doi.org/10.1037/a0020612

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*(1), 77–96. https://doi.org/10.1177/0022219414528540

Tarelli, I., Valtin, R., Bos, W., Bremerich-Vos, A., & Schwippert, K. (2012). IGLU 2011: Wichtige Ergebnisse im Überblick [IGLU 2011: Important results at a glance]. In W. Bos, I. Tarelli, A. Bremerich-Vos, & K. Schwippert (Eds.), *IGLU 2011 Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich* (pp. 11–26). Waxmann Verlag.

Temperley, D. (2009). Distributional stress regularity: A corpus study. *Journal of Psycholinguistic Research*, *38*(1), 75–92. https://doi.org/10.1007/s10936-008-9084-0

Thaler, V., Ebner, E. M., Wimmer, H., & Landerl, K. (2004). Training reading fluency in dysfluent readers with high reading accuracy: Word specific effects but low transfer to untrained words. *Annals of Dyslexia*, *54*(1), 89–113. https://doi.org/10.1007/s11881-004-0005-0

Torgesen, J. K., & Hudson, R. F. (2006). Reading fluency: Critical issues for struggling readers. In S. J. Samuels & A. E. Farstrup (Eds.), *What research has to say about fluency instruction* (pp. 130–158). International Reading Association.

Van Orden, G. C., & Kloos, H. (2005). The question of phonology and reading. In *The science of reading: A handbook.* (pp. 61–78). Blackwell Publishing. https://doi.org/10.1002/9780470757642.ch4

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*(1), 2–40. https://doi.org/10.1046/j.0021-9630.2003.00305.x

Wallot, S. (2017). Recurrence quantification analysis of processes and products of discourse: A tutorial in R. *Discourse Processes*, *54*(5-6), 382–405. https://doi.org/10.1080/0163853X.2017.1297921

Wallot, S., O'Brien, B. A., Haussmann, A., Kloos, H., & Lyby, M. S. (2014). The role of reading time complexity and reading speed in text comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(6), 1745–1765. https://doi.org/10.1037/xlm0000030

Wolters, A. P., Kim, Y.-S. G., & Szura, J. W. (2022). Is reading prosody related to reading comprehension? A meta-analysis. *Scientific Studies of Reading*, *26*(1), 1–20. https://doi.org/10.1080/10888438.2020.1850733

Yildiz, M., Yildirim, K., Ates, S., Rasinski, T., Fitzgerald, S., & Zimmerman, B. (2014). The relationship between reading fluency and reading comprehension in fifth-grade turkish students. *International Journal of School & Educational Psychology*, *2*(1), 35–44. https://doi.org/10.1080/21683603.2013.854187

Zarić, J., & Nagler, T. (2021). Reading comprehension on word- and sentence-level can be predicted by orthographic knowledge for German children with poor reading proficiency. *Reading and Writing*, *34*(8), 2031–2057. https://doi.org/10.1007/s11145-021-10126-9

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, *131*(1), 3–29. https://doi.org/10.1037/0033-2909.131.1.3

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*(3), 369–373. https://doi.org/10.1016/j.bandl.2004.10.010

List of Tables

Study 1

Study 2

Study 3

List of Figures

Study 1

Study 2

Study 3

Appendix

Appendix A: Material for Study 1

Due to copyright reasons, we cannot list the items in Study 1 or make them publicly available in the form of an online supplement. The items that were used in Study 1 are part of published tests and therefore the copyright is with the publisher.

Word recognition was measured with the test *Prozessbezogene Diagnostik von Lesefähigkeiten im Grundschulalter (ProDi-L)*. Commercially available under the following link https://www.testzentrale.de/shop/prozessbezogene-diagnostik-von-lesefaehigkeiten-im-grundschulalter.html

Reading comprehension was measured in the first cohort with the test *Ein Leseverständnistest für Erst- bis Sechstklässler (ELFE 1-6)*. Commercially available under the following link https://www.testzentrale.de/shop/ein-leseverstaendnistest-fuer-erst-bis-sechstklaessler.html

Reading comprehension was measured in the second cohort with the test *Ein Leseverständnistest für Erst- bis Siebtklässler – Version II (ELFE II)*. Commercially available under the following link https://www.testzentrale.de/shop/ein-leseverstaendnistest-fuer-erst-bis-siebtklaessler.html

Children received the word reading intervention *Lesen mit Willy Wortbär - Ein silbenbasiertes Training zur Förderung der Worterkennung beim Lesen*. Commercially available under the following link https://www.testzentrale.de/shop/lesen-mit-willy-wortbaer-91768.html

Appendix B: Material for Study 2

Due to copyright reasons, we cannot list the items in Study 2 or make them publicly available in the form of an online supplement. The items that were used in Study 2 are part of published tests and therefore the copyright is with the publisher:

Word recognition was measured with the test *Prozessbezogene Diagnostik von Lesefähigkeiten im Grundschulalter (ProDi-L)*. Commercially available under the following link https://www.testzentrale.de/shop/prozessbezogene-diagnostik-von-lesefaehigkeiten-im-grundschulalter.html

Reading comprehension was measured in the first cohort with the test *Ein Leseverständnistest für Erst- bis Sechstklässler (ELFE 1-6)*. Commercially available under the following link https://www.testzentrale.de/shop/ein-leseverstaendnistest-fuer-erst-bis-sechstklaessler.html

Appendix C: Material and Online Supplement Study 3

Table C1

Intercorrelations for all Important Variables for Grade 2 by Group

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Determinism of amplitude	–	.31	.58**	.43*	.74**	-.56**	-.04	-.17	-.05	-.10	.03	-.31	-.04	-.13	.26	.32	.30
2. Entropy of amplitude	.58**	–	-.01	.84**	.24	.11	-.05	-.12	-.03	-.19	-.14	-.06	.43*	-.12	.01	.03	.20
3. Laminarity of amplitude	.39*	-.04	–	.07	.69**	-.73**	.11	-.05	.13	.02	.33	-.32	-.06	.13	.04	.14	.29
4. Maximal line length of amplitude	.69**	.79**	.16	–	.32	.08	-.05	-.18	.06	-.26	-.11	-.02	.54**	.07	-.02	.02	.34
5. Recurrence rate of amplitude	.52**	.22	.45**	.24	–	-.73**	-.05	-.12	.11	-.07	-.01	-.40*	-.04	-.08	.12	.37*	.38*
6. Recurrence interval of amplitude	-.18	-.03	-.37*	-.00	-.63**	–	-.04	.14	-.02	-.08	.08	.37*	.43*	.16	-.16	-.23	.00
7. Determinism of pitch	.06	-.08	-.08	.17	-.12	.16	–	.42*	.64**	.87**	.22	.32	.14	.36*	-.48**	.06	.11
8. Entropy of pitch	-.24	-.07	-.04	.04	-.06	.13	.50**	–	.66**	.41*	.32	.21	.21	.03	-.32	.11	.06
9. Maximal line length of pitch	.07	.02	.05	.32	-.06	.29	.66**	.78**	–	.61**	.10	.26	.45*	.33	-.43*	.11	.18
10. Recurrence rate of pitch	-.03	-.16	-.05	.04	-.03	-.04	.63**	.42*	.55**	–	.16	.36	-.02	.27	-.49**	.08	.01
11. Average pause time	.06	-.13	-.00	.13	-.29	.47**	.46**	.43*	.61**	.44**	–	-.03	.14	.06	-.18	.24	.33

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
12. Recurrence interval of pauses	.13	-.25	.00	.03	.10	.19	.52**	.27	.44**	.49**	.53**	–	.11	.30	-.42*	-.13	.20
13. Final lengthening	-.23	-.38*	-.15	-.16	-.02	.11	.17	.05	.19	.22	.08	.26	–	.21	-.19	-.03	.36
14. Final pitch	-.35*	-.20	-.12	-.19	-.20	.18	.25	.29	.16	.35*	.19	.30	.10	–	-.25	-.31	.31
15. Accentuation	.34	.11	.18	.16	.09	-.24	-.36*	-.41*	-.37*	-.35*	-.19	-.20	.05	-.52**	–	.15	-.14
16. Appropriate pauses	.18	-.20	.16	-.01	-.06	.04	.09	-.16	.10	.10	.26	.23	.14	-.09	.15	–	.27
17. Inappropriate pauses	-.16	-.10	-.01	.09	-.33	.48**	.40*	.43*	.47**	.30	.51**	.31	-.01	.35*	-.43*	-.03	–

Note. The results for the tutees ($n = 33$) are shown above the Diagonal. The results for the tutors ($n = 34$) are shown below the diagonal.

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

Table C2

Intercorrelations for all Important Variables for Grade 4 by Group

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Determinism of amplitude	–	.05	.15	.38*	-.15	.17	.35*	.19	.28	.31	-.03	.19	.13	.07	-.14	.05	.17
2. Entropy of amplitude	.24	–	-.18	.75**	.03	-.03	-.21	.13	.03	-.08	.06	-.19	.00	-.19	-.11	-.12	-.08
3. Laminarity of amplitude	-.02	.10	–	-.02	.13	-.15	-.02	-.07	-.05	.06	-.13	-.13	-.05	-.18	.10	.19	.07
4. Maximal line length of amplitude	.54**	.74**	.16	–	-.27	.29	.01	.10	.23	.08	.11	.08	-.01	-.02	-.08	-.00	.30
5. Recurrence rate of amplitude	.37*	.21	.28	.33	–	-.85**	-.07	-.17	-.26	-.06	-.35*	-.19	-.19	-.24	-.12	-.00	-.19
6. Recurrence interval of amplitude	-.37*	-.23	-.17	-.40*	-.60**	–	.05	.05	.19	.12	.35*	.25	.28	.23	-.00	-.07	.38*
7. Determinism of pitch	.08	.39*	.05	.22	-.03	-.06	–	.37*	.68**	.78**	-.21	.22	.04	.13	.08	.26	.00
8. Entropy of pitch	-.04	.01	.10	.08	-.07	-.12	.29	–	.71**	.25	-.14	-.06	.02	-.31	.30	.06	-.29
9. Maximal line length of pitch	-.11	.25	.10	.12	-.26	.12	.72**	.51**	–	.66**	-.10	.28	.04	-.18	.17	.08	.01
10. Recurrence rate of pitch	-.07	.19	.16	.02	.15	.07	.56**	.14	.49**	–	-.12	.37*	.21	.17	-.04	.35*	.11
11. Average pause time	.15	-.35*	-.13	-.29	.22	-.24	-.13	.07	-.20	-.10	–	-.08	.08	.14	-.01	-.33	.29

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
12. Recurrence interval of pauses	-.04	.17	.13	.16	.12	.19	.07	.02	.06	.03	-.34	–	.37*	.28	-.12	.17	.54**
13. Final lengthening	-.11	.02	.10	.11	.27	-.16	-.18	.41*	.01	.01	.18	-.05	–	.05	-.18	.02	.33
14. Final pitch	.14	.07	-.14	.04	.30	-.03	-.09	-.14	-.25	.01	.16	-.05	.08	–	.01	.19	.14
15. Accentuation	.26	.23	-.17	.34*	.07	-.11	.23	-.03	.22	.30	-.16	.12	-.33	.08	–	.11	-.36*
16. Appropriate pauses	.37*	.42*	.17	.47**	.26	-.18	.33	-.05	.36*	.43*	-.17	.05	-.08	.13	.51**	–	-.13
17. Inappropriate pauses	.29	.04	.17	.20	.35*	-.34*	.10	.31	.10	.06	.33	.10	.20	.20	.15	-.08	–

Note. The results for the tutees (n = 35) are shown above the Diagonal. The results for the tutors (n = 34) are shown below the diagonal.

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

Table C3

List With Variables of Recurrence Quantification Analysis

1	average amplitude of pronunciation - mean
2	average amplitude of pronunciation - median
3	average amplitude of pronunciation - average diagonal line length
4	average amplitude of pronunciation - determinism rate
5	average amplitude of pronunciation - diagonal recurrence entropy
6	average amplitude of pronunciation - laminarity
7	average amplitude of pronunciation - maximum diagonal line length
8	average amplitude of pronunciation - maximum vertical line length
9	average amplitude of pronunciation - recurrence rate
10	average amplitude of pronunciation - recurrence time entropy
11	average amplitude of pronunciation - recurrence time type 1 (including the time of the pattern)
12	average amplitude of pronunciation - standard deviation
13	average amplitude of pronunciation - recurrence time type 2 (minus the time of the pattern)
14	average amplitude of pronunciation - trapping time
15	average pitch of pronunciation - mean
16	average pitch of pronunciation - median
17	average pitch of pronunciation - average diagonal line length
18	average pitch of pronunciation - determinism rate
19	average pitch of pronunciation - diagonal recurrence entropy
20	average pitch of pronunciation - laminarity
21	average pitch of pronunciation - maximum diagonal line length
22	average pitch of pronunciation - maximum vertical line length
23	average pitch of pronunciation - recurrence rate
24	average pitch of pronunciation - recurrence time entropy
25	average pitch of pronunciation - recurrence time type 1 (including the time of the pattern)
26	average pitch of pronunciation - standard deviation
27	average pitch of pronunciation - recurrence time type 2 (minus the time of the pattern)
28	average pitch of pronunciation - trapping time
29	average duration of pause - mean
30	average duration of pause - median
31	average duration of pause - average diagonal line length
32	average duration of pause - determinism rate
33	average duration of pause - diagonal recurrence entropy
34	average duration of pause - laminarity
35	average duration of pause - maximum diagonal line length
36	average duration of pause - maximum vertical line length

37	average duration of pause - recurrence rate
38	average duration of pause - recurrence time entropy
39	average duration of pause - recurrence time type 1 (including the time of the pattern)
40	average duration of pause - standard deviation
41	average duration of pause - recurrence time type 2 (minus the time of the pattern)
42	average duration of pause - trapping time
43	pronunciations per second - mean
44	pronunciations per second - median
45	pronunciations per second - average diagonal line length
46	pronunciations per second - determinism rate
47	pronunciations per second - diagonal recurrence entropy
48	pronunciations per second - laminarity
49	pronunciations per second - maximum diagonal line length
50	pronunciations per second - maximum vertical line length
51	pronunciations per second - recurrence rate
52	pronunciations per second - recurrence time entropy
53	pronunciations per second - recurrence time type 1 (including the time of the pattern)
54	pronunciations per second – standard deviation
55	pronunciations per second - recurrence time type 2 (minus the time of the pattern)
56	pronunciations per second - trapping time

The data that support the findings of this study are available in the repository of the Open Science Framework (OSF) at
https://osf.io/ep4bw/?view_only=f44d044d7bce4030b50a7c6e41f17ce6

Due to copyright reasons, we cannot list all items in Study 3 or make them publicly available in the form of an online supplement. Some items that were used in Study 3 are part of published tests and therefore the copyright is with the publisher.

Reading comprehension was measured in the first cohort with the test *Ein Leseverständnistest für Erst- bis Sechstklässler (ELFE 1-6)*. Commercially available under the following link
https://www.testzentrale.de/shop/ein-leseverstaendnistest-fuer-erst-bis-sechstklaessler.html

Story for the second grade

Vorgehen vor dem Vorlesen: Audacity starten und Kind mit Headset ausstatten. Vorstellung des Testleiters und Aufgabenstellung: „Bitte lies den Text hier gut betont vor! Danach habe ich noch zwei Fragen an dich.“. Dem Kind das Textblatt vorlegen.

Story – The first two sentences and the last sentence were not included in the analyses:

Besuch für Basti

Basti hat einen Freund, einen Zwerg mit roter Mütze. Aber der Zwerg hat kein Zuhause. Basti will ihm helfen. Er zieht seine Wolldecke vom Bett und versucht, sie über den Tisch zu legen. "Na los, hilf mal!", fordert er den Zwerg auf. Zusammen schaffen sie es. "Eine gute Zwergenhöhle", sagt der Zwerg und schleppt die Sofakissen hinein. "Hier werde ich jetzt wohnen." Das freut Basti sehr.

Vorgehen nach dem Vorlesen: Nach dem Lesen: das Textblatt umdrehen „Vielen Dank, dass hast du gut gemacht. Mal sehen, ob du mir auch noch zwei Fragen beantworten kannst.

Questions to the story:

1. Warum will Basti dem Zwerg helfen?
2. Was benutzen Basti und der Zwerg, um die Zwergenhöhle zu bauen?

Story for the fourth grade

Vorgehen <u>vor</u> dem Vorlesen: Audacity starten und Kind mit Headset ausstatten. Vorstellung des Testleiters und Aufgabenstellung: „Bitte lies den Text hier gut betont vor! Danach habe ich noch zwei Fragen an dich.“. Dem Kind das Textblatt vorlegen.

Story – The first two sentences and the last sentence were not included in the analyses:

<u>Das Geräusch der Grille</u>

Eines Tages kam ein Indianer in die große Stadt, um seinen Freund zu besuchen. Gemeinsam gingen sie die Straße entlang, als der Indianer plötzlich sagte: "Bleib einmal stehen. Hörst du auch das Zirpen der Grille?" "Nein", sagt sein Freund. "Alles, was ich höre, ist das Hupen der Autos. Dein Gehör ist eben besser geschult als meines." Da holte der Indianer ein 50-Cent-Stück aus der Tasche und warf es aufs Pflaster. Die Leute auf der Straße drehten sich nach dem Klimpern um. "Siehst du", sagte der Indianer. "Es stimmt nicht, dass mein Gehör besser ist als deines. Wir hören nur stets das gut, worauf wir zu achten gewohnt sind."

Vorgehen <u>nach</u> dem Vorlesen: Nach dem Lesen: das Textblatt umdrehen „Vielen Dank, dass hast du gut gemacht. Mal sehen, ob du mir auch noch zwei Fragen beantworten kannst.

Questions to the story:

1. Was hört der Indianer?
2. Warum kann der Freund die Grille nicht hören?

Acknowledgements

Zuallererst möchte ich mich bei meinem Betreuer Prof. Dr. Tobias Richter für die Möglichkeit bedanken, promovieren zu können. Darüber hinaus hat er mich mit seinen konstruktiven Rückmeldungen und seiner methodischen Expertise unterstützt. Bei Prof. Dr. Wolfgang Lenhard und PD Dr. Peter Marx bedanke ich mich für die Bereitschaft, die Begutachtung meiner Dissertation zu übernehmen. Außerdem, ich danke dem Bundesministerium für Bildung und Forschung für die Förderung dieser Arbeit (FKZ 01GJ1402B, FKZ 01GJ0985, FKZ 01GJ0986, FKZ 01GJ1206A, FKZ 01GJ1206B, und FKZ 01GJ1004).

Mein Dank gilt auch den Koautorinnen und Koautoren Dr. Bettina Müller, Dr. Julia Schindler, Dr. Maj-Britt Haffmans, Prof. Dr. Johannes Naumann und Prof. Dr. Sebastian Wallot, die mich – ebenso wie Prof. Dr. Tobias Richter – durch wertvolles Feedback bei der Abfassung sowie bei diversen Revisionen der vorliegenden Manuskripte unterstützt haben. Danken möchte ich zudem allen Lehrkräften der Grundschulen, den Kindern und deren Eltern für ihre Bereitschaft, an den Studien teilzunehmen und unsere Forschung zu unterstützen.

Weiterhin möchte ich mich an dieser Stelle auch bei einigen Leuten bedanken, die mich an der eine oder andere Stelle im Laufe meines Lebens unterstützt haben. Bei meinem Vater Kostas, der immer an mich glaubte, und bei Familie Satin – Silvia, Dennis, Silke und Radka – für die wertvolle Unterstützung als ich nach Deutschland ausgewandert bin.

Schließlich gilt mein besonderer Dank meinem ehemaligen Kollegen Jun.-Prof. Dr. Jan Lenhart und meiner derzeitigen Projektkollegin Janina Heß, die mir stets mit Rat und Tat zur Seite standen, meinem Bruder Dimitris und seiner wundervollen Familie und meiner Freundin, die mich auch während schwieriger Phasen der Promotion immer unterstützt haben.