This postprint was originally published by the APA as:
Hasenäcker, J., & Schroeder, S. (2022). **Transposed and substituted letter effects across reading development: A longitudinal study**. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 48*(8), 1202-1218.
https://doi.org/10.1037/xlm0001064.

Transposed and substituted letter effects across reading development: A longitudinal study

Jana Hasenäcker[1,2] & Sascha Schroeder[1,3]

[1]Max Planck Institute for Human Development, Berlin, Germany

[2]University of Erfurt, Erfurt, Germany

[3]Georg August University of Göttingen, Göttingen, Germany

Author Note

Jana Hasenäcker, MPRG Reading Education and Development (REaD), Max Planck Institute for Human Development. Current affiliation: Seminar für Sprachwissenschaft, University of Erfurt, Erfurt, Germany.

Sascha Schroeder, MPRG Reading Education and Development (REaD), Max Planck Institute for Human Development. Current affiliation: Georg-Elias-Müller-Institut für Psychologie, Georg August University of Göttingen, Göttingen, Germany.

Correspondence concerning this article should be addressed to: Jana Hasenäcker, Universität Erfurt, Seminar für Sprachwissenschaft, Nordhäuser Straße 63, 99084 Erfurt, Germany. E-Mail: janahasenaecker@gmail.com.

Abstract

Reading development involves several changes in orthographic processing. A key question is, “how does the coding of letters develops in children learning to read”. Masked priming effects of transposition and substitution primes have been taken to index the importance of letter position and identity coding. Somewhat contradicting results for developing readers have led to opposing theories. Here, we present new evidence from a large longitudinal study with over 100 children from grade 2 to 4. We investigate three different issues concerning transposition and substitution priming: (i) comparing priming both against an all-different and an identity baseline, (ii) testing priming effects for nonword targets, and (iii) taking into account inter-individual differences in orthographic knowledge. The analyses of the longitudinal data show, respectively, (i) an increase of priming effects over development in comparison to both baselines, (ii) identity priming for nonword targets in the elementary school years, and (iii) an additional impact of orthographic knowledge on priming effects for word, but not nonword, targets that is similar to the effect of grade. Taken together, our examination suggests that letter identity is coded relatively strictly, whereas letter position is coded relatively flexibly already early in reading development for words, but not for nonwords. We discuss how this pattern fits with different developmental models of orthographic processing.

Reading is a skill that typically needs years of practice to become proficient. One core component that has to be mastered is orthographic processing. For this, the encoding of the identity and position of letters is fundamental. Given the numerous attempts to computationally model letter coding in skilled reading and the fact that word reading is an acquired skill, the development of letter identity and position processing in reading acquisition has gained comparatively little attention.

There is now vast evidence that proficient readers are tolerant to trasnpositions and subslitutions of letters within words. In the preceding sentence, for example, the position of two letters was exchanged in one word (*trasnpositions*) and one letter was substituted by another letter in a second word (*subslitutions*), but nevertheless the words can be read and understood without great difficulty. This is because the skilled reading system codes letter position and identity flexibly, as has been found in several studies (e.g., Forster, Davis, Schoknecht, & Carter, 1987; Perea & Lupker, 2003). Most studies testing flexibility of letter coding with skilled adult readers have used the masked priming technique in which a word is preceded by a briefly (~50ms) presented prime that (a) includes a transposed letter (TL: *strom-STORM*), (b) comprises a substituted letter (SL: *starm-STORM*), or (c) consists of all different letters (DF: *flinz-STORM*). Results consistently show faster lexical decision times to words following TL primes compared to both SL and DF primes in skilled adult readers (e.g., Kinoshita, Castles, & Davis, 2009; Lupker & Davis, 2009; Lupker, Perea, & Davis, 2008; Perea & Lupker, 2003; Perea & Lupker, 2004). However, it is less clear how letter position and identity processing develop during reading acquisition and what drives this development.

Several studies using masked priming have been conducted with children to unravel the developmental trajectory of letter position coding (Table 1). Castles et al. (2007) were the first to investigate letter coding with masked priming in English-speaking 3rd-graders, 5th-graders and adults, comparing TL and 1SL primes to DF control primes. They found significant effects of both TL and SL compared to DF controls in 3rd-graders, but only effects of TL were still present in 5th-graders and no effects were found in adults. This prompted the authors to conclude that, throughout development, the reading system gets tuned towards more precise orthographic representations. This has given rise to the Lexical Tuning Hypothesis (Castles et al., 2007), which posits that the reading system allows for some degree of uncertainty in letter identity and position at the beginning of reading development because the representations of words in the orthographic lexicon are still underspecified. The more words children acquire in their orthographic lexicon, the more fine-tuned the representations need to become in order to distinguish between orthographic neighbors, such as *cat* and *hat* or *trial* and *trail*. The Lexical Tuning Hypothesis received support by two more studies that found decreasing TL effects across development, albeit the trajectories were slightly shifted depending on the language under investigation. In Spanish, Acha and Perea (2008) found substantial TL effects (as compared to SL primes) in beginning readers that decreased in size, but were still significant even in adults. In French, Lété and Fayol (2013) found no priming in 3rd-graders, TL and SL priming (compared to DF primes) in 5th-graders, and only TL priming in adults. In English, Kezilas, McKague, Kohnen, Badcock, and Castles (2017) showed TL effects both compared to SL and DF primes that did not change significantly across early, middle, and late primary school (Grade 2/3, Grade 4, Grade 5/6, respectively).

However, in a study with French 1st- to 5th-graders, Ziegler, Bertrand, Lété, and Grainger (2014) found the exact opposite developmental pattern, namely increasing TL effects (compared to SL primes) across grades. Based on this, the authors argue that orthographic processing becomes coarser throughout development. Converging evidence for increasing TL effects (compared to SL primes) was found in Italian 2nd-, 3rd-, and 5th-graders (Colombo, Sulpizio, & Peressotti, 2019) and German 2nd and 4th graders using eye tracking (Tiffin-Richards & Schroeder, 2015). Those findings are in line with the multiple-route model of reading (Grainger & Ziegler, 2011), which assumes that letter processing is position-specific in the beginning of reading acquisition and becomes more flexible throughout development. According to this model, the exact positions of letters are important as long as children decode written words phonologically by translating each letter into its corresponding sound. During development, children move away from this phonological procedure and rely increasingly on orthographic processing without the need to convert letters into sounds. The model assumes two orthographic routes: a fine-grained one that uses location-specific coding of letter sequences and a coarse-grained one that operates on non-contiguous location-invariant bigrams. For example, the word *STORM* would be coded by the non-contiguous open bigrams *ST, SO, SR, SM, TO, TR, TM, OR, OM, RM* in the coarse-grained route (cf., Grainger & van Heuven, 2003; Whitney, 2001). Hence, according to the multiple-route model, increased reliance on the coarse-grained route in development leads to increased letter coding flexibility.

Table 1. *Overview over studies investigating letter position priming effects in children.*

Authors	Language	Age groups	Prime types
Castles, Davis, Cavalot, & Forster (2007)	English	Grade 3, Grade 5, Adults	DF, 1SL, TL
Acha & Perea (2008)	Spanish	Beginning (~7 yrs.), Intermediate (~11 yrs.), Adults	2SL, TL
Lété & Fayol (2013)	French	Grade 3, Grade 5, Dyslexics, Adults	DF, 1SL, TL

Ziegler, Bertrand, Lété, & Grainger (2014)	French	Grades 1, 2, 3, 4, 5	2SL, TL
Kezilas, McKague, Kohnen, Badcock, & Castles (2017)	English	Early (Grade 2/3), Middle (Grade 4), Late (Grade 5/6), Adults	DF, 2SL, TL, ID
Colombo, Sulpizio, & Peressotti (2019)	Italian	Grades 2, 3, 5, Adults	2SL, TL

Note: DF = all letters different, 2SL = two substituted letters, 1SL = one substituted letter, TL = transposed letters, ID = identity

Importantly, the masked priming lexical decision studies with children shown in Table 1 differ from each other in terms of participants (i.e., varying age groups, sometimes further confounded with different onsets of schooling) and procedures (i.e., forward mask vs. sandwich priming, priming times varying from 50 to 70ms). Another potential source for differences in the developmental trajectory, pointed out by several authors, is the transparency of the respective language. Finally, varying control primes (DF, 2SL, 1SL, ID) further complicate cross-experiment comparisons. Consequently, how letter position and identity processing develop during reading acquisition, and what drives this development, is still unsolved.

The present study takes a longitudinal approach to elucidate how the developing reading system deals with letter coding. We present longitudinal data from a group of 100 German children that were tested at the end of grade 2, 3 and 4. The longitudinal nature goes beyond previous studies by truly tracking developmental changes in the same individuals, thus avoiding some problems of cross-sectional studies and focusing with increased power on the grades during which previous studies found the most striking changes in priming effects. Moreover, the study adds to previous ones by addressing three particularly important issues of letter position and identity coding in development. The first issue concerns the baseline to which TL priming effects should be compared; an issue raised by Kezilas, et al. (2017) that is still underexplored, but crucial for the interpretation of the effects. The second issue relates to priming effects in nonwords, which have been largely disregarded in previous studies, but can provide additional insights especially with regard to the locus of the effects, that is, whether it has a lexical or prelexical origin (cf. Mousikou, Kinoshita, Wu, & Norris, 2015). The third issue bears on the role of interindividual differences in specific skills for which grade is often used as an index, but which might have independent effects on the developmental trajectory (cf. Hasenäcker, Beyersmann, & Schroeder, 2020, for similar reasoning concerning morphological priming). Each of these issues will be discussed in detail.

Comparison baseline

Both adult and child studies on TL priming effects have used different control conditions. In order to pit letter position and letter identity against each other, SL primes have been most commonly used, replacing two letters such that the overlap of correctly positioned letters across prime and target is equal in both TL and SL primes (e.g., the pairs ***strom-STORM*** and ***stacm-STORM*** each have three letters that share both identity and position). From a developmental perspective, the TL-SL comparison is not very informative, because a change in this difference across development could be a result of changes either in the processing of transpositions (i.e., changes in letter position coding) or in the processing of substitutions (i.e., changes in letter identity coding) or an interaction between them (see Kezilas et al., 2017, for an in-depth discussion).

Some studies have additionally made use of DF primes, in which all letters of the prime differ from the target. Thus, the letter overlap between DF prime and target amounts to zero and the TL prime can be compared to a null-priming baseline. In this case, priming from TL primes compared to priming from DF primes can be seen as a *benefit* effect: the benefit from partially overlapping letters (***str**om-**STOR**M* or ***st**a**rm-ST**O**RM***) in comparison to no overlapping letters (*flinz-STORM*). Note that *benefit* here refers to the faster processing of a target preceded by a partially overlapping prime compared to the processing of a target preceded by a non-overlapping prime. Of course, within one type of prime, there can still be both beneficial and detrimental mechanisms at work. For example, the shared letters between TL prime and target can be beneficial, while at the same time the misplaced letters can be detrimental (cf. Peressotti & Grainger, 1999). Thus, from a developmental perspective, it is difficult to tell whether a change in the TL-DF comparison is due to a change in processing of transpositions per se or due to a change in the use of the remaining correctly-positioned shared letters (***str**om-**STOR**M*) across prime and target. It could either be that transposed letters become more or less detrimental for processing or, alternatively, that correctly positioned overlapping letters become more efficiently used throughout reading development.

Kezilas et al. (2017) argued in favor of another baseline prime, namely ID, which is the exact same word as the target (***storm-STORM***). This allows one to conceptualize the transposition of letters as a *cost* effect in comparison to the actual word. *Cost* here refers to the slower processing of a target preceded by a partially overlapping prime compared to the processing of a target preceded by a fully overlapping prime. Again, this does not deny the possibility that there can be both beneficial and detrimental mechanisms at work within a certain type of prime, as explained above for the TL prime. The key asset of using an ID baseline is that the only difference between the ID and TL prime is the position of letters, whereas the identity of letters is fully preserved: the primes *strom* and *storm* both contain the same five letters as the target *STORM*, but differ in whether these are in the correct position or not. Such a comparison is therefore much more straight-forward when testing the development of letter position coding.

Crucially, the lexical tuning hypothesis and the multiple route model make divergent developmental predictions for the TL-DF and the TL-ID comparisons (cf. Kezilas et al., 2017). The lexical tuning hypothesis predicts that the TL-DF difference decreases, while the TL-ID difference increases as letter position coding becomes more precise throughout development. By contrast, the multiple route model predicts that the TL-DF difference increases as letter position coding becomes coarser throughout development. The model does not make explicit assumptions about the development of the TL-ID difference in its original version. However, one key principle of the model is that the developing reading system increasingly relies on coarser location-invariant orthographic coding using open bigrams (cf. Grainger & Ziegler, 2011). Because the open bigrams activated by TL and ID primes are largely similar, the reliance on this route with increasing reading experience would suggest that the difference between TL and ID primes becomes smaller. Tracking both TL-ID and TL-DF differences across reading development thus provides a preliminary test of the two opposing hypotheses regarding letter position coding. In the present study, we therefore follow Kezilas et al. (2017) in taking into account the different baselines jointly.

In order to gain a full picture of letter coding throughout reading development, substituted letter priming needs to be investigated in comparison to both ID and DF baseline primes. In most previous studies, including the one by Kezilas et al. (2017), SL primes have been solely used as a comparison baseline to TL primes. In these studies, substitutions of two letters (2SL) have been most commonly used in order to keep the number of position-locked letter overlap equal across primes (e.g., 3/5: ***st**al**m**-**STORM**, **st**ro**m**-**STORM***), while changing the number of position-independent letter overlap (3/5: ***stalm-STORM*** vs. 5/5: ***strom-STORM***). As discussed above, however, this comparison is not fully able to disentangle developmental changes of letter identity vs. position coding. Consequently, in our study, we compared SL priming to both ID and DF controls. For this specific case, we reasoned that a substitution of only one letter (1SL) would be more insightful. 1SL primes have higher overlap of position-locked letters with the ID prime than TL primes do (4/5: ***starm-STORM*** vs. 3/5 ***strom-STORM***), whereas they have lower overlap of position-independent letters (4/5: ***starm-STORM*** vs. 5/5 ***strom-STORM***). Thus, this creates a very useful asymmetry between the 1SL and the TL prime in comparison to ID, rendering it a stronger test case for letter identity vs. position coding.

The lexical tuning hypothesis assumes that letter identity, like letter position, becomes more precise throughout development. In fact, the seminal formulation of this hypothesis was based on the finding that the 1SL-DF difference decreases throughout development (Castles et al., 2007), demonstrating that readers are increasingly more sensitive to letter identity. Therefore, in a similar fashion, the 1SL-ID difference can be expected to increase across development in the lexical tuning framework, just like TL-ID does. However, this has not been explicitly tested so far, as Kezilas et al. (2017) only report the TL-ID and TL-SL comparison, but not the 1SL-ID comparison. The multiple-route model does not make any explicit assumption about the development of substituted letter priming. However, the coarse-grained route, which is supposedly used by experienced readers, operates on open bigrams. It should thus be flexible in terms of letter position, but less so in terms of letter identity coding. Consequently, this model would most likely not predict that the SL-ID difference decreases similarly to the TL-ID difference.

Taken together, using both a DF and an ID prime as comparison baselines for TL and SL priming, instead of pitting the latter against each other, can provide a more comprehensive picture of the changes of letter position and identity coding in reading development. Hence, in the present study, we adopt this approach.

Nonword Priming

With regard to the models discussed so far, the lexical tuning hypothesis explicitly assumes that changes in priming effects have a lexical origin in that they are driven by the development of the orthographic lexicon. In this framework, lexical representations become more precise with orthographic lexicon growth in order to reduce confusability between orthographic neighbors. Importantly, *nonwords* do not have corresponding representations in the orthographic lexicon that could become more fine-tuned. Thus, in the framework of the lexical tuning hypothesis, no priming effects for nonword targets or developmental changes thereof are predicted as a direct consequence of the growing lexicon. Note that this does not rule out changes at the visual processing level under this hypothesis. However, those would not be directly driven by the growth and fine-tuning of the orthographic lexicon that are at the heart of the hypothesis.

In the multiple route model, there are two different procedures for letter processing: serial decoding to translate letters into phonemes (sublexical route) and parallel activation of letters to activate orthographic representations (lexical route). The latter type of processing is based on open bigram coding (e.g., Grainger & van Heuven, 2003; Whitney, 2001) using non-contiguous letter pairs. TL priming effects in this route arise based on overlapping open bigrams between prime and target. The sublexical route does not operate on the basis of open bigrams, but needs ordered letter sequences as an input to generate a sequence of ordered phonemes (e.g., Grainger & van Heuven, 2003; Grainger & Ziegler, 2011; Whitney & Cornelissen, 2008). Therefore, Mousikou et al. (2015) have argued that TL priming effects should not be expected for nonword targets, at least when the task is reading aloud. Similarly, the model would predict that beginning readers show no or reduced TL priming effects for words when they mainly rely on the sublexical route in word recognition. However, some amount of priming of nonwords can arise through the coarse-grained route, at least in lexical decision. Primes that partly or completely overlap with the nonword target could pre-activate open bigrams, which could give a head-start to the decision of whether the encountered item is a word (i.e., matches a whole-word orthographic presentation, to put it in the terms of the model) or not.

Some researchers have also proposed that the presence of TL effects for nonwords suggests that letter flexibility arises at the level of visual perception (e.g., Gomez, Ratcliff, & Perea, 2008; Kinoshita & Norris, 2009; Norris & Kinoshita, 2012), rather than at an orthographic or lexical level. The Overlap model (Gomez et al., 2008) and the noisy channel model (Norris & Kinoshita, 2012; see also Norris, Kinoshita, & van Casteren, 2010) suggest that TL priming effects are a consequence of perceptual noise very early in processing that results in an underspecified prelexical representation that is used as the input for both the lexical and sublexical route. Due to this very early locus, letter transposition and substitution primes should affect nonwords similarly to words. These models, however, make no assumptions about development and have thus been largely neglected in the developmental literature on letter coding (for an exception see Perea, Jiménez, & Gomez, 2016). Important changes affecting the magnitude, precision, and invariance at those lower visual-perceptual levels can, however, be assumed on the basis of neuroimaging studies that show a specialization for letter processing in the visual cortex with reading acquisition (e.g., Dehaene, Cohen, Morais, & Kolinsky, 2015; Pegado et al., 2014; Szwed et al., 2011).

It is surprising that only two studies so far have investigated TL priming effects in nonwords across development. Both Acha and Perea (2008) and Ziegler et al. (2014) found TL compared to SL priming for nonwords that was similar in magnitude for all age groups considered. However, both studies solely compared TL and 2SL primes against each other without a DF or ID baseline. As established before, this can be problematic. Other developmental studies have neglected priming of nonwords entirely. However, TL priming of nonwords has been a hotly debated issue in the adult literature, especially with regard to the locus of the effect, that is, whether letter flexibility happens at a visual-perceptual or orthographic-lexical level of processing. Investigating the development of letter position and identity priming of nonword targets thus can provide interesting additional insights into the locus and nature of TL and SL priming effects.

Orthographic Knowledge

For the development of letter position coding, the lexical tuning hypothesis explicitly predicts that orthographic knowledge is a crucial driving skill (Castles et al., 2007; Marinus, Kezilas, Kohnen, Robidoux, & Castles, 2018). According to this hypothesis, the pressure for more precise representations is higher the more words children have in their sight word vocabulary. At the beginning of reading acquisition, the lexical identification system can afford to be rather broadly-tuned. As children’s orthographic lexicon grows, the system must adapt by fine-tuning the representations in the orthographic lexicon to minimize confusion between similarly spelled words (e.g., *form*, *from*, *farm*).

While grade can be a proxy for the development of the orthographic lexicon (over grades, the orthographic lexicon should grow), it is not a direct measure of it. In addition, variability within grades is typically as high or even higher than between grades. Hence, the assumption of reading research that development happens uniformly across grades is too simplistic. For example, in the context of morphological processing, evidence accumulates that the developmental trajectory is modulated not only by grade, but also by reading proficiency and vocabulary knowledge (Beyersmann, Grainger, Casalis, & Ziegler, 2015; Hasenäcker et al., 2020; Hasenäcker, Schröter, & Schroeder, 2017). This underlines the importance of examining the influence of interindividual variability in orthographic knowledge on the development of visual word processing and letter coding. A stronger test of the lexical tuning hypothesis would therefore be to measure inter-individual differences in orthographic knowledge and systematically examine its effect on the development of transposed and substituted letter priming beyond grade level. A few studies have successfully taken a similar approach with regard to inter-individual differences in orthographic knowledge on priming in adult readers (e.g., Adelman et al., 2014; Andrews & Hersch, 2010; Andrews & Lo, 2012; Welcome & Trammel, 2017). However, to our knowledge, no study has attempted to more directly measure the impact of differences in orthographic knowledge on the development of priming effects. Some previous studies have, however, related differences in reading ability – rather than orthographic knowledge – to priming effects in children (e.g., Colombo et al., 2019; Ziegler et al., 2014). Both Ziegler et al. (2014) and Colombo et al. (2019) have correlated the size of TL effects to measures of reading ability. In a recent study, Gomez, Marcet, and Perea (2021) showed that interindividual differences in reading ability among 6^{th}-graders modulated TL effects in a lexical decision task without priming: better readers were less likely to confuse TL nonwords (*mohter*) with real words (*mother*).

Although reading ability and orthographic knowledge tend to correlate, they are not identical, especially in a transparent language like German, where high word reading accuracy can be achieved based on a phonological decoding procedure. To test the specific prediction of the lexical tuning hypothesis, namely that orthographic knowledge is the driving force for changes in letter coding, we took orthographic knowledge into account. Furthermore, we aimed to go beyond a correlation of orthographic knowledge and priming at one single point in time by investigating this relationship at several time points. Our longitudinal design and large sample size present an ideal test case for this.

There is no direct pure measure of orthographic knowledge and measuring it is especially challenging in a rather transparent orthography like German. However, some tasks can be considered to capture different aspects of orthographic knowledge. First, correct spelling requires the recourse to orthographic representations, although this is less so in transparent orthographies, in which many, but not all, words can also be spelled by drawing on letter-sound regularities. Nevertheless, spelling is most commonly interpreted as a measure of the quality of the orthographic lexicon and weak spelling skills have been associated with increased letter-position uncertainty in adults (Marinus et al., 2018). Second, the difference between word and nonword reading aloud is often assumed to be an indicator of lexical development, as both words and nonwords can be read using phonological decoding, but only words can be also read via direct orthographic recognition (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). Advantages of word over nonword reading thus stem not only from existing phonological, but – importantly – also from existing orthographic representations. Thirdly, vocabulary and orthographic knowledge are closely related: only children who know many words can also have orthographic representations for many words. Especially in a language in which most words can be read correctly via phonological decoding and similar phonology equates to similar orthography, vocabulary might be a good proxy for some aspects of orthographic knowledge.

Hence, in the present study, we tracked several reading-related skills in our participants in each grade and used a combined measure of spelling, word and nonword reading difference, and vocabulary in order to encompass different aspects of orthographic knowledge and thus capture the concept as precisely as possible. Taking into account interindividual differences in the growth of the orthographic lexicon in addition to the number of school years, allows us to test whether lexical tuning drives changes in letter coding during development.

The present study

In the following, we present a longitudinal study conducted with children from grade 2 to 4 to investigate the developmental trajectory of letter position coding. To this end, we used a masked priming lexical decision task with four prime conditions, namely DF, 1SL, TL, and ID. In the analysis, we compare the development of priming effects across the elementary school years in words and, importantly, also in nonwords. Additionally, the influence of orthographic knowledge on the developmental trajectory of letter coding is tested.

Our study is unique in its truly developmental approach. Instead of testing small groups of children from different grades, we followed the development of a large group of children covering the normal variety of reading skills and orthographic knowledge. Alongside the transposed and substituted letter priming effects, we tracked the development of several reading related skills over the years, allowing us to directly investigate the contribution of orthographic knowledge in addition to grade as a driving factor of development.

Method

Participants

The study was part of the longitudinal project *Orthographic Processing in Reading Acquisition (OPeRA)* in which children from four elementary schools in the Berlin area participated. For the present study, data from 104 children (57 girls, 47 boys) was available from three testing points in which the first one took place at the end of grade 2, the second exactly one year later at the end of grade 3 and the third another year later at the end of grade 4. All testing sessions took place in the schools during regular school hours. The study was approved by the ethics committee of the Max Planck Institute for Human Development as well as the local school board. Written consent for participation was obtained from the parents prior to the beginning of the project and oral consent was asked from each child at the beginning of each testing session.

Materials

Forty-eight, German nouns were chosen from the childLex corpus (Schroeder, Würzner, Heister, Geyken, & Kliegl, 2014) as target words. All target words were 5 letters long and one-syllabic and had a mean normalized log10 lemma frequency of 1.33 (SD=0.57). For each target word, four primes were created: (1) an identity prime (ID) that was exactly the same as the target word (e.g., *sturm-STURM*), (2) a transposed letter prime (TL) in which two adjacent letters were swapped (e.g., *strum-STURM*), (3) a substituted letter prime (SL) in which one letter was substituted by another one (e.g., *starm-STURM*), and (4) an all-letter-different prime (DF) that was a pronounceable pseudoword not sharing any letters with the target (e.g., *flinz-STURM*). TL primes always involved transpositions between a consonant and a vowel that was always internal, either involving the second and third or the third and fourth letter. SL primes always involved the middle letter. Primes were matched as closely as possible on OLD20, bigram- and trigram frequency (all $t<2$, all $p>.05$).

In addition to the word targets, 48 pronounceable nonwords were created to serve as nonword targets. Again, all of them were 5 letters long and monoyllabic. For each nonword target, four primes were created, ID, TL, SL, DF, in the same way as described for the target words.

In order to avoid repeating targets within the experiment, we used a Latin square design with four lists, such that each participant saw each target once with one of the four primes. However, across participants, each target was presented in combination with all four primes.

Procedure

Participants were tested individually in a separate room in their school. The experiment was run on a 15” laptop monitor with a refresh rate of 60 Hz. All stimuli were presented on black background in white 20-point Courier New font. The sandwich priming technique (Lupker & Davis, 2009; see also Colombo et al., 2019) was used, because it has been shown to increase measurement sensitivity compared to standard masked priming and is therefore being increasingly preferred with children (Colombo, et al., 2019; Ziegler et al., 2014). Each trial started with a fixation cross (+) presented in the middle of the screen for 500ms, immediately followed by a target preview in lowercase letters (e.g., *sturm*) for 30ms, which was replaced by the prime in lowercase letters (e.g., *strum*) presented for 50ms. Then the target word followed in uppercase letters (e.g., *STURM*) and remained on the screen until the participant made a lexical decision by pressing either the *D* (marked red) or the *K* (marked green) key on a standard keyboard to indicate whether the word is a nonword or a word, respectively. The participants were instructed to indicate their answer as fast and as accurate as possible. They were not informed about the prime. Prior to the experiment, the participants conducted four practice trials with feedback. After that, they conducted the main experiment without feedback. After half of the trials, an experimenter-timed break was introduced.

In addition to the experimental task, reading-related skills were tested at each time point using standardized tasks. Spelling was tested using the Hamburger Schreibprobe (HSP; May, 2002). This is a standardized dictation task in which single words and short sentences are read aloud by the examiner and the children are asked to correctly write them down. Grade-specific versions (differing in difficulty and number of words) were used in accordance with each testing point. For each word, the number of correctly written graphemes was counted.

Word and nonword reading were tested using the one-minute reading tests of the Salzburger Lese- und Rechtschreibtest (SLRT-II; Moll & Landerl, 2010). In this standardized reading test, 156 words are presented in eight columns and the child's task is to accurately read out loud as many words as possible in one minute. The nonword part is exactly the same except that it contains pronounceable nonwords instead of words. The number of correctly read words/nonwords serves as the test score. The difference between word and nonword reading was calculated by subtracting the nonword test from the score in the word test.

Vocabulary was tested using the vocabulary part of the CFT 20-R (Weiß, 2006). This task is a pen-and-paper multiple-choice task: children had to select a synonym or hypernym for a word from 5 given alternatives (e.g. *eagle – nest, tree, bird, sparrow, sky*). The test encompassed 30 trials of increasing difficulty over trials and children were given 5 minutes to complete as many as possible. The number of correctly solved trials is used as the test score.

Analysis

Data analyses were carried out with the statistical software R (RDevelopmentCoreTeam, 2008) using linear mixed-effects models with the lme4 package (Bates, Mächler, Bolker, & Walker, 2015). Response times to word and nonword trials were analyzed separately. Five children were excluded from data analysis because they had mean accuracy values below 60% in at least one of the testing sessions. Prior to the response time analysis, data was cleaned for each grade separately by removing incorrect responses (words: 2nd grade 17.78%, 3rd grade 11.82%, 4th grade 10.54%; nonwords 2nd 14.71%, 3rd 10.61%, 4th 8.72%), response times below 200ms or above 6000ms (2nd), 5000ms (3rd), 4000ms (4th) (words: 2nd 1.77%, 3rd 0.89%, 4th 0.85%; nonwords: 2nd 2.47%, 3rd 3.17%, 4th 2.71%). For outlier trimming, model criticism based on a simple model including random effects for subject and item was used, excluding all data points with residuals exceeding 2.5 standard deviations for the main analyses (words: 2nd 2.01%, 3rd 2.42%, 4th 2.28%; nonwords: 2nd 1.79%, 3rd 1.71%, 4th 2.21%). Table 2 shows the mean response times after outlier trimming. Next, z-scores were computed for each participant in each grade by subtracting from each response time the mean response time divided by the standard deviation of the respective participant in the respective grade. Z-scores have been commonly used in similar studies (e.g., Kezilas et al., 2017; Ziegler et al., 2014) in order to ensure that the response time analyses were not confounded by differences in overall processing speed between grades (cf. Faust, Balota, Spieler, & Ferraro, 1999; Ziegler et al., 2014; Zoccolotti, De Luca, Di Filippo, Judica, & Martelli, 2008).

Table 2. *Mean raw Response Times (in Milliseconds) to Words and Nonwords in All Prime Conditions per Grade (Standard Errors in Parentheses) and Differences between Conditions (in Milliseconds).*

	DF	SL	TL	ID	Difference to DF		Difference to ID	
					SL	TL	SL	TL
			Words					
Grade 2	1961 (33)	1941 (34)	1858 (32)	1950 (35)	20	103	9	92
Grade 3	1382 (22)	1343 (22)	1316 (22)	1304 (23)	39	66	-39	-12
Grade 4	1037 (14)	1018 (15)	954 (14)	934 (14)	19	83	-84	-20
			Nonwords					
Grade 2	2720 (44)	2736 (47)	2671 (44)	2564 (43)	-16	49	-172	-107
Grade 3	1816 (28)	1807 (28)	1776 (27)	1750 (29)	9	40	-57	-26
Grade 4	1281 (19)	1253 (19)	1242 (19)	1244 (19)	28	39	-9	2

Then a linear mixed-effects model was fitted to the word and nonword response times separately with Prime Type (DF, SL, TL, ID) as a contrast-coded categorical variable, Grade (2^{nd}, 3^{rd}, 4^{th}) as a continuous variable (centered at grade 3). In addition, the model comprised Subject and Item as random intercepts. Results for the overall effects tests using contrast coding and Type III sum of squares (using the Anova function in the car package) are reported. To compare the different prime types against each other, post-hoc comparisons were carried out using cell means coding and single df contrasts with the glht function of the multcomp package (Hothorn et al., 2008) and were evaluated using a normal distribution. For completeness, we also ran all analyses with log-transformed response times as the dependent variable. Both z-score and log-transformed response times analyses showed the same overall pattern. The former are reported in the following, the latter are reported in the Appendix.

For the analysis of interindividual differences in orthographic knowledge, we used the measures of reading-related skills tested with standardized tests as described above: spelling (HSP), word and nonword reading (SLRT), and vocabulary (CFT-20R). The results of the standardized tests in each grade and their correlations are shown in Table 3. In order to obtain a measure of the orthographic knowledge within each grade, we performed a principal component analysis on the scaled and centered scores for each grade separately (cf. Andrews & Lo, 2012, for a similar approach with adult readers). We extracted the first principal component which accounted for 69% of variance in grade 2 and 3 and 68% in grade 4. All three tasks had similar intermediate factor loadings in all grades (grade 2: spelling 0.56, word-nonword reading difference 0.62, vocabulary 0.55; grade 3: spelling 0.56, word-nonword reading difference 0.59, vocabulary 0.58; grade 4: spelling 0.55, word-nonword reading difference 0.57, vocabulary 0.61). The factor scores for each participant in each grade served as a composite measure of Orthographic Knowledge and were used as a time-varying predictor in the linear mixed-effects model described above. Orthographic Knowledge was added as a fixed effect (scaled per grade) and in interactions with Prime Type and Grade.

Table 3. *Means (SDs and Ranges in Parentheses) of the Tests of Spelling, Word Reading, Nonword Reading, and Vocabulary and Correlations between the Measures per Grade.*

			Spelling	Word reading	Nonword reading	Vocabulary
Grade 2	Means (SD, Range)		0.86 (0.11, 0.43-1)	50.34 (20.12, 1-102)	31.14 (11.56, 4-72)	7.45 (2.93, 1-13)
	Correlations	Word reading	0.52	-	-	-
		Nonword read.	0.41	0.86	-	-
		Vocabulary	0.33	0.64	0.56	-
Grade 3	Means (SD, Range)		0.90 (0.07, 0.67-	68.48 (20.15, 25-	38.48 (13.13, 9-77)	14.71 (5.13, 1-27)

			0.99)	104)		
	Correlations	Word reading	0.56	-	-	-
		Nonword read.	0.42	0.82	-	-
		Vocabulary	0.41	0.63	0.45	-
Grade 4	Means (SD, Range)		0.93 (0.07, 0.49-1)	80.37 (20.99, 13-122)	44.62 (13.68, 9-82)	18.01 (5.15, 4-27)
	Correlations	Word reading	0.60	-	-	-
		Nonword read.	0.48	0.84	-	-
		Vocabulary	0.54	0.68	0.55	-

Results

Words. The response time analysis for word trials revealed a significant main effect of Prime Type ($\chi^2(3)=137.25$, $p<.001$). Post-hoc contrasts showed that all prime types differed significantly from each other except for the ID-TL comparison; DF-ID: $b=-.241$, $t=-9.86$, $p<.001$, SL-DF: $b=-.084$, $t=-3.41$, $p<.001$, TL-DF: $b=-.229$, $t=-9.40$, $p<.001$, SL-TL: $b=-.146$, $t=-6.02$, $p<.001$, SL-ID: $b=.157$, $t=6.48$, $p<.001$, TL-ID: $b=.011$, $t=0.47$, $p=.637$.

As response times were z-transformed within each grade, there was no main effect of Grade ($\chi^2(1)=0.10$, $b=-.003$, $t=-0.309$, $p=.758$). Crucially, however, the effect of Prime Type was modulated by Grade ($\chi^2(3)=31.88$, $p<.001$), indicating that the developmental trajectory differed for the different prime types. Decomposing this interaction by using post-hoc contrasts showed that the difference between DF and ID was already present in grade 2 and increased with increasing grade (grade 2: $b=-.082$, $t=-2.08$, $p=.037$; grade 3: $b=-.241$, $t=-9.86$, $p<.001$; grade 4: $b=-.399$, $t=-10.54$, $p<.001$). The SL-DF difference became significant from grade 3 onward (grade 2: $b=-.052$, $t=-1.33$, $p=.184$; grade 3: $b=-.084$, $t=-3.41$, $p<.001$; grade 4: $b=-.115$, $t=-3.01$, $p=.003$). The TL-DF difference was significant in grade 2 and increased over grades (grade 2: $b=-.162$, $t=-4.12$, $p<.001$; grade 3: $b=-.229$, $t=-9.40$, $p<.001$; grade 4: $b=-.297$, $t=-7.82$, $p<.001$).

Considering ID as a baseline, the SL-ID difference emerged in grade 3 and increased towards grade 4 (grade 2: *b*=.030, *t*=0.76, *p*=.449; grade 3: *b*=.157, *t*=6.48, *p*<.001; grade 4: *b*=.285, *t*=7.58, *p*<.001). The TL-ID difference reached significance in an unexpected direction in grade 2 (TL faster than ID) and was significant in the expected direction (ID faster than TL) in grade 4 (grade 2: *b*=-.080, *t*=-2.05, *p*=.040; grade 3: *b*=.011, *t*=0.47, *p*=.637; grade 4: *b*=.103, *t*=2.74, *p*=.006). Figure 1 shows the z-transformed response times as a function of Grade.

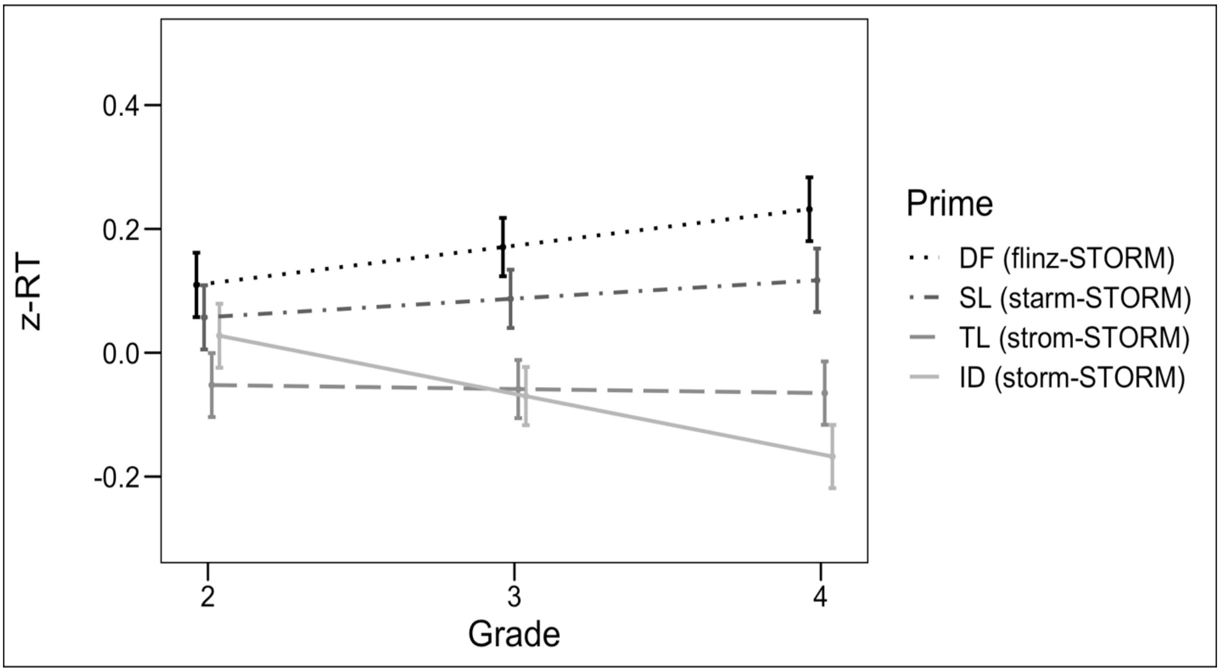

Figure 1. Mean predicted response times (z-transformed) in each prime condition for word targets as a function of grade.

Nonwords. The response time analysis for nonword trials revealed a significant main effect of Prime Type ($\chi^2(3)=20.69$, *p*<.001). Post-hoc contrasts decomposing this effect showed that DF and ID differed significantly from each other (*b*=-.039, *t*=-4.20, *p*<.001). Neither the SL-

DF nor the TL-DF difference were significant (SL-DF: *b*=-.005, *t*=-0.66, *p*=.507, TL-DF: *b*=-.011, *t*=-1.13, *p*=.259). However, both the SL-ID and the TL-ID difference was significant (SL-ID: *b*=.034, *t*=3.54, *p*<.001, TL-ID: *b*=.028, *t*=3.07, *p*=.002). The main effect of grade was not significant ($\chi^2(1)$=0.01, *b*=-.001, *t*=-0.097, *p*=.923). The effect of Prime Type was not modulated by Grade ($\chi^2(2)$=2.93, *p*=.403), indicating that there was no significant developmental change over grades. Figure 2 shows the z-transformed response times as a function of Grade.

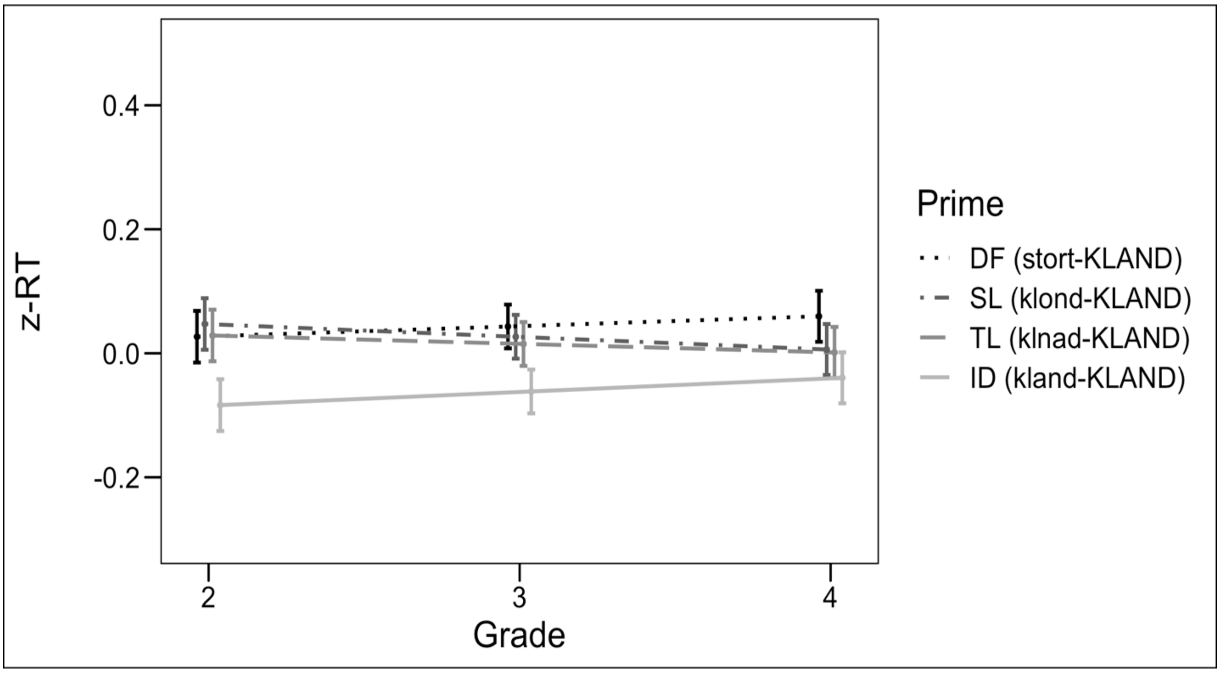

Figure 2. Mean predicted response times (z-transformed) in each prime condition for nonword targets as a function of grade.

Interindividual differences. To test the influence of interindividual differences in orthographic knowledge on the developmental trajectory of priming effects, we added the composite Orthographic Knowledge score to the previous models as a fixed effect and in interaction with Prime Type and Grade. For the word data, adding Orthographic Knowledge did not change the pattern of results with regard to the effects of Grade, Prime Type, and their interaction; the main effect of Grade was not significant ($\chi^2(1)=0.08$, $p=.775$) and the previously significant effects remained significant (Prime Type: $\chi^2(1)=131.23$, $p<.001$; Grade x Prime Type: $\chi^2(1)=32.81$, $p<.001$). Importantly, although the main effect of Orthographic Knowledge was not significant ($\chi^2(1)=0.10$, $b=-.003$, $t=-0.315$, $p=.753$), the interaction of Prime Type and Orthographic Knowledge was significant ($\chi^2(1)=18.47$, $p<.001$), indicating that the priming effects were modulated by the orthographic knowledge of the children. Breaking down this interaction showed that the difference between DF and ID was significant in both children with low (-1SD) and high (+1SD) orthographic knowledge, but was stronger for the latter (low: $b=-.130$, $t=-3.61$, $p<.001$; high: $b=-.339$, $t=-9.98$, $p<.001$). The SL-DF difference was only present in children with high, but not with low orthographic knowledge (low: $b=-.016$, $t=-0.44$, $p=.662$; high: $b=-.144$, $t=-4.21$, $p<.001$). The TL-DF difference was significant in all children, but was also stronger for children with higher orthographic knowledge (low: $b=-.148$, $t=-4.12$, $p<.001$; high: $b=-.300$, $t=-8.86$, $p<.001$). Turning to ID as a baseline, the SL-ID difference was significant in all children, but was again stronger for children with higher orthographic knowledge (low: $b=.114$, $t=3.19$, $p=.001$; high: $b=.196$, $t=5.81$, $p<.001$). The TL-ID difference was not significant either in low or in high orthographic knowledge children (low: $b=-.018$, $t=-0.51$, $p=.607$; high: $b=.039$, $t=1.16$, $p=.247$). Figure 3 shows the z-transformed response times as a function of Orthographic Knowledge. There was no significant three-way interaction of Prime Type, Grade and Orthographic Knowledge ($\chi^2(3)=3.37$, $p=.338$).

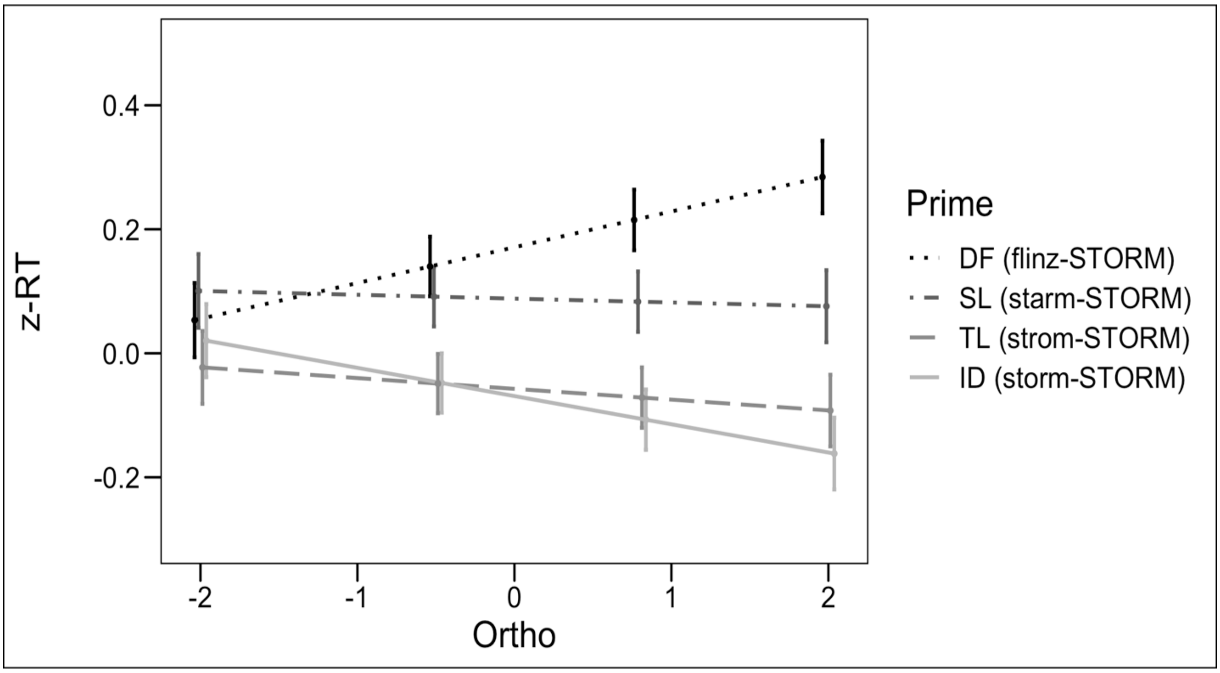

Figure 3. Mean predicted response times (z-transformed) in each prime condition for word targets as a function of orthographic knowledge.

Also for the nonword data, adding Orthographic Knowledge did not change the pattern of results with regard to the effects of Grade, Prime Type, and their interaction; Prime Type remained significant ($\chi^2(1)=20.68$, $p<.001$), and the interaction of Prime Type and Grade remained non-significant ($\chi^2(1)=3.10$, $p=.377$). There was no main effect of Orthographic Knowledge ($\chi^2(1)=0.01$, $p=.935$), nor was any interaction involving Orthographic Knowledge significant (Prime Type x Ortho: $\chi^2(3)=0.73$, $p=.867$; Grade x Ortho: $\chi^2(1)=3.37$, $p=.067$; Prime Type x Grade x Ortho: $\chi^2(3)=5.39$, $p=.146$). Figure 4 shows the z-transformed response times as a function of Orthographic Knowledge.

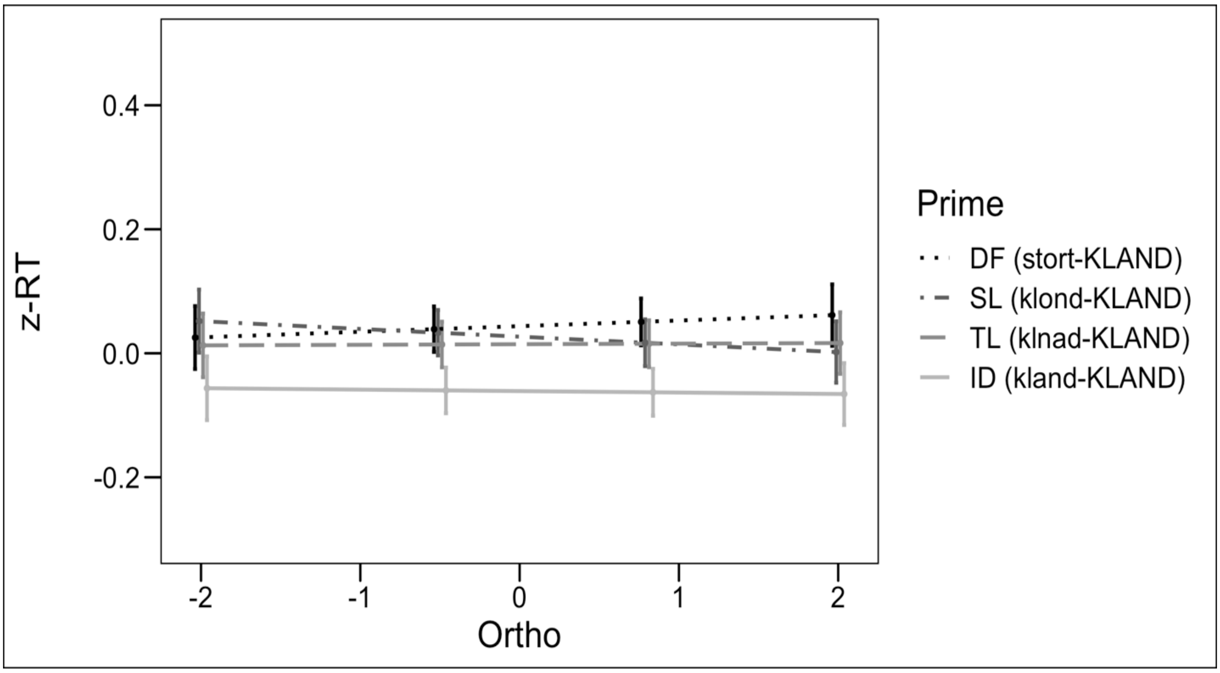

Figure 4. Mean predicted response times (z-transformed) in each prime condition for nonword targets as a function of orthographic knowledge.

Discussion

In the present experiment, we addressed the question of how the reading system processes letter position and identity during reading development and what drives these potential changes. To this end, we conducted a longitudinal examination of letter position and identity coding by tracking masked priming effects in a large group of children over three years. This investigation went beyond previous studies not only because it was longitudinal, thus avoiding some of the problems of cross-sectional studies, but also because it examined three important aspects that have been neglected so far.

One concerns the baseline to which effects are compared; an issue recently raised by Kezilas et al. (2017). Many previous studies on letter coding have compared TL primes to SL primes and/or to DF primes. Kezilas et al. (2017) recently argued that the more important baseline is one with complete overlap with the target, not only because it is more natural, but also because it allows researchers to disentangle effects of position and identity. We followed the reasoning of Kezilas et al. (2017) and used an ID baseline.

Another important aspect of our study concerns priming effects for nonword targets, which have been investigated by very few developmental studies so far (for exceptions see Acha & Perea, 2008; Ziegler et al., 2014; who, however, only compared TL and SL directly). However, nonword targets can provide additional insights about the locus and nature of the priming effects (cf. Mousikou et al., 2015). Also for nonword targets, we used both comparison baselines (DF and ID).

A third aspect concerns growth of the orthographic lexicon, which has been suggested to be the driving force for developmental changes in letter coding, especially in the framework of the lexical tuning hypothesis (Castles et al., 2007). We thus tested how priming effects are modulated by interindividual differences in orthographic knowledge over and above grade.

In the following, we will discuss our results in relation to those three aspects. We start with the nonwords, as we found the least complex pattern of priming there. From there, we move on to the words, which showed a more elaborate pattern that needs to be discussed in depth.

In the responses to nonword targets, we found that children showed no significant differences between SL or TL primes in comparison to the completely unrelated DF baseline: responses to nonword targets following these primes took equally long. Although the lack of TL priming for nonwords contrasts with adult studies that do report such an effect (e.g., Kinoshita & Norris, 2009; Mousikou et al., 2015), it is in line with the lexical tuning framework (Castles et al., 2007), which assumes a primarily lexical origin of such effects and would hence not precit them for nonwords. The importance of precise letter position and identity coding in nonword reading also resonates with the ordered input coding assumed in the phonological (nonlexical) route of the multiple route model (Grainger & Ziegler, 2011).

Interestingly, children did show significantly faster responses to nonwords after ID primes as compared to any other prime type. Hence, nonword rejection is only facilitated if the preceding prime matches the target completely in terms of both letter position and identity. This is especially peculiar as it cannot be due to any information the prime itself gives away: the SL nonword prime *klond* is as much a nonword as the ID nonword prime *kland* (in contrast, the ID word prime *storm* is a real word and can thus already bias the system towards a “word”-response). The ID condition is somewhat special in our experiment as it is the only condition in which no changes occur from the target preview to the prime in sandwich priming. Adding up the 30 ms of target preview and the 50 ms of prime leads to an effective prime duration of 80 ms. This alone could have potentially led to increased priming effects. However, one has to keep in mind that such increased presentation times of 80 ms also apply to the majority of letters that remain unchanged in the TL and SL condition. Hence, the differences between ID and TL/SL must be driven by the inconsistent information that is presented in the 50 ms prime time between target preview and actual target. The processing advantage must then stem from the fact that no changes in letter position or identity occur in the ID condition.

Concerning developmental changes in nonword priming, we do not find a significant modulation of the priming pattern by grade, despite Figure 2 suggesting that the differences with ID become slightly smaller. We cannot completely rule out that developmental changes are a) too subtle and overall response times too long (cf. Table 2) and/or b) occur only later in development and were therefore not captured in our investigation. To address the issue whether the absence of a significant interaction is due to insensitivity of data or really indicates the absence of developmental change, we additionally conducted a Bayes Factor analysis using the BayesFactor package in R (Morey & Rouder, 2018). We compared a model including the interaction with grade to a model without this interaction and obtained a BF close to zero ($BF10 = 0.041 \pm 0.02\%$), which can be counted as “very strong” evidence against the interaction (Lee & Wagenmakers, 2013). Hence, we can assume that there is no indication of developmental changes in the priming effects for nonwords in our study. Similarly, there is no indication of orthographic knowledge modulating the pattern of priming effects for nonword targets.

Based on our nonword data, we cannot distinguish between the multiple-route model and the lexical tuning hypothesis. Importantly, though, we can conclude that the advantage for ID primes for nonword targets in developing readers remains unchanged during the elementary school years (cf. Gomez & Perea, 2020), showing no developmental changes.

Turning to the word data, the results averaged across all grades indicate that targets preceded by TL or SL primes were faster compared to DF control primes, supporting a certain general degree of preactivation of the word even when letter identity or letter position did not match exactly (i.e., when overlap between prime and target was only partial). Moreover, there was no significant difference between TL primes and ID control primes – they were indistinguishable. At the same time, SL primes were significantly slower than ID control primes. This is particularly crucial in disentangling letter position and identity priming: 1SL and ID primes have more position-locked overlap (namely, 4/5 letters: **st**a**rm-st**o**rm**) than TL and ID primes do (3/5 letter: **str**o**m-st**or**m**), while the position-independent letter overlap is higher for TL (5/5 letters) than SL primes (4/5 letters). Hence, this is a strong indicator that letter position is more flexible early in German reading development, while the system seems to code letter identity much more precisely.

In contrast to the pattern of nonword priming, which remained constant, the pattern of word priming was modulated by grade, indicating that it underwent changes throughout development. Closer inspection showed that both SL and TL effects compared to a DF baseline *increased* over the school years, which is usually taken as indication for letter coding becoming more *flexible*. However, SL and TL effects compared to an ID baseline also *increased* over the school years, which is usually taken as indication for letter coding becoming more *precise*[1].

When compared to previous results, we can observe some similarities in particular developmental trends. Most strikingly, the TL-ID increase, which becomes significant in the middle of elementary school, is in line with Kezilas et al. (2017). The TL-DF increase is in line with observations by Lété and Fayol (2013), whereas Kezilas et al. (2017) report no developmental changes of this contrast. It is possible that our data, being longitudinal, were more sensitive to such a change than Kezila's cross-sectional data. Another contrast to previous results concerns the SL-DF increase, where Castles et al. (2007) reported a decrease. Because they also

[1] Note that we found a significant TL-ID effect in the unexpected direction (TL faster than ID) in grade 2. This should most likely be interpreted as noise, also connected to the fact that 2nd-graders made many errors and thus many data points had to be excluded from the response time analysis. This might compromise the reliability of the data in this grade and it should hence be interpreted with caution. To make sure, however, that the observed developmental trends were not due to noisy grade 2 data, we reran the respective analyses without the grade 2 data. The pattern of effect, and most importantly the interactions involving grade, remained significant at $p<0.05$. We are thus confident to report a developmental trend despite potentially increased noise in the grade 2 data.

used a 1SL (not 2SL) prime, the type of SL prime, and therefore amount of letter overlap with the target, can be ruled out as a confound. The SL-ID development has never been tested before. Although Kezilas et al. (2017) included both prime types, they did not report this exact contrast. Inspection of their figures suggests an increase throughout development, which would be in line with our pattern. However, without statistical analyses to compare, this is just speculative. Other studies summarized in Table 1 (Acha & Perea, 2008; Ziegler et al., 2014; Colombo et al., 2019) did not include DF or ID as comparison baselines and can thus not be used for direct comparison.

Concerning orthographic knowledge, we found that it modulates the priming of word targets. Importantly, the way it modulated word priming was very similar to the modulation we observed by grade: for different levels of orthographic knowledge *within* grades, we see a similar pattern as the one *across* grades (cf. Figure 1 and 3). This implies that the developmental changes in letter position and identity coding are indeed driven by growing orthographic knowledge, for which number of school years are a good proxy, and not only by general maturation. Both SL-DF and TL-DF effects increased with orthographic knowledge. Moreover, SL-ID differences increased with orthographic knowledge. Notably, TL-ID differences were not significant for any level of orthographic knowledge. Even though Figure 3 suggests that they increased, they never reached significance. Also, the modulation of the TL-ID difference by orthographic knowledge was not statistically significant, although Figure 3 suggests a trend towards a growing effect. Again, it is not clear whether the modulation was just too subtle to be detected, so we conducted a post-hoc Bayes Factor analysis (Morey & Rouder, 2018) comparing the TL-ID contrast in a model with and without the interaction with orthographic knowledge. We obtained a BF close to zero ($BF10 = 0.075 \pm 0\%$), indicating "strong" evidence for a model without the interaction (Lee & Wagenmaker, 2013). This is problematic because this specific modulation is one core prediction of the lexical tuning hypothesis. An item-specific measure of orthographic knowledge might be necessary to detect a reliable modulation. This would be an interesting test case for future studies.

With regard to the models of reading development, the trend of the effects is indecisive. At first glance, both models seem to fit partially, but neither completely. The increasing TL-DF difference is in line with the multiple-route model, while the increasing TL-ID and SL-ID differences are consistent with the lexical tuning hypothesis. Although Kezilas et al. (2017) argue that the TL-ID comparison is the most informative one to adjudicate between the models, we similarly want to highlight our finding regarding the TL-DF contrast. Hence, despite our initial motivation for the study, we find ourselves in a situation where our experimental results do not allow us to endorse or reject either of the models completely. The models fail to capture the current data because they do not account for the diverging developmental pattern depending on which baseline, ID or DF, is used. Instead, they implicitly assume that processing of ID and DF primes remains stable across development and they disregard that this can be subject to change too.

In fact, the most striking developmental finding, which we did not set out to test, but which we made thanks to the use of different comparison baselines, is that the most marked developmental changes seem to occur in the ID and DF priming conditions. This becomes very clear from Figure 1 (see also Figure 3). Such a pattern can also be observed in the data of Kezilas et al. (2019), although they do not comment on it. This finding indicates that not only the processing of TL and SL primes changes, but also – or even more so – the processing of completely overlapping (ID) or completely non-overlapping (DF) primes. This is even more interesting when we compare it to the pattern of nonword priming, where we found a special advantage for ID primes as well, but no changes across development (neither as captured by grade nor by orthographic knowledge). Hence, the changes we see for ID primes of words must be due to a mechanism that is specific to words and lies beyond the encoding advantage of ID primes that we observed for nonwords. The increase in ID priming over development thus likely involves changes at the lexical level, potentially related to developmental changes in the strength of the underlying lexical representations or the letter-to-word connections (cf. McClelland & Rumelhart, 1981).

For the DF prime, we can observe the opposite trend as for the ID prime: it seems to become more inhibitory over development. Again, this could be related to development of the strength of lexical representations that lead to stronger inhibition effects from lexical competitors. Thus, the same developmental mechanism could explain both the increase in facilitation from completely overlapping primes as well as the emerging inhibition effects from inconsistent primes.

A similar divergent pattern between identity and all-different primes was observed by Jacobs, Grainger, and Ferrand (1995) in an incremental masked priming study with adult readers using primes of varying intensity. Incremental priming can provide additional information about how a prime affects target processing as it becomes increasingly available. If response times increase with increasing prime availability, then it is likely that inhibitory processes are at work (see also Ziegler, Ferrand, Jacobs, Rey, & Grainger, 2000). For higher levels of prime intensity, Jacobs et al. (1995) found stronger facilitation by identity primes and stronger inhibition by all-different primes. The more lexical information from identical primes was activated, the more it facilitated the subsequent identification of the target. At the same time, the more information from incompatible primes was activated, the more target processing was inhibited, supposedly via inhibitory connections at the lexical level.

Previous studies on transpositions and substitutions have mainly focused on how preactivation from consistent primes affected target processing. However, there is accumulating evidence that inhibition from inconsistent information is equally important and should be given more attention in models of orthographic processing (e.g,, Lupker, Spinelli, & Davis, 2020). Clearly, inhibition dynamics might also be especially important to explain changes in priming effects throughout development (Kezilas et al., 2017). The idea that maturing lexical representations change the sensitivity of the orthographic system and drive priming effects in the ID and DF condition into opposite directions is generally consistent with the lexical tuning hypothesis and also compatible with the multiple route model. The question is then, how do facilitation and inhibition from consistent or inconsistent information change because of lexicon growth or an increase in the strength of lexical representations? One promising approach would be to investigate the correspondence between the emerging priming effects during reading development (i.e., when the lexicon grows) and the increase of the same effects when using an incremental priming technique in adults (i.e., when the lexicon is stable; Jacobs et al., 1996; Ziegler et al., 2000). This would allow researchers to disentangle the effects of bottom-up information from prime availability and top-down activation from the orthographic lexicon.

Another important issue to be discussed in the context of our study pertains to cross-linguistic differences. All studies covered in Table 1 investigated Indo-European languages, but with different degrees of orthographic transparency. Colombo et al. (2019; see also Lété & Fayol, 2013) discuss how the ease and success of phonological decoding versus lexical recognition might impact the developmental timeline of letter position coding. Another potentially even more powerful factor could be the density of an orthography, that is, how many anagrams a language has (Frost, 2012; Lally, Taylor, Lee, & Rastle, 2020; Lerner, Armstrong, & Frost, 2014). Flexibility arises, or is constrained, depending on what best serves word recognition in a specific language. Hence, if a language features many transposition neighbors, the pressure to code letter position more precisely is higher than when there are few and the same might apply to the number of substitution neighbors. It would therefore be extremely useful to derive those statistics for the languages examined so far and relate them to the observed developmental trajectories.

Conclusion

Our study brings new insights into the debated issue of how letter position and identity processing develop. For nonwords, we found few priming effects, and these remained stable across levels of reading development or orthographic knowledge. For words, we found substantial developmental changes in masked priming effects. Letter identity was coded relatively strictly even in beginning readers. With regard to the processing of letter position, we found increasing priming effects compared to the DF baseline, which indicates increasing flexibility across grades. However, priming effects also increase in comparison to the ID baseline condition, which indicates that the processing system becomes more strict. Both effects were modulated by orthographic knowledge, suggesting – in accordance with the absence of effects for nonwords – that they are primarily lexical.

The developmental pattern fits both the lexical tuning hypothesis and the multiple route hypothesis partly in ways we predicted. However, it fits neither of them fully. This demonstrates that the models, as well as the experimental studies so far, potentially overlooked changes in the processing of all-different and identity primes. Our results suggest that processing of inconsistent and consistent information that leads to diverging inhibition vs. facilitation during lexical processing. These opposing effects might be connected to structural changes of the orthographic lexicon. The processing system gets better at using overlapping information between prime and target, but at the same time becomes more sensitive to subtle differences between orthographic strings. The interplay between the two seemingly opposing principles of flexibility and precision is ubiquitous in the reading system: It must reliably detect invariances in the orthographic input (e.g., the same letter written in different sizes or fonts) while simultaneously discerning even small differences (e.g., activate different letter representations for "e" and "c"; Dehaene et al., 2005). During reading development, the system becomes specialized to orthographic input strings while treating them also more flexibly on various hierarchical levels (Dehaene et al., 2015; Szwed et al., 2011). It can be assumed that similar mechanisms are relevant for the coding of letter identity and position and any comprehensive theory of orthographic processing must account for them.

References

Acha, J., & Perea, M. (2008). The effect of neighborhood frequency in reading: Evidence with transposed-letter neighbors. *Cognition, 108(1),* 290–300. doi:10.1016/j.cognition.2008.02.006

Adelman, J. S., Johnson, R. L., McCormick, S. F., McKague, M., Kinoshita, S., Bowers, J. S., … Davis, C. J. (2014). A behavioral database for masked form priming. *Behavior Research Methods, 46(4)*, 1052–1067. doi:10.3758/s13428-013-0442-y

Andrews, S., & Hersch, J. (2010). Lexical precision in skilled readers: Individual differences in masked neighbor priming. *Journal of Experimental Psychology: General*, 139(2), 299–318. doi:10.1037/a0018366

Andrews, S., & Lo, S. (2012). Not all skilled readers have cracked the code: Individual differences in masked form priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 38(1), 152–163. doi:10.1037/a0024953

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Usinglme4. *Journal of Statistical Software, 67(1)*. doi:10.18637/jss.v067.i01

Beyersmann, E., Grainger, J., Casalis, S., & Ziegler, J. C. (2015). Effects of reading proficiency on embedded stem priming in primary school children. *Journal of Experimental Child Psychology, 139,* 115–126. doi:10.1016/j.jecp.2015.06.001

Castles, A., Davis, C., Cavalot, P., & Forster, K. (2007). Tracking the acquisition of orthographic skills in developing readers: Masked priming effects. *Journal of Experimental Child Psychology, 97(3)*, 165–182. doi:10.1016/j.jecp.2007.01.006

Colombo, L., Sulpizio, S., & Peressotti, F. (2019). The developmental trend of transposed letters effects in masked priming. *Journal of Experimental Child Psychology, 186,* 117–130. doi:10.1016/j.jecp.2019.05.007

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. doi:10.1037/0033-295x.108.1.204

Dehaene, S., Cohen, L., Morais, J., & Kolinsky, R. (2015). Illiterate to literate: behavioural and cerebral changes induced by reading acquisition. *Nature Reviews Neuroscience, 16(4),* 234-244.

Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Filho, G. N., Jobert, A., … Cohen, L. (2010). How Learning to Read Changes the Cortical Networks for Vision and Language. *Science, 330(6009),* 1359–1364. doi:10.1126/science.1194140

Faust, M. E., Balota, D. A., Spieler, D. H., & Ferraro, F. R. (1999). Individual differences in information-processing rate and amount: Implications for group differences in response latency. *Psychological Bulletin, 125(6),* 777–799. doi:10.1037/0033-2909.125.6.777

Forster, K. I., Davis, C., Schoknecht, C., & Carter, R. (1987). Masked priming with graphemically related forms: Repetition or partial activation? *The Quarterly Journal of Experimental Psychology Section A, 39(2)*, 211–251. doi:10.1080/14640748708401785

Frost, R. (2012). Towards a universal model of reading. *Behavioral and Brain Sciences,* 35(5), 263–279. doi:10.1017/s0140525x11001841

Gomez, P., & Perea, M. (2020). Masked identity priming reflects an encoding advantage in developing readers. *Journal of Experimental Child Psychology, 199*, 104911. doi:10.1016/j.jecp.2020.104911

Gomez, P., Marcet, A., & Perea, M. (2021). Are better young readers more likely to confuse their mother with their mohter? *Quarterly Journal of Experimental Psychology, Online First Publication.* doi:10.1177/17470218211012960

Gomez, P., Ratcliff, R., & Perea, M. (2008). The overlap model: A model of letter position coding. *Psychological Review, 115(3)*, 577–600. doi:10.1037/a0012667

Grainger, J., & van Heuven, W. J. B. (2003). *Modeling letter position coding in printed word perception.* In P. Bonin (Ed.), Mental lexicon: „Some words to talk about words“ (pp.1-23). New York: Nova Science.

Grainger, J., & Ziegler, J. C. (2011). A Dual-Route Approach to Orthographic Processing. *Frontiers in Psychology, 2*. doi:10.3389/fpsyg.2011.0005

Hasenäcker, J., Beyersmann, E., & Schroeder, S. (2016). Masked Morphological Priming in German-Speaking Adults and Children: Evidence from Response Time Distributions. *Frontiers in Psychology, 7*. doi:10.3389/fpsyg.2016.00929

Hasenäcker, J., Beyersmann, E., & Schroeder, S. (2020). Morphological priming in children: Disentangling the effects of school-grade and reading skill. *Scientific Studies of Reading, 24(6),* 484-499. doi:10.1080/10888438.2020.1729768

Hasenäcker, J., Schröter, P., & Schroeder, S. (2017). Investigating developmental trajectories of morphemes as reading units in German. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43(7)*, 1093–1108. doi:10.1037/xlm0000353

Jacobs, A.M., Grainger, J., & Ferrand, L. (1995). The incremental priming technique: A method for determining within-condition priming effects. *Perception & Psychophysics, 57,* 1101-1110.

Kezilas, Y., McKague, M., Kohnen, S., Badcock, N. A., & Castles, A. (2017). Disentangling the developmental trajectories of letter position and letter identity coding using masked priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43(2)*, 250–258. doi:10.1037/xlm0000293

Kinoshita, S., & Norris, D. (2009). Transposed-letter priming of prelexical orthographic representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 35(1), 1–18. doi:10.1037/a0014277

Kinoshita, S., Castles, A., & Davis, C. (2009). The role of neighbourhood density in transposed-letter priming. *Language and Cognitive Processes, 24(4)*, 506–526. doi:10.1080/01690960802289213

Lee, M., & Wagenmakers, E. (2014). Bayesian Cognitive Modeling: A Practical Course. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139087759

Lerner, I., Armstrong, B. C., & Frost, R. (2014). What can we learn from learning models about sensitivity to letter-order in visual word recognition? *Journal of Memory and Language, 77,* 40–58. doi:10.1016/j.jml.2014.09.002

Lété, B., & Fayol, M. (2013). Substituted-letter and transposed-letter effects in a masked priming paradigm with French developing readers and dyslexics. *Journal of Experimental Child Psychology, 114(1)*, 47–62. doi:10.1016/j.jecp.2012.09.001

Lupker, S. J., & Davis, C. J. (2009). Sandwich priming: A method for overcoming the limitations of masked priming by reducing lexical competitor effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35(3)*, 618–639. doi:10.1037/a0015278

Lupker, S. J., Perea, M., & Davis, C. J. (2008). Transposed-letter effects: Consonants, vowels and letter frequency. *Language and Cognitive Processes, 23(1)*, 93–116. doi:10.1080/01690960701579714

Lupker, S. J., Spinelli, G., & Davis, C. J. (2020). Is zjudge a better prime for JUDGE than zudge is? A new evaluation of current orthographic coding models. *Journal of Experimental Psychology: Human Perception and Performance, 46,* 1252-1266. doi:10.1037/xhp0000856

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review, 88*, 375-407.

Morey, R. D. & Rouder, J. N. (2018). BayesFactor: Computation of Bayes Factors for Common Designs. R package version 0.9.12- 4.2. https://CRAN.R-project.org/package=BayesFactor.

Mousikou, P., Kinoshita, S., Wu, S., & Norris, D. (2015). Transposed-letter priming effects in reading aloud words and nonwords. *Psychonomic Bulletin & Review, 22(5)*, 1437–1442. doi:10.3758/s13423-015-0806-7

Norris, D., & Kinoshita, S. (2008). Perception as evidence accumulation and Bayesian inference: Insights from masked priming. *Journal of Experimental Psychology: General,* 137(3), 434–455. doi:10.1037/a0012799

Norris, D., & Kinoshita, S. (2012). Reading through a noisy channel: Why there's nothing special about the perception of orthography. *Psychological Review,* 119, 517–545. doi:10.1037/a0028450

Norris, D., Kinoshita, S., & van Casteren, M. (2010). A stimulus sampling theory of letter identity and order. *Journal of Memory and Language,* 62(3), 254–271. doi:10.1016/j.jml.2009.11.002

Pegado, F., Comerlato, E., Ventura, F., Jobert, A., Nakamura, K., Buiatti, M., … Dehaene, S. (2014). Timing the impact of literacy on visual processing. *Proceedings of the National Academy of Sciences, 111(49),* E5233–E5242. doi:10.1073/pnas.1417347111

Perea, M., & Lupker, S. J. (2003). Does jugde activate COURT? Transposed-letter similarity effects in masked associative priming. *Memory & Cognition, 31(6)*, 829–841. doi:10.3758/bf03196438

Perea, M., & Lupker, S. J. (2004). Can CANISO activate CASINO? Transposed-letter similarity effects with nonadjacent letter positions. *Journal of Memory and Language, 51(2)*, 231–246. doi:10.1016/j.jml.2004.05.005

Perea, M., Jiménez, M., & Gomez, P. (2016). Does location uncertainty in letter position coding emerge because of literacy training? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42(6),* 996–1001. doi:10.1037/xlm0000208

Peressotti, F., & Grainger, J. (1999). The role of letter identity and letter position in orthographic priming. *Perception & Psychophysics, 61(4)*, 691–706. doi:10.3758/bf03205539

Schroeder, S., Würzner, K.-M., Heister, J., Geyken, A., & Kliegl, R. (2014). childLex: a lexical database of German read by children. *Behavior Research Methods, 47(4)*, 1085–1094. doi:10.3758/s13428-014-0528-1

Szwed, M., Dehaene, S., Kleinschmidt, A., Eger, E., Valabrègue, R., Amadon, A., & Cohen, L. (2011). Specialization for written words over objects in the visual cortex. *Neuroimage, 56(1),* 330-344.

Tiffin-Richards, S. P., & Schroeder, S. (2015). Children's and adults' parafoveal processes in German: Phonological and orthographic effects. *Journal of Cognitive Psychology, 27 (5)*, 531-548. doi:10.1080/20445911.2014.999076

Welcome, S. E., & Trammel, E. R. (2017). Individual differences in orthographic priming relate to phonological decoding skill in adults. *Cognitive Processing,* 18(2), 119–128. doi:10.1007/s10339-017-0793-x

Whitney, C. (2001). How the brain encodes the order of letters in a printed word: The SERIOL model and selective literature review. *Psychonomic Bulletin & Review, 8(2)*, 221–243. doi:10.3758/bf03196158

Whitney, C., & Cornelissen, P. (2008). SERIOL Reading. *Language and Cognitive Processes, 23(1)*, 143–164. doi:10.1080/01690960701579771

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. doi:10.1037/a0035187

Ziegler, J. C., Ferrand, L., Jacobs, A. M., Rey, A., & Grainger, J. (2000). Visual and phonological codes in letter and word recognition: Evidence from incremental priming. *The Quarterly Journal of Experimental Psychology, 53,* 671-692. doi:10.1080/713755906

Zoccolotti, P., De Luca, M., Di Filippo, G., Judica, A., & Martelli, M. (2008). Reading development in an orthographically regular language: effects of length, frequency, lexicality and global processing ability. *Reading and Writing, 22(9),* 1053–1079. doi:10.1007/s11145-008-9144-8

Appendix

Table A1. *Complete list of word and nonword targets and primes used in Experiment 1 and 2.*

Target	SL Prime	TL Prime	DF Prime	Target Lexicality
BLECH	blich	belch	krint	word
BLOCK	bluck	bolck	frund	word
BRAND	bsand	barnd	zwomk	word
BRUCH	brnch	burch	zwoln	word
DRUCK	drock	durck	psolt	word
DURST	dumst	drust	balch	word
FLECK	flrck	felck	trasb	word
FLUCH	flich	fulch	ksint	word
FLUSS	flass	fulss	tramc	word
GLÜCK	glsck	gülck	pfarz	word
GRAMM	grumm	garmm	jsunz	word
GRIFF	groff	girff	pnolm	word
GRILL	grbll	girll	psunf	word
KAMPF	kaspf	kmapf	husgt	word
KELCH	kench	klech	bodsz	word
KNALL	knull	kanll	frumz	word
KRACH	krech	karch	spemk	word
MILCH	mikch	mlich	sarft	word
PLATZ	plstz	paltz	grufb	word
TRICK	trock	tirck	kwond	word
TRITT	trbtt	tirtt	spalf	word
TROLL	trall	torll	knads	word
WURST	wumst	wrust	minzd	word
ZWECK	zwock	zewck	srond	word
BRUST	brest	brsut	knerl	word
DRAHT	driht	drhat	blifk	word
FROST	frast	frsot	gnark	word
FRUST	frzst	frsut	knerb	word
GLANZ	glonz	glnaz	pform	word
GRUND	grend	grnud	zwemf	word
KLANG	klung	klnag	stump	word
KRAFT	krsft	krfat	spold	word
KRANZ	krinz	krnaz	dsirm	word
PFAND	pfond	pfnad	glort	word

PFERD	pfird	pfred	glint	word
PRINZ	pronz	prniz	gloms	word
SPORT	spgrt	sprot	dremf	word
STAND	stlnd	stnad	pfirl	word
START	sturt	strat	gruln	word
STERN	storn	stren	rhomz	word
STIFT	stuft	stfit	kruhd	word
STIRN	stzrn	strin	pfamz	word
STROH	stnoh	storh	pfned	word
STUHL	strhl	sthul	brodf	word
STURM	stkrm	strum	flonz	word
STURZ	stirz	struz	plims	word
TROST	trast	trsot	knamb	word
ZWERG	zworg	zwreg	snomp	word
GLACK	glock	galck	pfort	nonword
BROLF	bralf	borlf	knahd	nonword
KRINN	krsnn	kirnn	stulm	nonword
GRUCK	grick	gurck	blims	nonword
TALST	tafst	tlast	lunkd	nonword
PLOCK	plnck	polck	frunz	nonword
FLIRN	flurn	filrn	stumz	nonword
BRINN	bronn	birnn	fnorz	nonword
PLUCK	plzck	pulck	dranf	nonword
DRIMM	dromm	dirmm	blorz	nonword
TRELL	trall	terll	zwamk	nonword
BROMM	brzmm	bormm	knist	nonword
KALMT	kabmt	klamt	torks	nonword
KILPF	kidpf	klipf	fumch	nonword
DROLL	drell	dorll	stenz	nonword
BLACH	blnch	balch	pfind	nonword
MIRFT	minft	mrift	ponck	nonword
PLARN	plorn	palrn	stomz	nonword
TRETZ	tratz	tertz	fnals	nonword
BRULL	brsll	burll	pladt	nonword
FRINN	fronn	firnn	gloms	nonword
WINST	wimst	wnist	zurml	nonword
ZWUNN	zwlnn	zuwnn	srimr	nonword
KLAND	klond	klnad	stort	nonword
DRAST	drist	drsat	knimf	nonword
PRONT	prant	prnot	zwalf	nonword

FLOST	flist	flsot	prich	nonword
GROLZ	grmlz	grloz	smulf	nonword
FLINZ	flenz	flniz	grems	nonword
BRIND	brund	brnid	glumk	nonword
KLIRG	klerg	klrig	bnedt	nonword
BRAFT	brnft	brfat	smolk	nonword
GRONF	granf	grnof	plist	nonword
STURN	stgrn	strun	zwosd	nonword
KNERD	knird	knred	pfizt	nonword
PRILF	prelf	prlif	gnekt	nonword
SPOLT	spilt	splot	brifd	nonword
STAST	stust	stsat	prulf	nonword
STORL	sterl	strol	bneck	nonword
STELD	strld	stled	frabt	nonword
STENT	stont	stnet	mlorf	nonword
STONF	stenf	stnof	brezd	nonword
STRUN	stmun	sturn	pfliz	nonword
STAHM	stohm	stham	zlokn	nonword
SPIRM	spzrm	sprim	rhans	nonword
DRAPS	drups	drpas	knugz	nonword
PRAST	prist	prsat	gnimf	nonword
ZWELF	zwalf	zwlef	mrask	nonword

Analysis using log-transformed response times

We originally also ran all analyses on log-transformed response times as the dependent variable. For completeness, we present those analyses here as well.

Words. The response time analysis for word trials revealed a significant main effect of Grade ($\chi^2(1)=7234.80$, $b=-.315$, $t=-85.06$, $p<.001$). This effect indicated that responses became faster with increasing grades. More importantly, there was a main effect of Prime Type ($\chi^2(3)=129.47$, $p<.001$). Post-hoc contrasts showed that all prime types differed significantly from each other except for the ID-TL comparison; DF-ID: $t=-9.62$, $p<.001$, SL-DF: $t=-3.90$,

$p<.001$, TL-DF: t=-9.42, $p<.001$, SL-TL: t=-5.54, $p<.001$, SL-ID: t=5.74, $p<.001$, TL-ID: t=0.21, p=.835.

Crucially, the effect of Prime Type was modulated by Grade ($\chi^2(3)$=21.66, $p<.001$), indicating that the developmental trajectory differed for the different prime types. Decomposing this interaction by using post-hoc contrasts showed that the difference between DF and ID was already present in grade 2 and increased with increasing grade (grade 2: t=-2.68, p=.007; grade 3: t=-9.62, $p<.001$; grade 4: t=-9.62, $p<.001$). The SL-DF difference emerged in grade 3 (grade 2: t=-1.85, p=.065; grade 3: t=-3.90, $p<.001$; grade 4: t=-3.10, p=.002). The TL-DF difference was present in grade 2 and increased over grades (grade 2: t=-4.35, $p<.001$; grade 3: t=-9.42, $p<.001$; grade 4: t=-7.61, $p<.001$). Considering ID as a comparison baseline, the SL-ID difference emerged in grade 3 and increased towards grade 4 (grade 2: t=0.83, p=.407; grade 3: t=5.74, $p<.001$; grade 4: t=6.55, $p<.001$). The TL-ID difference was only significant in grade 4 (grade 2: t=-1.68, p=.093; grade 3: t=0.21, p=.835; grade 4: t=2.02, p=.044). Figure A1 shows the log-transformed response times as a function of Grade.

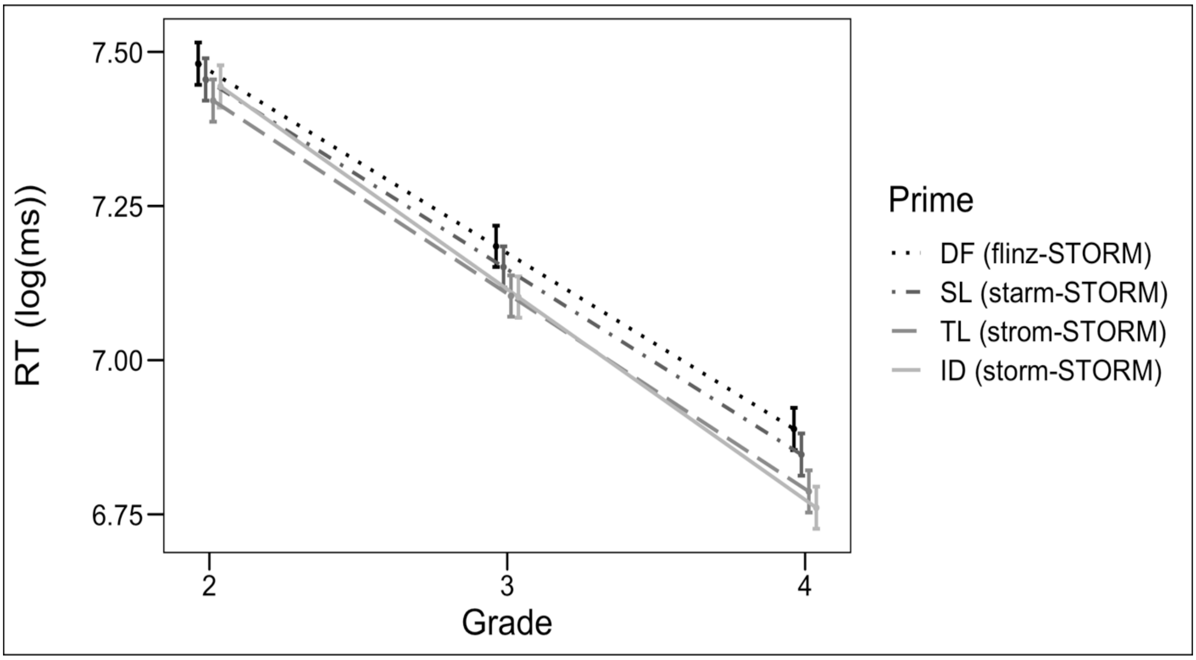

Figure A1. Mean predicted response times (in log(ms)) in each prime condition for word targets as a function of grade.

Nonwords. The response time analysis for nonword trials revealed a significant main effect of Grade ($\chi^2(1)$=9139.22, *b*=-0.360, *t*=-95.60, *p*<.001). This effect indicated that responses became faster with increasing grades. More importantly, there was a main effect of Prime Type ($\chi^2(3)$=24.08, *p*<.001). Post-hoc contrasts decomposing this effect showed that DF and ID differed significantly from each other (*t*=-4.47, *p*<.001). Neither the SL-DF nor the TL-DF difference were significant (SL-DF: *t*=-0.53, *p*=.593, TL-DF: *t*=-1.24, *p*=.214). However, both the SL-ID and the TL-ID differences were significant (SL-ID: *t*=3.94, *p*<.001, TL-ID: *t*=3.23, *p*=.001). The effect of Prime Type was not modulated by Grade ($\chi^2(2)$=3.92, *p*=.271), indicating that there was no developmental change over grades. Figure A2 shows the log-tranformed response times as a function of Grade.

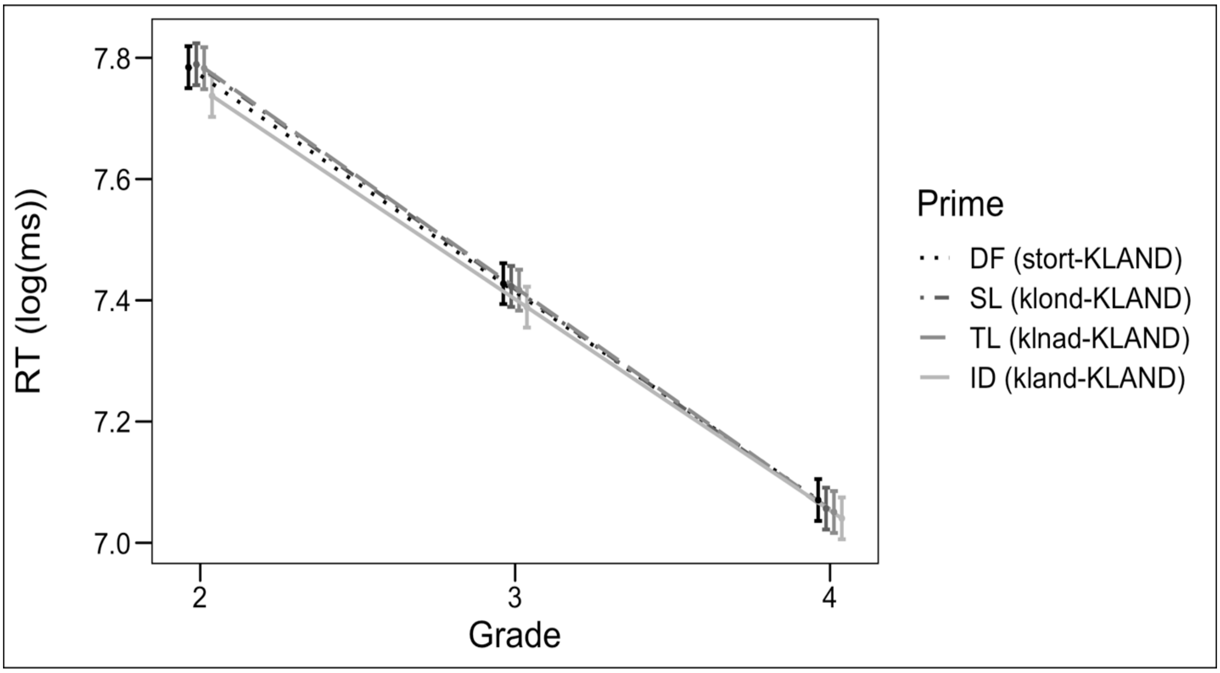

Figure A2. Mean predicted response times (in log(ms)) in each prime condition for nonword targets as a function of grade.

Interindividual differences. To test the influence of interindividual differences in orthographic knowledge on the developmental trajectory of priming effects in each grade, we added the composite Orthographic Knowledge score to the previous models as a fixed effect and in interaction with Prime Type and Grade. For the word data, adding Orthographic Knowledge did not change the pattern of results with regard to the effects of Grade, Prime Type, and their interaction; they all remained significant (Grade: $\chi^2(1)=7469.01$, $p<.001$; Prime Type: $\chi^2(1)=126.80$, $p<.001$; Grade x Prime Type: $\chi^2(1)=22.63$, $p<.001$). In addition, Orthographic Knowledge itself was a significant predictor: response times were faster for children with higher Orthographic Knowledge ($\chi^2(1)=108.74$, $p<.001$). The interaction of Grade and Orthographic Knowledge was also significant ($\chi^2(1)=158.11$, $p<.001$), indicating that the difference in response times between children with better and poorer Orthographic Knowledge was more pronounced in earlier than in later grades. Importantly, the interaction of Prime Type and Orthographic Knowledge was also significant ($\chi^2(1)=15.03$, $p=.002$), indicating that the priming effects were modulated by the orthographic knowledge of the children. Breaking down this interaction showed that the difference between DF and ID was significant in both children with low (-1SD) and high (+1SD) orthographic knowledge, but stronger for the latter (low: $t=-3.81$, $p<.001$; high: $t=-9.63$, $p<.001$). The SL-DF difference was only present in children with high, but not with low orthographic knowledge (low: $t=-0.88$, $p=.381$; high: $t=-4.50$, $p<.001$). The TL-DF difference again was significant in both children with low (-1SD) and high (+1SD) orthographic knowledge, but stronger for the latter (low: $t=-4.49$, $p<.001$; high: $t=-8.65$, $p<.001$). Considering ID as a comparison baseline, the SL-ID difference was also significant in both children with low (-1SD) and high (+1SD) orthographic knowledge, and again stronger for the latter (low: $t=-4.49$, $p<.001$; high: $t=-8.65$, $p<.001$). Finally, the TL-ID difference was neither significant in low nor in high orthographic knowledge children (low: $t=-0.69$, $p=.490$; high: $t=1.02$, $p=.309$). Figure A3 shows the log-transformed response times as a function of Orthographic Knowledge. There was no significant three-way interaction of Prime Type, Grade and Orthographic Knowledge ($\chi^2(3)=6.55$, $p=.088$).

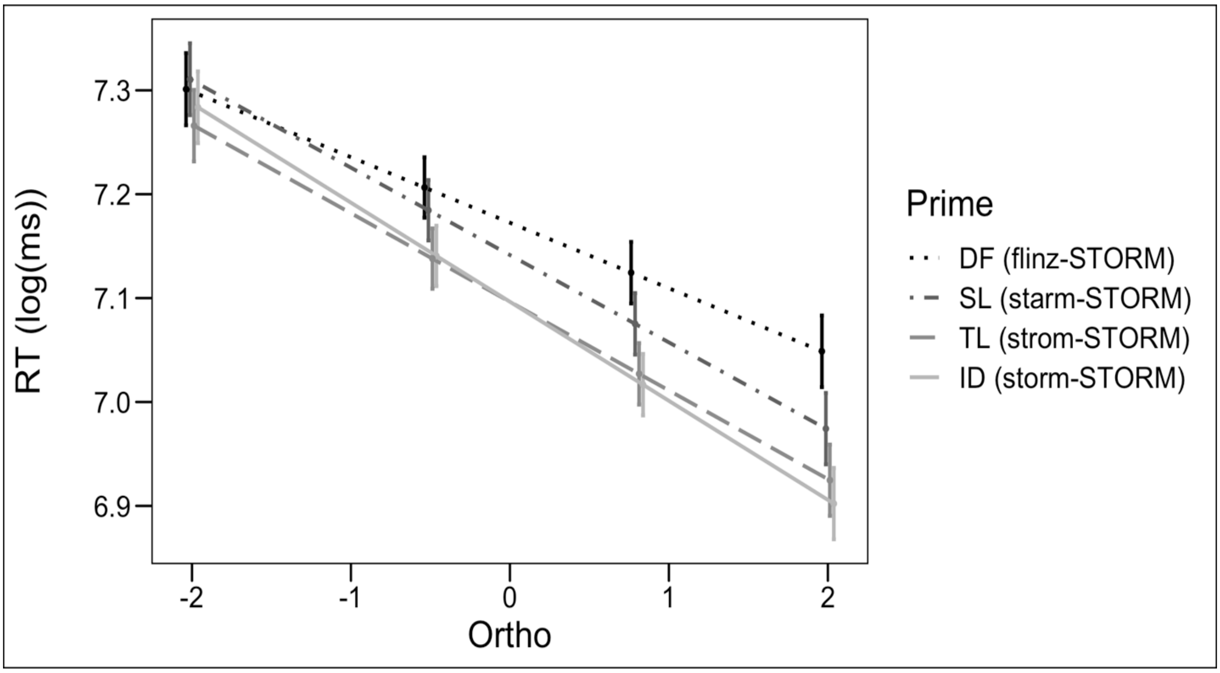

Figure A3. Mean predicted response times (in log(ms)) in each prime condition for word targets as a function of grade.

Also for the nonword data, adding Orthographic Knowledge did not change the pattern of results with regard to the effects of Grade, Prime Type, and their interaction; Grade and Prime Type remained significant (Grade: $\chi^2(1)=9217.93$, $p<.001$; Prime Type: $\chi^2(1)=24.64$, $p<.001$), and the interaction remained insignificant ($\chi^2(1)=4.19$, $p=.241$). The main effect of Orthographic Knowledge was significant ($\chi^2(1)=94.56$, $p<.001$), indicating faster response times for children with higher Orthographic Knowledge, as was the interaction of Grade and Orthographic Knowledge ($\chi^2(1)=10.53$, $p=.0.001$), indicating a more pronounced difference in response times between children with better and poorer Orthographic Knowledge in earlier than in later grades. No interaction involving Orthographic Knowledge and Prime Type was significant (Prime Type x Ortho: $\chi^2(3)=1.40$, $p=.705$; Prime Type x Grade x Ortho: $\chi^2(3)=4.56$, $p=.207$).