



Ludwig-Maximilians-Universität München

Master's Thesis submitted in partial fulfilment for the degree:
M.Sc. Psychology: Learning Sciences

Supervisor: Dr. Xenia Schmalz

Does knowledge of different orthographic regularities rely on a domain-general statistical learning mechanism? An SEM analysis of a large-scale database

Darcourt, Alvaro

Munich, August 2022

Abstract

Reading involves several cognitive processes responsible for dealing with linguistic information at the phonological, morphological, lexical, and sublexical levels. A crucial aspect of reading proficiency is fluency, which combines accuracy, automaticity, and prosody. Automaticity frees up cognitive resources that allow the reader to focus on comprehending a written text. Here, we examined how morphological awareness, lexical and sublexical knowledge, and age predict reading fluency. Additionally, we explored the potential role of statistical learning in the relationship between morphological awareness and sublexical knowledge. To answer these questions, we used structural equation modeling. Overall, we found that lexical knowledge was highly, positively, and significantly predictive of reading fluency. Age was also a relevant predictor of this literacy skill. In contrast, neither morphological awareness nor sublexical knowledge, variables related to statistical learning, were significant predictors of fluency. In addition, we found no evidence that statistical learning, implemented as a correlation between morphological awareness and sublexical knowledge, played any role in the predictive model tested. We also obtained psychometric evidence that the task employed to measure morphological awareness has a multidimensional structure and limited reliability. From a methodological point of view, the results of the present study point to the importance of evaluating the dimensionality and reliability of the measurements used. From a theoretical point of view, we conclude that lexical knowledge constitutes a crucial pillar of fluency.

Keywords: morphological awareness, orthographic knowledge, reading fluency, statistical learning

Character count: 90,432

Word count: 13,514

Does knowledge of different orthographic regularities rely on a domain-general statistical learning mechanism? An SEM analysis of a large-scale database

Reading involves several cognitive processes of varying complexity (Zarić et al., 2021). Word-reading operations, such as decoding individual words and vocabulary knowledge, occur at the most basic level. The most complex level involves understanding relationships between sentences and processing texts. Moving from the former to the latter level will depend on the reader successfully integrating and interpreting the meaning of multiple words in specific contexts (Zarić et al., 2021).

A key component in reading proficiency development is fluency, which reflects how fast the reader translates text into speech (Hudson et al., 2005; Kuhn et al., 2010). Reading fluency combines accuracy, automaticity, and prosody: it enables fast and effortless word recognition with appropriate pace, phrasing, and intonation during oral and silent reading (Kuhn et al., 2010). A fluent reader can read proficiently for extended periods, does not require constant practice to maintain high performance, and can generalize his or her fluency to different texts (Hudson et al., 2005). Reading fluency is closely related to comprehension (Hudson et al., 2005; Kuhn et al., 2010). The strong relationship between fluency and comprehension makes sense when considering that automatization in word recognition allows the reader to free cognitive resources and redirect them to focus on higher-order processes such as comprehension (LaBerge & Samuels, 1974). Thus, fluent readers require shorter eye fixations, can perform longer jumps between fixations, and require fewer regressions than non-fluent readers (NICHD, 2000).

Fluent word-reading requires phonological, morphological, and orthographic processing skills. This research examined how these skills predict reading fluency, a critical aspect of literacy. Additionally, we explored the role played in this process by statistical

learning, the implicit ability to detect underlying regularities in the continuous stream of sensory stimuli (Frost et al., 2015).

The Dual Route Cascaded (DRC) model

Several computational models account for reading proficiency. The Dual Route Cascaded (DRC) model (Coltheart et al., 2001; Coltheart, 2005) has effectively explained two essential tasks in language processing: visual word recognition and reading aloud. This model consists of two routes: a lexical route and a nonlexical route.

The lexical route of the DRC model facilitates the pronunciation of a word by triggering several processes (Coltheart et al., 2001). First, the features of the letters that make up the word activate the letter units of the word (in parallel and across all letter positions). Then, these letters activate the word entry in the orthographic lexicon (i.e., a knowledge system about word forms). Third, the word entry in the orthographic lexicon, in turn, activates the corresponding word entry in the phonological lexicon. Finally, that word entry in the phonological lexicon activates the word's phonemes (in parallel, across all phoneme positions). Relevantly, the lexical route assumes that when reading a word, it is automatically matched against a mental lexicon that stores information about the spelling and pronunciation of letter strings that reflect actual words; thus, nonwords cannot be read correctly using this procedure because they are not stored in these lexicons (Coltheart, 2005).

On the other hand, the nonlexical route converts a sequence of letters into a sequence of phonemes using grapheme-phoneme correspondence rules (Coltheart et al., 2001). Some of these rules are context-sensitive (e.g., "c" becomes "/s/" when the following letter is "e," "i," or "y;" otherwise, it becomes "/k/") while others are position-sensitive (e.g., "y" is assigned the phoneme "/j/," "/l/," or "/ai/" depending on whether it is in an initial, medial, or final position). A third category includes phonotactic and morphemic rules. An example of a morphemic rule accounts for the regularity associated with the phoneme /s/ in a final position,

which phonotactic rules cannot account for. When such a phoneme is written “s” and denotes the plural morpheme, the phoneme must become “/z/” when preceded by several specific phonemes, even if this is not required by phonotactic considerations (Coltheart, 2005). The whole process involved in the nonlexical route is carried out serially from left to right, initially taking into account only the first letter of the sequence, then the first two letters, then the first three, and so on (Coltheart, 2005). Unlike the lexical route, the nonlexical route can correctly convert printed nonwords into sound (Coltheart, 2005).

The DRC model explains visual word recognition and reading aloud through a lexical and sublexical route. Both routes incorporate linguistic information processing at the phonological, lexical, sublexical, and morphological levels. Below we describe each type of processing and review the evidence linking them to performance on different literacy tasks.

Phonological awareness, a crucial predictor of language skills

Previous evidence shows that phonological processing plays a crucial role in learning to read (Byrne, 1996; National Reading Panel, 2000; Diamanti et al., 2017). One aspect of phonological processing that has received considerable attention in the past 50 years is *phonological awareness*, the ability to explicitly and deliberately access and manipulate language sounds (Stahl & Murray, 1994; Byrne, 1996). Phonological awareness is expressed, for instance, in the ability to rhyme, match initial consonants, and count the number of phonemes in spoken words (Stahl & Murray, 1994).

Phonological awareness is a prerequisite for the acquisition of the alphabetic principle (Byrne, 1996), decoding proficiency (National Reading Panel, 2000), and spelling in various orthographies (see Diamanti et al., 2017). Considering the importance of this skill for literacy development, it is not surprising that several studies have found that it predicts progress in learning to read in childhood (for a review, see National Reading Panel, 2000), a relationship that often persists after controlling for the effect of other several variables. While phonological

awareness is a crucial predictor of reading development and performance, other language skills are also critical in predicting learning and performance in reading and spelling. One of these skills is morphological awareness.

Morphological awareness

Morphemes are the smallest units of meaning within words (Deacon & Kirby, 2004) and contain phonological, semantic, and syntactic information (Görge et al., 2021). *Morphological awareness* (MA) reflects the understanding of the morphological structure of words and the ability to manipulate this structure (Carlisle, 1995). Such manipulation makes possible the creation of new words, for instance, by using derivational (e.g., “worker” = “work” + “er”), inflectional (e.g., “working” = “work” + “ing”), or compound (e.g., “worksheet” = “work” + “sheet”) morphology (Carlisle, 1995).

According to Deacon and Kirby (2004), MA plays a crucial role in many reading tasks. First, it is essential for single-word reading; for instance, a child seeking to pronounce the “ea” in “reading” and “react” will succeed by relying on morphemic boundaries. Second, morphology helps uncover the meaning of single words: the base and affix morphemes in “reading” and “react” provide helpful information about their meaning. Third, morphological problem-solving contributes to meaning construction in texts. Finally, MA might also be helpful in tasks involving pseudoword reading. In general, pseudowords are useful because they allow us to experimentally test what children can do with unfamiliar words. In addition, pseudowords also seem to have morphemic structures, just like actual words. For instance, it is possible to state that the pseudoword “lagician” is composed of a root (“logic”) and a suffix (“ian”).

Empirical evidence for the contribution of morphological awareness to literacy skills

Morphological awareness develops early (i.e., as early as six years) and in parallel to spoken language acquisition (Deacon et al., 2008; Carlisle et al., 2010). Furthermore, evidence

shows that it is a relevant predictor of performance in various literacy skills. For example, Deacon and Kirby (2004) investigated the role of MA in learning to read in a sample of English-speaking Canadian children. Specifically, they tested the impact of MA (assessed in grade 2) in single-word reading, pseudoword reading, and reading comprehension in grades 3, 4, and 5. After controlling for prior measures of phonological awareness, reading ability, and verbal and nonverbal intelligence, they found that MA contributes longitudinally to pseudoword reading and reading comprehension but not to single-word reading.

Desrochers and colleagues (2018) examined the relative contribution of MA (assessed at the beginning of grade 2) to the development of various literacy outcomes (assessed at the end of grade 2) in three alphabetic languages of different orthographic consistency (English, French, and Greek). They found that, after controlling for phonological awareness and rapid automatized naming, MA was a unique predictor of reading comprehension and spelling (English, French, and Greek), reading fluency (English and French), and reading accuracy (English). Additional multi-group analyses showed that the contribution of MA to literacy skills did not significantly differ by language.

Finally, Diamanti and colleagues (2017) researched longitudinally the role of phonological awareness and MA in developing literacy in the Greek language, which has a relatively consistent orthography. After controlling for expressive and receptive vocabulary skills, they found that MA contributed uniquely and significantly to all literacy outcomes (word and pseudoword reading accuracy, reading comprehension, and spelling) except for reading fluency.

To summarize, the empirical evidence suggests that MA is an essential antecedent of literacy in several languages. Moreover, the results linking MA to such outcomes include longitudinal studies and extend to pseudoword reading. However, the role of the MA would not necessarily be the same in all languages. One aspect that could impact the extent to which

MA influences different literacy skills is the depth or consistencies of orthographies (see Mousikou et al., 2020), the degree to which written language symbols (e.g., graphemes or letters) correspond to spoken language units (e.g., phonemes).

Morphological awareness in consistent and non-consistent orthographies

Morphological awareness contributes to word recognition (i.e., reading accuracy) in orthographies of varying consistency or transparency, such as English, Italian, Portuguese, and French (see Görgen et al., 2021). The relevance of MA in studies with English-speaking populations, particularly in the case of reading accuracy, makes sense considering that English is a non-consistent orthography (Görgen et al., 2021). In such orthographies, the lower grapheme-phoneme consistency (i.e., a single grapheme or letter can be pronounced in several ways) implies that resorting to larger units (i.e., whole words and morphemes) would be more effective than relying on sublexical units (i.e., individual graphemes) when successfully decoding a word (Nagy et al., 2003).

Unlike the association between MA and reading accuracy, the relationship between morphological skills and reading fluency has been less explored (Görgen et al., 2021). Studying such a relationship in a consistent orthography (e.g., German or Finnish) would be interesting because reading fluency and accuracy are not as closely related as in non-consistent orthographies (Görgen et al., 2021). This is because, in transparent orthographies, reading accuracy would reach ceiling levels after the first year of instruction (Seymour et al., 2003), thus allowing reading fluency to be assessed independently of reading accuracy (Görgen et al., 2021).

So far, studies that explored this relationship in consistent orthographies found that MA significantly predicts reading fluency (see Görgen et al., 2021). This finding suggests that to recognize unfamiliar words quickly, resorting to more extended written units would be more effective than decoding such words by breaking them down into isolated graphemes (Carlisle

& Stone, 2005). Such a strategy would be especially effective in processing long compound words, frequent in orthographies such as German (e.g., “Straßenbahnhaltestelle” – “street-train-stop-place” = “tram stop”; Görgen et al., 2021).

Few studies have examined the associations between MA and literacy skills in German school children. In one of them, Görgen et al. (2021) found that, after controlling for nonverbal cognitive skills, age, sublexical orthographic sensitivity, and phoneme awareness, MA was the most robust predictor of reading fluency and spelling. However, MA predicted more variability in spelling (6.8%) than in reading fluency (4.0%) scores. Additionally, the contribution of MA to spelling significantly increased from third to fourth grade. The authors suggest that the particular importance of MA in spelling would lie in the strong asymmetry of German orthography, which has a markedly lower phoneme-grapheme consistency (i.e., spelling direction) than grapheme-phoneme consistency (i.e., reading direction). In contrast, Haase and Steinbrink (2022) results suggest that MA plays a more limited role in German. This study found that phonological awareness was the main predictor of reading comprehension and word and pseudoword reading fluency (the other predictors were rapid naming, verbal memory, and vocabulary, all non-significant) and that MA did not predict additional variance in these outcomes.

Although MA is a relevant cognitive skill to explain literacy performance in writing systems of varying consistency, it is crucial to remember that results coming from single orthographies would not be necessarily generalizable to others (Haase & Steinbrink, 2022). Relevant to our study, German has a complex and varied morphology, especially concerning inflections and compounds (König & Gast, 2012). From a measurement standpoint, this suggests that MA would not necessarily be a unidimensional construct but would be composed of multiple factors, each reflecting a different type of morphology. If this is so, it is reasonable to think that different aspects of German morphology (e.g., inflectional, derivational, and

compound morphology) might differentially relate to literacy outcomes (Haase & Steinbrink, 2022).

Measuring morphological awareness: Some methodological issues

Finally, a methodological aspect that might impact the results obtained by using MA tasks is the insufficient psychometric properties reported in the literature. While some studies have obtained appropriate reliability values for MA tasks (Tighe & Binder, 2015; Diamanti et al., 2017; Haase & Steinbrink, 2022), others have reported alpha coefficients below .65 (Deacon & Kirby, 2004; Spencer et al., 2015; Desrochers et al., 2018; Görgen et al., 2021; James et al., 2021). Such low-reliability values suggest that the instruments used for measuring this construct contain significant statistical noise. Additionally, the extended underreporting of reliability values in the literature for MA tasks (Bratlie et al., 2022) exacerbates these psychometric problems. Considering these limitations and the fact that MA involves resorting to different types of morphology (e.g., derivational or inflectional), it will be essential to analyze the dimensionality of these measures and report their psychometric properties when using them.

Orthographic knowledge

In addition to phonological and morphological processing, one factor that has received attention for its importance in literacy development is *orthographic knowledge* or *processing*. Although the term *orthography* alludes to correct writing or spelling (Apel et al., 2019), orthographic knowledge is far from having an unambiguous definition in the literature (Apel et al., 2019; Rothe et al., 2014). While some authors have focused on the individual *ability* or *sensitivity* to the patterns of written language (Deacon et al., 2012) and the orthographic structure of words (Georgiou et al., 2008), others have emphasized the resulting *knowledge* of valid letter patterns (Perfetti, 1984) and regularities in the visual and orthographic aspects of written language (Roman et al., 2009). Although the lack of consistency in the definition and

measurement of orthographic knowledge has limited the understanding of its role in the development of literacy skills (Conrad et al., 2013), a cross-cutting aspect of this construct is that it reflects an understanding of the conventions of a written system (Deacon et al., 2012).

Word-specific orthographic knowledge

There is consensus that orthographic knowledge is a bi-dimensional construct (Apel et al., 2019; Conrad et al., 2013). The first dimension is *lexical* or *word-specific orthographic knowledge* (WOK) and reflects the representations stored in memory about the correct spelling of words and word units (Barker et al., 1992). A widespread way to measure WOK is an orthographic selection task (Olson et al., 1994) in which the child must demonstrate knowledge of specific words by selecting the correct word from two phonologically plausible alternatives, a pseudohomophone (e.g., “tertle”) and a correctly written word (e.g., “turtle”). Several studies have shown that WOK is a relevant predictor of reading and spelling skills in opaque orthographies such as English (Conrad et al., 2013; Deacon et al., 2012) and in transparent languages such as Dutch (Bekebrede et al., 2009). Regarding German, a recent study showed that WOK contributes to reading both at the basic (i.e., word reading) and complex (i.e., sentence and text reading) levels (Zarić et al., 2021).

General orthographic knowledge

The second dimension is *sublexical* or *general orthographic knowledge* (GOK). It involves an understanding of the rules or patterns of representing a sound or affix with letters, the rules about where certain letters may or may not appear in specific positions in words, as well as the norms that stipulate which letters may or may not be combined (Apel et al., 2019). General orthographic knowledge is usually assessed using a word similarity task in which the child must choose which of two alternative nonwords most closely resembles a real word (Rothe et al., 2014). This task usually poses one of the following two scenarios. In one of them, it is possible to assess sensitivity to frequent letter patterns by using two nonwords, one of

which contains a doublet that frequently occurs in written word forms while the other nonword contains a doublet that occurs rarely or never (e.g., “yill” - “yihh”; Cassar & Treiman, 1997). A second scenario also involves two nonwords, one of which contains a double consonant in a legal position, while the second word features a double consonant in an orthographically illegal position (e.g., “baff” – “bbaf”; Cunningham & Stanovich, 1993; Treiman, 1993). In German, the study of the contribution of GOK to reading performance has yielded mixed results. Rothe et al. (2014) and Zarić and colleagues (2021) found that GOK explained a significant amount of unique variance in reading and spelling performance at the end of first grade. In contrast, Görgen et al. (2021) reported that the amount of variance explained by GOK was significant for spelling but not for reading.

The relationship between word-specific and general orthographic knowledge

Unlike WOK, which consists of *explicit* representations about words or word units stored in memory, GOK involves *implicit* knowledge of the conventions about which letter combinations take place in a language. For example, many English words begin with the letter combination “st” while none begin with “sb” (Conrad et al., 2013). Empirical evidence shows that GOK develops earlier than WOK (see Apel et al., 2019) and significantly increases during the transition from kindergarten to first grade (Rothe et al., 2014). In addition to developing earlier, GOK would influence the development of WOK and is, in turn, influenced by the statistical orthographic properties regarding which letter combinations occur within a given language (see Pacton et al., 2005; Deacon et al., 2008). However, the specific mechanisms linking both types of orthographic processing to reading and spelling remain largely unexplored.

Recently, Rothe and colleagues (in preparation) found that WOK facilitates the relationship between GOK and reading and spelling. Furthermore, the correlations in this study confirmed that literacy skills show higher associations with WOK than with GOK. Likewise,

this study obtained psychometric evidence on the importance of distinguishing between legal vs. illegal items for WOK and GOK. They also found that the relationship between spelling and reading with WOK and GOK varies by item type, with higher correlations for literacy skills with pseudohomophones (vs. words) and illegal (vs. legal) pseudowords.

While there is ample evidence of the predictive role of phonological, morphological (MA), and orthographic (GOK and WOK) information processing skills in performance on various reading and spelling tests, the mechanisms at the basis of the acquisition and mastery of such skills have been less explored in the empirical literature. One mechanism that could help to understand the development of these skills is statistical learning.

Statistical learning, a potential mechanism explaining language skills acquisition

Statistical learning (SL) is the implicit ability that allows cognitive systems to automatically detect underlying structures in the continuous stream of sensory stimuli and extract the distributional properties of that information (Frost et al., 2015). Initially proposed as an attempt to explain the infant's ability to detect hidden words in a continuous artificial language (see Saffran et al., 1996), SL theories emerge as a domain-general theoretical alternative to the Chomskyan domain-specific explanation of language acquisition. Instead of assuming an innate, modular, neurobiologically programmed human capacity, SL assumes a general mechanism at the basis of learning and processing of sensory stimuli of various kinds over time and space (Frost et al., 2015).

While there is ample evidence that the SL underlies several cognitive functions, such as segmentation of continuous auditory input, visual search, contextual cueing, visuomotor learning, conditioning, and generally any predictive behavior (see Frost et al., 2015), its role in the acquisition and development of language skills is less clear.

Statistical learning and sensitivity to sublexical and morphological regularities

Two types of regularities have previously been linked to statistical learning, orthographic regularities at the sublexical level and morphological regularities. More specifically, it has been argued that children's spelling is influenced from a very early age by sublexical and morphological patterns and that writers draw not only on rule-based knowledge but also on implicit and automatic processing of statistical regularities (Pacton et al., 2005; Deacon et al., 2008; Deacon & Leung, 2013). The implicit and statistically regularity-based nature of both abilities contrasts, for example, with that of lexical-level processing (WOK), which requires individual strategies based on memorization and self-teaching (Share, 2008).

Concerning statistical knowledge of sublexical regularities (i.e., GOK), evidence suggests that children are sensitive from a very early age to which letters may or may not be doubled, beyond their knowledge of letter-sound mappings and in contradiction to models suggesting that phonological representations dominate the spelling of very young children (Deacon et al., 2008). For example, Treiman (1993) reported that children use doublets in spelling pseudowords (e.g., "ff") more recurrently in word-intermediate and word-final positions than in word-initial places, in line with implicit conventions of the English language. Cassar and Treiman (1997) found that very young English-speaking children (i.e., six years and older) show sensitivity to letters that could and could not be used as doublets: specifically, they were more likely to choose (as words) pseudowords containing two identical vowels (e.g., "geed" vs. "gaad") or consonants (e.g., "baff" vs. "bahh") that frequently went together in their language. Additionally, Thompson et al. (1999) found that, in the absence of explicit decoding training, children make spelling errors that reflect the statistics of the texts used in their classes. More recently, Apel et al. (2019) pointed out that children employ statistical learning (i.e., implicit knowledge of orthographic regularities of words, such as bigram frequencies and standard vs. unfrequent orthographic patterns) in the acquisition of general orthographic

knowledge. Overall, this evidence suggests that people are sensitive from a very early age to sublexical conventions based on statistical regularities that are not usually explicitly addressed during their school instruction.

Regarding the role of SL in learning morphological regularities, there is evidence that the spelling of very young writers reflects the morphemic structure of words (Deacon & Leung, 2013). For example, very young children are more likely to include the spelling of the penultimate sound of a word when the sound is part of a root than when it is not (e.g., /n/ in “rained” vs. “brand”; Treiman et al., 1994) and are more able to spell the initial elements in two-morpheme words than in one-morpheme words (e.g., “turn” in “turning” and “turnip”; Deacon & Bryant, 2005). Similar to what happens with GOK, the evidence for the influence of morphemic structure on spelling is at odds with rule-based models of morphological learning, according to which very young children’s spelling would be determined almost exclusively by phonological aspects (Deacon & Leung, 2013). In addition, as reported by Kemp & Bryant (2003), even adults do not exclusively use morphological rules even when possible: adult spelling also reflects sensitivity to different regularities of written language (e.g., the effect of adjacent sounds in pseudoword spelling).

Assessing statistical learning and some conflicting results

A common way to assess SL involves using visual tasks to measure sensitivity to statistical patterns. Some studies have examined the relationship of such sensitivity to reading performance. For example, after controlling for age in a sample of children and adults, Arciuli and Simpson (2012) found that participants with greater sensitivity to the statistical patterns of a visual task also showed better reading performance. In the same vein, Frost et al. (2013) found a positive correlation between reading ability and performance on one of these SL visual tasks in an English-speaking sample that were learning Hebrew as a second language.

While these studies suggest a relationship between SL and reading ability, other authors have reported results that challenge this conclusion. In an attempt to replicate the findings of Arciuli and Simpson (2012) and Frost et al. (2013), Schmalz et al. (2019) not only found no correlations between the two tasks used to measure SL with reading ability and bigram sensitivity but also found no significant associations between both SL tasks. The authors point out that the difficulty in replicating the results obtained by Arciuli and Simpson (2012) and Frost et al. (2013) is partly due to the low correlations between SL tasks. Additionally, Schmalz et al. (2019) suggest that such an absence of a relationship between SL tasks could be related to psychometric problems in both measurements.

The psychometric problems present in many of the tasks used to measure individual differences in SL derive from the fact that practically all of them were initially conceived to be used in group-level studies (Siegelman et al., 2017). Specifically, these tasks show three significant limitations: (1) a low number of trials during the test phase (or the repeated use of the same targets during the test); (2) a chance-level performance in most of the samples used, which leads to most of the data points reflecting noise; and (3) that almost all the test items used after the familiarization phase are identical and of the same level of difficulty. According to Siegelman and colleagues (2017), these three factors would make SL tasks particularly vulnerable to measurement error effects, especially low reliability and validity problems.

Aims of the present study and hypotheses

To summarize, phonological, morphological, and orthographic skills are relevant predictors of literacy skills in different languages. While the importance of phonological awareness is undisputed, this is not necessarily the case for MA and orthographic knowledge. Evidence from the German language suggests that the contribution of MA to spelling is higher than to reading fluency; furthermore, recent results indicate that, unlike PA, MA would not necessarily be a relevant predictor of reading comprehension and fluency of words and

pseudowords. A possible explanation for these conflicting results is that measures of MA are particularly prone to capturing statistical noise. Alternatively, MA scores might reflect an underlying multidimensional structure. Concerning orthographic knowledge, GOK has correlated with literacy skills to a lesser extent than WOK. Likewise, while the relationship of WOK with reading in German is positive, GOK has shown mixed results. Furthermore, there is also contradictory evidence regarding the role of SL as a predictor of literacy skills. Finally, measures of SL show severe psychometric limitations.

The present study aims to model the relationships between MA, orthographic processing at the lexical (WOK) and sublexical (GOK) levels, and age with reading fluency. To this end, we use structural equation modeling on a large-scale dataset with data from 2,624 children. Using confirmatory factor analyses, we investigated whether the responses to the three tasks used to measure MA fit a unidimensional or a three-dimensional model (i.e., derivational, inflectional, and plural formation). Regarding the SEM model, we hypothesize that higher MA (either as a unidimensional or multidimensional construct; [H]ypothesis 1), lexical processing (a latent variable made up of a WOK and a spelling task; H2), GOK (H3), and age (H4) positively predict reading fluency. Additionally, we expect the value of the SEM regression path for GOK to be lower than for WOK (H5).

Finally, considering that critical methodological limitations hamper SL measurement, we chose to evaluate the role of this mechanism using a novel approach. Instead of using a particular task that measures SL, we introduced an additional parameter in our structural model for a correlation between MA and GOK (skills that SL would facilitate). In this regard, if SL is a relevant mechanism underlying both MA and GOK, a model that includes this relationship should better fit the data than a model that does not (H6).

Method

Participants

We invited 52,734 families with children in the third or fourth grade to participate in an extensive survey conducted in the German federal states of Hesse ($n = 25,000$) and Bavaria ($n = 27,734$). They had to do this by logging onto a web-based application where, over the course of five sessions, each lasting between 30 and 45 minutes, the children's scholastic abilities and psychopathological profile would be evaluated. The Bavarian regional registry offices and the Hessian Ministry of Culture chose the families that were contacted randomly. Every exam and questionnaire was processed independently from home using a tablet or smartphone within eight weeks. The study, which was carried out following the most recent version of the Declaration of Helsinki and in compliance with national law, received approval from the ethics committees of the DIPF | Leibniz Institute for Research and Information in Education, Frankfurt am Main (Project ID: FoeDises; approval date: April 2, 2017) and the University Hospital of the Ludwig Maximilian University Munich (Project ID: 438-16; approval date: August 25, 2016).

Of the 4,621 families who downloaded the app, 4,542 consented to the study's procedures and provided details about their kids' demographics. We took out 1,598 kids whose information did not pass the usual test plausibility checks or was incomplete by the fifth session on all pertinent examinations. For plausibility checks, we rejected data with unreasonably quick or slow response times (for more information, see Visser et al., 2020). Additionally, we eliminated 26 children with an IQ score of 70 or lower and 230 kids who did not answer every question of the IQ test (CFT20-R; Weiß & Weiß, 2008). Finally, we eliminated 63 cases whose parents claimed the infant suffered from neurological disorders, chromosomal problems, hearing or visual issues, or both. In sum, we excluded 1,918 cases, resulting in a final sample

of 2,624 children. Participants were evenly distributed regarding grade and gender (3rd Grade: Boys = 608, Girls 630; 4th Grade: Boys = 687, Girls = 699).

Procedure

For the present study, the software company Meister Cody GmbH transformed all standardized psychometric tests and pencil-and-paper questionnaires into an online tablet/smartphone version. Families downloaded and used the web-based application independently on their own devices for eight weeks. The application of the instruments to children took place during five sessions. Each session took place on different days and lasted 30-45 minutes. Once the application began, the children received guidance from the wizard Meister Cody, who provided detailed information per task. At the beginning of each test, the children had to answer two to four practice items before it began. Also, depending on the task, an automatic reminder was triggered if the child paused participation for 15 or 25 seconds.

Measures

Reading fluency

We assessed reading fluency performance using digitized versions of two standardized tests. The first was the “Wuerzburger Silent Reading Test-Revised” (WLLP-R; Schneider et al., 2011; parallel-test reliability: $r = .93$ for 3rd grade and $r = .82$ for 4th grade). Children read up to 180 simple, high-frequency words silently and choose the correct picture (out of four options) representing each word. The task lasted 5 minutes. Each child’s total number of correct images was calculated and used in the subsequent statistical analyses. The second standardized reading test was the “Verlaufsdiagnostik Sinnerfassenden Lesens” [Diagnosis of progress in reading comprehension]’ (VSL; Walter, 2013; parallel-test reliability: $.77 < r < .86$). Participants must read a text in which there is a parenthesis every seven words with three possible options. The objective is to circle the appropriate word. The test has 20 booklet-type

texts. We calculated the total number of correct answers for each child and used this score in subsequent analyses.

Spelling

We assessed spelling ability using the extended version of the standardized “Weingarten spelling test for basic vocabulary” (WRT 3+ for 3rd graders; Birkel, 2007a; parallel-test reliability: $r > .91$, and WRT4+ for 4th graders; Birkel, 2007b; parallel-test reliability: $r > .90$). In this test, children typed a target word (3rd grade: 55 words; 4th grade: 60 words) using the virtual keyboard on their devices’ screen. Both the target word and the sentence frames were read aloud. Children could correct their typing and had to confirm each word by ticking in a box. The total number of correctly spelled target words was calculated and standardized to summarize the spelling performance of the 3rd and 4th graders. Because the standards in the manual are based on the pencil-and-paper test, we developed standards based on the entire sample that responded to the test in the web-based application (for more information on the development of the standards, see Visser et al., 2020).

Word-specific orthographic knowledge (WOK)

We measured WOK with an orthographic word decision task. After telling children that they would see words, their task consisted in indicating whether the stimulus was spelled correctly (green tick) or not (red cross). Two types of stimuli that sound like real words were presented: words and pseudohomophones. Regarding the former, orthographically and phonologically familiar German words were used. Regarding the pseudohomophones, these were phonologically familiar, although orthographically unfamiliar. All words were high-frequency, according to the childLex corpus (Schroeder et al., 2015). Pseudohomophones were derived from these words, but had a phonologically identical grapheme exchanged. Both types of stimuli were matched in the number of letters, bigram frequencies, and trigram frequency. For this, childLex was used as a reference. The test consisted of 30 words and 30

pseudohomophones. The 60 stimuli were divided into two blocks (15 words and 15 pseudowords per block) and presented in a pseudorandomized fashion to control that no more than four stimuli of the same type were presented consecutively. Each stimulus was presented alone on the screen once. Before the start of the task, practice items were presented. The correct responses were calculated and used in the subsequent statistical analyses. The alpha coefficient of the total task was adequate ($\alpha = .83$). However, internal consistency was higher for pseudohomophones ($\alpha = .81$) than for words ($\alpha = .58$).

General orthographic knowledge (GOK)

We assessed GOK with a pseudoword orthographic decision task. After telling children that they would see pseudowords, they were asked to indicate whether the stimulus could be a German word (green tick) or not (red cross). Two types of orthographically and phonologically unfamiliar stimuli were used: legal pseudowords (30 stimuli) and illegal pseudowords (30 stimuli). Regarding the legal pseudowords, these contained strings of letters obeying German orthography. Illegal pseudowords, on the other hand, included letter strings that violated German orthography. The pseudowords were derived from the words used in the WOK task, to which one grapheme per syllable was exchanged. Legal pseudowords and illegal pseudowords contained the same number of letters. The 60 items were divided into two blocks (15 frequent pseudowords and 15 infrequent pseudowords per block). Each stimulus was presented once in the middle of the screen. The presentation took place in a pseudorandomized fashion to control that no more than four stimuli of the same type were presented consecutively. Practice items appeared at the beginning of the task. The number of correct responses was computed and used in subsequent analyses. The alpha coefficient was adequate for the complete task ($\alpha = .84$). Within it, internal consistency was higher for legal pseudowords ($\alpha = .89$) than for illegal pseudowords ($\alpha = .74$).

Morphological awareness (MA)

We assessed MA using pseudowords in three domains, consisting of 8 items each. The first domain, *plural formation*, measured the ability to select the correct plural form from three options. After showing a picture of a fantastic creature and introducing it (e.g., *Das ist ein Roto* - That is a Roto), children had to select the correct name for two other identical creatures while watching the corresponding picture (e.g., *Das sind zwei ...* These are two “Roto” – “Roton” – “Rotos”). The second domain, *derivation*, assessed the ability to recognize derivational changes in words. Children read and listened to a sentence containing a pseudoword (e.g., *Wir kresen* - We kresen). After this, adapted sentence frames were presented and asked to select the appropriate derived form of the previously used pseudoword that best matched the frame (e.g., *Das war eine tolle ...* - This was a great ... “Kreserei” – “Krest” – “Kresel”). For the third domain, *verb inflection*, children were asked to identify the correct inflection of a pseudoverb. First, the infinitive of a verb (e.g., *dannen*) was shown orally and in writing. After this, a sentence frame appeared, and the child had to choose the appropriate inflected verb from three alternatives (e.g., *Ben ...* “dannt” – “dannst” – “dannen”). All pseudowords followed the structure of actual German words. The answer alternatives resembled the target answers. However, they were either illegal in the specific sentence frame, irregular, infrequent, or did not reflect proper word forms according to German language conventions. We report the complete psychometric analysis of this task in the Results section.

Statistical analysis

Descriptive statistics and correlations

We computed descriptive statistics (*M* and *SD*) for all study variables for grades 3 and 4 and the total sample. We also estimated Pearson correlations to know the magnitude of the associations between study variables in the total sample.

Confirmatory factor analysis

Using the total sample ($N = 2,624$), we conducted confirmatory factor analyses (CFA) to examine the dimensionality of the scores from the three morphological domains (plural formation, verb inflection, and derivation). First, we fitted a one-dimensional and a three-dimensional model. We retained the best-fit model and made some modifications to it. Due to the ordinal nature of the data, we employed the diagonally weighted least squares (DWLS) estimator in its robust weighted least squares-mean and variance adjusted (WLSMV) variant. We used the following cut-off points as a reference to evaluate the fit indices of each model: Comparative Fit Index (CFI) $\geq .95$, Root Mean Square Error of Approximation (RMSEA) $\leq .06$, and Standardized Root Mean Square Residual (SRMR) $\leq .08$ (Hu & Bentler, 1999). Even though these cut-off points are widely used in the empirical literature, we chose to follow a flexible approach in their interpretation because using rigid rules in SEM might lead to type I errors (Marsh et al., 2004). We did not emphasize the interpretation of the p -value associated with the chi-square test due to its high sensitivity to sample size. However, we must highlight that a significant p -value suggests the presence of model misspecifications (Ropovik, 2015). Likewise, we used the scaled chi-square difference test proposed by Satorra (2000) for comparing nested models. Finally, we report standardized factor loadings (λ), factor correlations (ϕ), and McDonald's omega (ω ; McDonald, 1999) reliability coefficients of the final model.

Predicting reading fluency using Structural Equation Modeling

Then, we fitted several models using structural equation modeling (SEM) to estimate how morphological (MA, H1), lexical processing (H2), and sublexical orthographic knowledge (GOK, H3), as well as age (H4), predict reading fluency performance. We also expected the value of the SEM regression path for GOK to be lower than for WOK (H5). Additionally, we tested if SL plays a role in some of these relationships (H6).

For this, we randomly divided the total sample into two subsamples, the first containing 80% of the cases (*training set*) and the second 20% (*testing set*). The training set ($n_1 = 2,109$) was used to estimate model parameters and obtain global fit indices for the SEM models. The testing set ($n_2 = 515$) was used to cross-validate the results.

Because some of the tests consisted of many items, we used *parceling* (Little et al., 2002; Matsunaga, 2008) to generate the model's observed variables. For MA, sum scores from each domain (after the modifications made during the CFAs) were used as indicators of the latent factor. We proceeded similarly with the rest of the latent variables. For lexical processing, we used two indicators. One of them, WOK, corresponds to the sum of scores obtained from an orthographic word decision task and the second consists of the standardized scores in a standardized spelling test (WRT 3+/WRT 4+). The total GOK score was included in the model as a single observed variable. On the other hand, the reading fluency latent variable was made up of the total score in two standardized tests, WLLP-R and VSL. Thus, each indicator of the latent variables MA, WOK, and reading fluency corresponded to the total score or standardized score of the items that made up the corresponding test. Finally, it is worth mentioning that for both lexical and sublexical orthographic knowledge, we used only the illegal items of their corresponding tasks (WOK: pseudohomophones, GOK: illegal pseudowords). The rationale behind this decision is that the association of these types of stimuli with reading fluency was higher than in the case of the legal items (WOK: words, GOK: legal pseudowords).

We report robust versions of the RMSEA and the CFI fit indices based on the mean-corrected Satorra-Bentler test statistic. Due to the continuous nature of the manifest variables and the possible absence of multivariate normality, we used maximum likelihood estimation with robust standard errors and a Satorra-Bentler scaled test statistic (MLM). Again, we refer to the cut-off points for the fit indices suggested by Hu and Bentler (1999) and adopt a flexible

approach to their interpretation (Marsh et al., 2004). Model comparison between nested models also considered the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC).

We evaluated two theoretical models using the training data. The first model included all regression paths from language skills and age to reading fluency (H1 to H5). In the second model, we tested the hypothesis that SL would be relevant to the performance on two tasks, MA and GOK. To this end, we added a parameter corresponding to the association between the two constructs, expecting that this model would fit the data better than the first one (H6). Finally, we refitted both models and compared their fit. Once again, we emphasized the AIC and BIC values for model comparison.

All analyses were conducted using the R language (ver. “4.2.0”; R Core Team, 2022). We employed the psych package (ver. “2.2.5”; Revelle, 2022) for computing Cronbach’s alpha internal consistency coefficients. We used the lavaan package (ver. “0.6.11”; Rosseel, 2012) for estimating and comparing the CFA and SEM models. For estimating McDonald’s reliability coefficients, we employed the semTools package (ver. “0.5.6”; Jorgensen et al., 2022), and for plotting the path diagram figures, we used semPlot (ver. “1.1.5”; Epskamp, 2022).

Results

Descriptive statistics and correlations

Descriptive statistics (M and SD) for grades 3 and 4 and the total sample are shown in Table 1. Also, Table 2 shows the correlations between all study variables.

Table 1

Descriptive statistics (M and SD) for grades 3 & 4 and the total sample

	3rd grade (N = 1,238)	4th grade (N = 1,386)	Total (N = 2,624)
Age in months	110.6 (4.3)	121.9 (4.5)	116.6 (7.2)
Reading 1 (WLLP-R)	93.5 (16.2)	104 (15.0)	99.1 (16.4)
Reading 2 (VSL)	22.1 (7.1)	27.4 (7.7)	24.9 (7.9)
Spelling (WRT) ^a	53.8 (10.2)	52.4 (9.8)	53.1 (10.0)
MA	19.5 (2.6)	20.2 (2.2)	19.9 (2.4)
Plural formation	5.7 (1.2)	6 (1.2)	5.9 (1.2)
Inflection	7.2 (1.1)	7.5 (0.9)	7.3 (1.0)
Derivation	6.6 (1.2)	6.8 (1.1)	6.7 (1.2)
WOK	53.3 (5.6)	55.5 (4.2)	54.5 (5.0)
Words	27.9 (2.2)	28.5 (1.6)	28.2 (1.9)
Pseudohomophones	25.4 (4.1)	27 (3.2)	26.3 (3.7)
GOK	46.4 (7.5)	48 (7.1)	47.2 (7.3)
Legal pseudowords	23.5 (5.7)	24.4 (5.7)	23.9 (5.8)
Illegal pseudowords	22.9 (4.1)	23.6 (3.8)	23.3 (3.9)

Note. ^a t-score

Table 2*Correlations between study variables with confidence intervals*

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Age in months													
2. Reading 1 (WLLP-R)	.26** [.22, .29]												
3. Reading 2 (VSL)	.26** [.22, .29]	.72** [.70, .74]											
4. Spelling (WRT)	-.09** [-.13, -.06]	.41** [.38, .44]	.48** [.45, .51]										
5. MA.	.10** [.06, .14]	.31** [.28, .35]	.31** [.28, .35]	.39** [.36, .42]									
6. Plural formation	.09** [.05, .13]	.20** [.16, .23]	.22** [.18, .25]	.28** [.24, .31]	.71** [.69, .73]								
7. Inflection	.07** [.03, .11]	.29** [.25, .32]	.28** [.24, .32]	.34** [.31, .38]	.71** [.69, .72]	.26** [.22, .29]							
8. Derivation	.06** [.02, .10]	.19** [.16, .23]	.18** [.14, .22]	.23** [.19, .27]	.72** [.70, .74]	.20** [.16, .24]	.33** [.30, .37]						
9. WOK	.17** [.13, .20]	.42** [.39, .46]	.45** [.42, .48]	.54** [.52, .57]	.41** [.37, .44]	.25** [.22, .29]	.37** [.34, .40]	.26** [.22, .29]					
10. Words	.11** [.07, .15]	.28** [.25, .32]	.30** [.26, .33]	.36** [.32, .39]	.30** [.27, .33]	.18** [.14, .21]	.28** [.24, .31]	.20** [.16, .24]	.77** [.76, .79]				
11. Pseudohomophones	.17** [.13, .21]	.42** [.39, .45]	.46** [.42, .49]	.55** [.52, .57]	.39** [.36, .42]	.25** [.21, .29]	.36** [.32, .39]	.24** [.20, .28]	.95** [.94, .95]	.53** [.50, .55]			
12. GOK	.09** [.05, .13]	.11** [.07, .15]	.13** [.09, .17]	.19** [.15, .22]	.21** [.17, .24]	.13** [.09, .17]	.18** [.14, .22]	.14** [.10, .18]	.22** [.18, .25]	.18** [.14, .22]	.20** [.16, .23]		
13. Legal pseudowords	.06** [.03, .10]	.06** [.03, .10]	.07** [.03, .10]	.09** [.05, .13]	.16** [.12, .20]	.11** [.07, .14]	.14** [.10, .17]	.10** [.07, .14]	.10** [.06, .14]	.10** [.06, .14]	.08** [.04, .12]	.85** [.84, .86]	
14. Illegal pseudowords	.07** [.03, .11]	.11** [.07, .15]	.14** [.10, .18]	.21** [.18, .25]	.15** [.11, .19]	.08** [.05, .12]	.13** [.10, .17]	.11** [.07, .15]	.26** [.22, .29]	.19** [.15, .22]	.25** [.21, .28]	.62** [.60, .65]	.11** [.07, .15]

Note. Values in square brackets indicate the 95% confidence interval for each correlation.* $p < .05$. ** $p < .01$.

Confirmatory factor analysis

Morphological awareness

Unidimensional vs. three-dimensional models

First, we tested and compared two measurement models using the responses to the MA items. In Model 1, we tested a unidimensional structure. In contrast, in model 2, we assessed a three-dimensional structure based on the three domains assessed in the task (plural formation, verb inflection, and derivation). The fit of the one-dimensional model was deficient, $\chi^2_{\text{Scaled}} (252, N = 2,624) = 1,056.46, p < .001, \text{CFI}_{\text{Scaled}} = .78, \text{RMSEA}_{\text{Scaled}} = .04, 90\% \text{ CI } [.03, .04], \text{SRMR} = .08$, whereas the three-dimensional model fit the data better, $\chi^2_{\text{Scaled}} (249, N = 2,624) = 705.77, p < .001, \text{CFI}_{\text{Scaled}} = .87, \text{RMSEA}_{\text{Scaled}} = .03, 90\% \text{ CI } [.03, .04], \text{SRMR} = .08$. This improvement was statistically significant, $\chi^2_{\text{Scaled difference}} (3, N = 2,624) = 247.77 p < .001$.

Final two-dimensional model

Considering that the model with three latent factors performed better, we used it as a basis for exploring and implementing modifications in its internal structure. Thus, after observing that six of the eight items that made up the plural formation dimension had low standardized loadings ($\lambda < .20$), we opted to remove this dimension and re-estimate the model with the two remaining factors, inflection, and derivation. This two-dimensional model (Model 3) showed an improved fit to the data, $\chi^2_{\text{Scaled}} (103, N = 2,624) = 245.13, p < .001, \text{CFI}_{\text{Scaled}} = .93, \text{RMSEA}_{\text{Scaled}} = .02, 90\% \text{ CI } [.02, .04], \text{SRMR} = .07$. Upon inspection of the two-dimensional model, we identified poor standardized loadings for items 2 ($\lambda = .26, p < .001$) and 5 ($\lambda = -.00, p = .96$) from the derivation factor. For this reason, we chose to re-estimate the model by omitting both items. The fit to the data of this final model (Model 4) was adequate, $\chi^2_{\text{scaled}} (76, N = 2,624) = 144.85, p < .001, \text{CFI}_{\text{Scaled}} = .96, \text{RMSEA}_{\text{Scaled}} = .02, 90\% \text{ CI } [.01, .02], \text{SRMR} = .06$. The correlation between both latent factors was high ($\phi = .83, p < .001$) and all standardized loadings (inflection = $.42 < \lambda < .75$; derivation = $.45 < \lambda < .70$) were statistically

significant at the .001 level. Both the total instrument ($\omega = .62$) and its subdimensions (inflection: $\omega = .47$; derivation: $\omega = .48$) shown to have reliabilities below the rule of thumb of .70.

Finally, we recalculated the sum scores of the derivation and inflection items without items 2 and 5 of the first factor for their inclusion in further analysis. Due to the high correlation between both dimensions, we included both sum scores as indicators of a single MA latent variable in the subsequent structural models.

Predicting reading fluency using Structural Equation Modeling

Training set

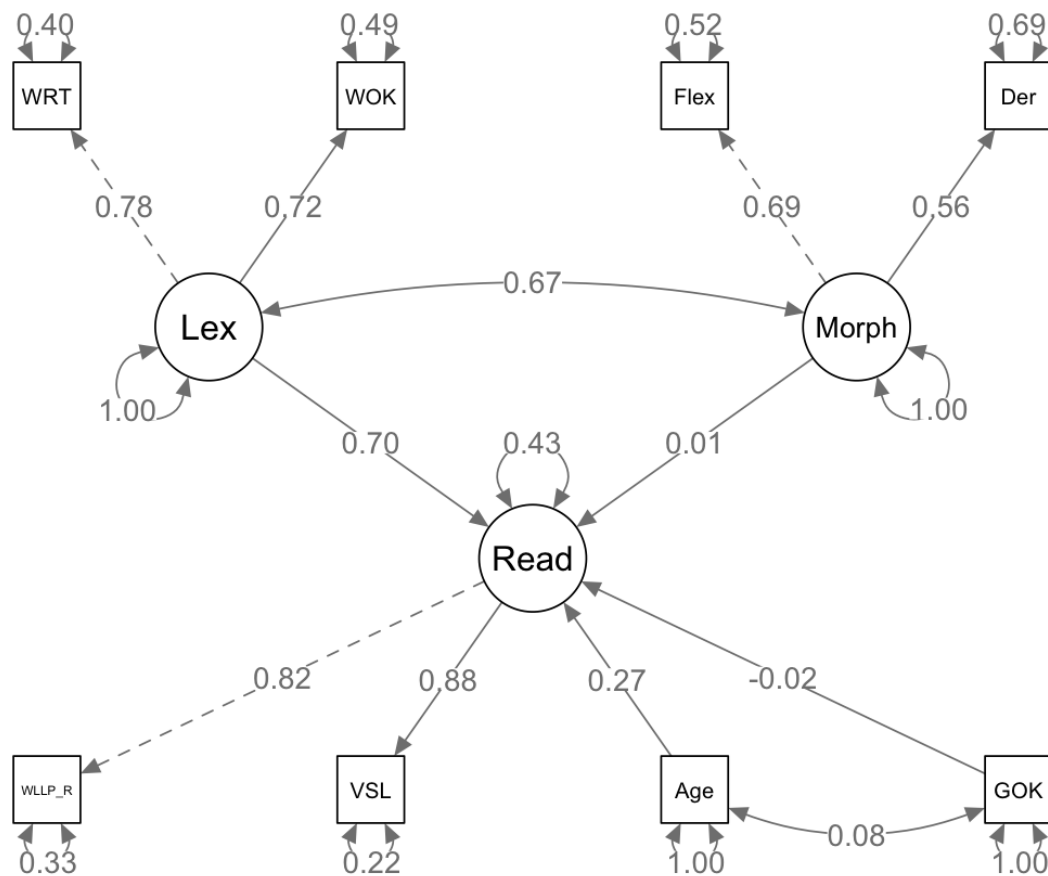
We estimated two SEM models using the training set containing 80% of the data. In the first model (Model 1, see Figure 1), we evaluated the predictive relationships of lexical processing, MA, GOK and age with reading fluency and obtained the following fit indices, $\chi^2_{\text{Scaled}} (16, n_1 = 2,109) = 355.78, p < .001, \text{CFI}_{\text{Robust}} = .92, \text{RMSEA}_{\text{Robust}} = .10, 90\% \text{ CI } [.09, .11], \text{SRMR}_{\text{Bentler}} = .08$. Only lexical processing ($\beta = .70, p = .001$) and age ($\beta = .27, p < .001$) significantly predicted reading fluency performance. In contrast, neither MA ($\beta = .01, p = .841$) nor GOK ($\beta = -.02, p = .382$) were relevant predictors of reading fluency. Finally, the association between MA and lexical processing ($\phi = .67, p < .001$) and between age and GOK ($\phi = .08, p < .001$) were statistically significant.

In Model 2 (see Figure 2), we evaluated the same predictive relationships, adding a parameter corresponding to the correlation between MA and GOK. In this way, we could test the role of SL in the structural model. This model obtained a similar fit to Model 1 in most indices, $\chi^2_{\text{Scaled}} (15, n_1 = 2,109) = 343.31, p < .001, \text{CFI}_{\text{Robust}} = .92, \text{RMSEA}_{\text{Robust}} = .11, 90\% \text{ CI } [.10, .12], \text{SRMR}_{\text{Bentler}} = .08$. Likewise, the correlation that sought to assess the role of SL in MA and GOK was significant, although small ($\phi = .06, p = .034$). Since most indices performed very similarly between Model 1 and Model 2, we looked at the AIC and BIC of each

model. Considering both fit indices, Model 1 (AIC = 93,399; BIC = 93,512) and Model 2 (AIC = 93,396.66; BIC = 93,515) showed an almost identical fit to the data. We then checked for statistical differences between the performance of both models using a Scaled Chi-Squared Difference Test. Although we obtained a statistically significant difference in favor of Model 2, $\chi^2_{\text{Scaled difference}}(1, N = 2,624) = 3.99$ $p = .04578$, this difference was marginal.

Figure 1

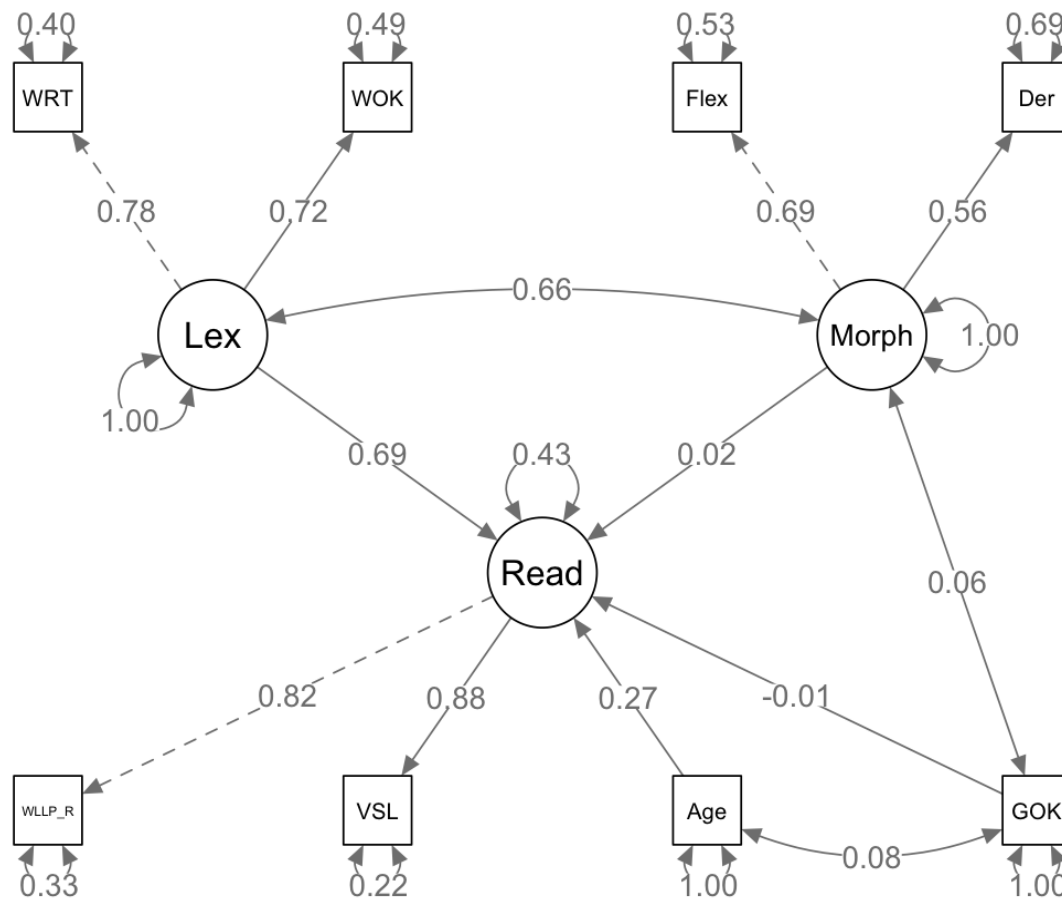
Model 1. Reading fluency predicted by linguistic skills and age, using the training set



Note. Latent variables: *Lex* = Lexical processing, *Morph* = Morphological awareness, *Read* = Reading fluency. Observed variables: *WRT* = Spelling test, *WOK* = Word-specific orthographic knowledge, *Flex* = Morphological inflection, *Der* = Morphological derivation, *WLLP_R* = Reading fluency test 1, *VSL* = Reading fluency test 2, *GOK* = General orthographic knowledge

Figure 2

Model 2. Reading fluency predicted by linguistic skills and age, with an additional SL parameter for the correlation between MA and GOK, using the training set



Note. Latent variables: *Lex* = Lexical processing, *Morph* = Morphological awareness, *Read* = Reading fluency. Observed variables: *WRT* = Spelling test, *WOK* = Word-specific orthographic knowledge, *Flex* = Morphological inflection, *Der* = Morphological derivation, *WLLP_R* = Reading fluency test 1, *VSL* = Reading fluency test 2, *GOK* = General orthographic knowledge

Testing set

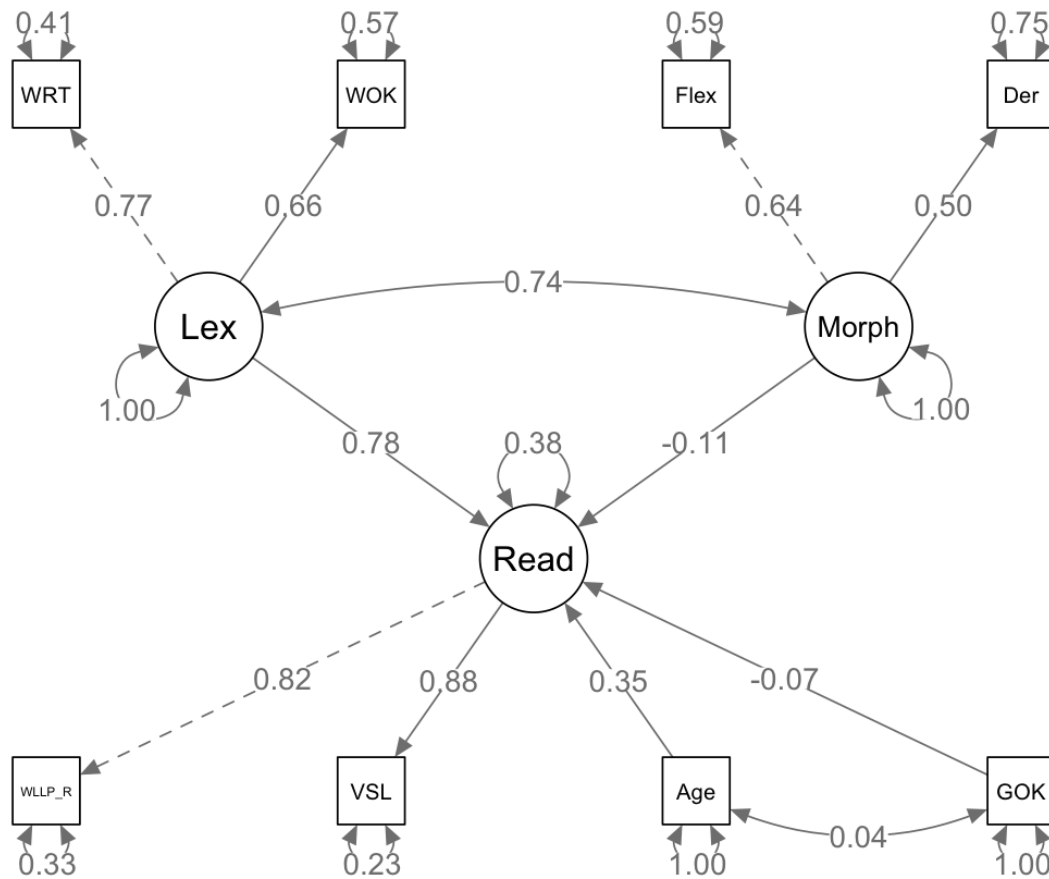
We re-evaluated and compared models 1 (Figure 3) and 2 (Figure 4) on the testing set, which contained 20% of the data. The fit of the first model, $\chi^2_{\text{Scaled}}(16, n_2 = 515) = 99.99, p < .001$, $\text{CFI}_{\text{Robust}} = .90$, $\text{RMSEA}_{\text{Robust}} = .11$, 90% CI [.09, .13], $\text{SRMR}_{\text{Bentler}} = .09$, was similar to

that previously obtained in the training set. Likewise, the statistical significance of almost all parameters of model 1 obtained in the training set was stable in the testing set, with two exceptions: first, the correlation between age and GOK was no longer significant in this dataset ($\phi = .04, p = .364$), and second, unlike the non-significant result of the GOK path coefficient to reading fluency in the training set, this predictive coefficient was significant in the testing set ($\beta = -.07, p = .040$). Regarding the fit of model 2, $\chi^2_{\text{Scaled}} (15, n_2 = 515) = 97.92, p < .001$, $\text{CFI}_{\text{Robust}} = .90$, $\text{RMSEA}_{\text{Robust}} = .11$, 90% CI [.09, .13], $\text{SRMR}_{\text{Bentler}} = .08$, it did not differ much from that obtained using the training set. In contrast to the training set, the correlations between age and GOK ($\phi = .03, p = .523$) and between MA and GOK were not statistically significant in the testing set ($\phi = .09, p = .121$). Nevertheless, the statistical significance of all other parameters was the same as in the training set. Finally, AIC and BIC values for Model 1 (AIC = 22,706; BIC = 22,791) and Model 2 (AIC = 22,706, BIC = 22,795) were almost identical and the Scaled Chi-Squared Difference Test yielded a non-significant result, $\chi^2_{\text{Scaled difference}} (1, N = 2,624) = 1.94, p = .1636$.

In summary, although the results of both models obtained in the training set could not be fully replicated using the testing set, we did not get robust evidence for the relevance of a correlation parameter between MA and GOK. Thus, no statistical evidence was obtained for the involvement of an SL mechanism using SEM.

Figure 3

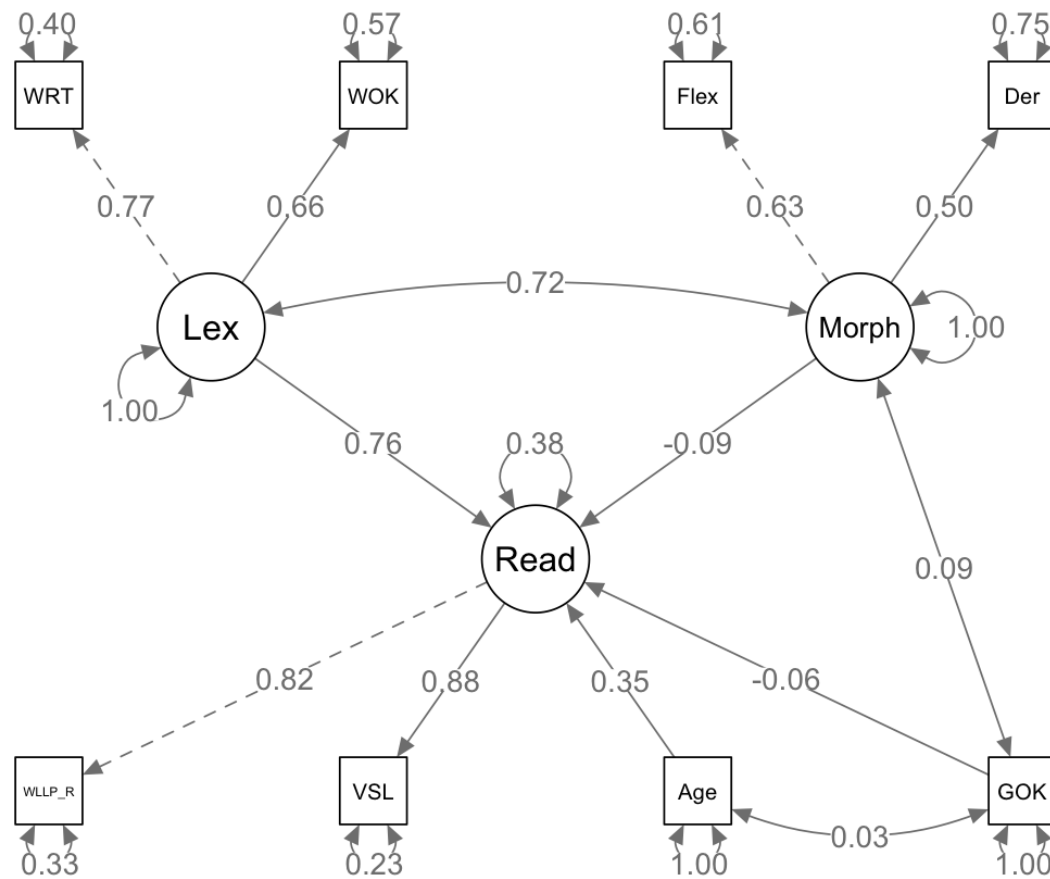
Model 1. Reading fluency predicted by linguistic skills and age, using the testing set



Note. Latent variables: *Lex* = Lexical processing, *Morph* = Morphological awareness, *Read* = Reading fluency. Observed variables: *WRT* = Spelling test, *WOK* = Word-specific orthographic knowledge, *Flex* = Morphological inflection, *Der* = Morphological derivation, *WLLP_R* = Reading fluency test 1, *VSL* = Reading fluency test 2, *GOK* = General orthographic knowledge

Figure 4

Model 2. Reading fluency predicted by linguistic skills and age, with an additional SL parameter for the correlation between MA and GOK, using the testing set



Note. Latent variables: *Lex* = Lexical processing, *Morph* = Morphological awareness, *Read* = Reading fluency. Observed variables: *WRT* = Spelling test, *WOK* = Word-specific orthographic knowledge, *Flex* = Morphological inflection, *Der* = Morphological derivation, *WLLP_R* = Reading fluency test 1, *VSL* = Reading fluency test 2, *GOK* = General orthographic knowledge

Discussion

The present study investigated the predictive role of MA (H1), lexical (H2), and sublexical (H3) orthographic knowledge and age (H4) on reading fluency performance. For this purpose, we used structural equation modeling. We also expected the value of the regression path for GOK to be lower than for WOK (H5). Complementarily, we tested the potential role of SL in the hypothesized predictive model in a novel way, including a parameter corresponding to the correlation between MA and GOK, both variables previously linked to the statistical learning of orthographic regularities (Pacton et al., 2005; Apel et al., 2019), and assessed if the fit of this model increased compared to the one without this parameter (H6).

Psychometric analysis of the morphological awareness test

First, considering that previous literature has pointed out several problems with MA measures, we examined the dimensionality of the test used to measure this construct in our sample. The confirmatory factor analyses suggest that the MA data was not unidimensional. In contrast, the three-factor model showed a better fit to the data; however, the plural formation factor showed poor psychometric performance (six out of eight items had factor loadings below .20), so we decided to eliminate it. After removing two additional items, we obtained an excellent fit to the data. However, when assessing the reliability of the two remaining scales using the McDonald omega internal consistency coefficient, values were below .50.

The results of these analyses have several implications. At the theoretical level, they suggest that the processes at the basis of performance in the different morphological tasks are not necessarily the same, even if they are highly correlated. They also indicate that the recognition of derivational changes and the identification of the appropriate inflection are more likely to be operationalized and assessed effectively than the ability to choose the correct plural form using pseudowords. At the methodological level, these findings are in line

with the low-reliability coefficients reported by previous studies (Deacon et al., 2004; Spencer et al., 2015; Desrochers et al., 2018; Görgen et al., 2021; James et al., 2021) and point to the fact that the measurement of morphological processing is often particularly affected by statistical noise. Beyond this, the results of the present research highlight the importance of analyzing the internal structure of measurements used in empirical studies before assuming their unidimensionality and making analytical decisions based on such assumptions (e.g., using total scores in statistical models). Finally, our results highlight the need to develop future measures that are internally consistent and capture individual differences in the ability to understand and manipulate the internal structure of words using different types of morphology (i.e., derivation, inflection, plural formation, and compound).

The predictive role of linguistic skills in reading fluency performance

Next, using structural equation modeling, we tested the ability of MA (H1), lexical (H2) and sublexical (H3) processing, and age (H4) to positively predict reading fluency performance. In both models assessed in the training and testing set, only lexical processing (H2) and age (H4) were statistically significant predictors of reading performance. In contrast, neither GOK nor morphological processing predicted this literacy skill.

On the one hand, these results provide evidence for the importance of word and word unit knowledge (WOK) for performance on reading fluency tests. In that sense, they align with results obtained by previous studies (Conrad et al., 2013; Deacon et al., 2012; Rothe et al., in preparation). On the other hand, the high regression coefficient obtained for the lexical processing latent variable might suggest something different. Some authors have suggested that tasks measuring lexical orthographic processing are similar to the outcomes they should predict (see Deacon et al., 2012). If that is true, the ability of WOK to predict word reading might reflect an overlap in the way the tasks employed capture performance on both constructs (i.e., a *method effect*) than an association between two distinct variables. In our

study, the lexical processing latent variable comprised two indicators reflecting the total score obtained in two tests: the first measuring word-specific orthographic knowledge (WOK) and the second spelling (WRT 3+/4+). In the first one, children had to indicate whether the stimulus presented (i.e., a pseudohomophone) was correctly spelled. In the second one, they had to type a target word previously read aloud along with a sentence frame. In the reading fluency tasks, the child had to select, from a set of options, the picture that best represented a common written word of their spelling (WLLP-R) or the correct word in a sentence frame with a blank space (VSL). Considering the different ways of measuring reading fluency and lexical processing, we can safely assume that the high association between the two variables in our data is not a methodological artifact. On the contrary, these results suggest that both skills require retrieving explicit lexical information stored in long-term memory.

The absence of significant effects for MA and GOK does not imply that these skills are irrelevant to literacy development. About MA, which reflects the understanding of the morphological structure of words and the ability to manipulate this structure actively (Carlisle, 1995), one possibility is that the procedural nature of this type of linguistic processing is relatively independent of reading fluency as operationalized in this research. In the present study, reading fluency might reflect an explicit knowledge of whole words rather than the ability to manipulate their internal structure dynamically. This is supported by the high relationship between lexical processing and reading fluency. Additionally, the high association between MA and lexical processing suggests that the tasks that measured both constructs, and the constructs themselves, also have common aspects. However, MA would not have been able to compete with lexical processing as a relevant predictor of reading fluency, most likely due to its sensitivity to statistical noise, expressed in the low-reliability levels of its two subscales.

It is also possible that the more transparent character of German orthography (compared, for example, to English) may have played a role in the absence of significant effects for MA. If this is so, it would be feasible to assume that MA would have played a more dominant role if the criterion variable had been reading accuracy and not reading fluency, also considering that both are relatively independent in orthographies of higher consistency. For example, Görgen et al. (2021), using the same dataset as us, obtained significant results for MA using hierarchical regression models on the same data set used in the present study. However, three aspects differentiate that study from ours: first, they assumed the unidimensional character of MA, whereas in our case, we hypothesized (and tested) a multidimensional composition and removed one factor (e.g., plural formation); second, Görgen et al. (2021) did not include WOK in their models; and third, they included phonological awareness and non-verbal intelligence as predictors. Their results might have differed if they included lexical processing as a predictor. Finally, our results align with those of Haase and Steinbrink (2022), who also obtained no significant effects for MA.

Finally, the null effect of MA on reading fluency when controlling for lexical processing might suggest a mediation effect between these three variables. For example, considering this result, it would be feasible to hypothesize that the relationship between MA and reading fluency requires WOK as a facilitating variable. Thus, in order to manipulate the structure of an unknown stimulus (i.e., a pseudoword), an effective strategy to accomplish this task would be to activate similar real word entries in the orthographic lexicon. In this way, drawing on specific word knowledge would serve as a bridge between MA and reading performance. Although this is beyond the scope of the present research, it will be interesting for future research to test such a causal mechanism.

Concerning GOK, a skill that reflects implicit knowledge about the frequency and legality of letter combinations within words, this variable not only showed a lower regression

coefficient than lexical-level processing (in line with H5), but its contribution to reading fluency was statistically non-significant. This result coincides with previous evidence that the association between WOK and literacy skills is higher than that shown by GOK (Georgiou et al., 2008; Rothe et al., 2014), primarily when pseudohomophones and illegal pseudowords are used to measure both types of orthographic processing, respectively (Rothe et al., in preparation). This makes sense if one considers that, while lexical orthographic representations are indispensable both for automatic lexical access during reading and for the production of correctly spelled words during spelling, the impact of GOK on literacy would require the intervention of other intermediate cognitive processes (for a review of the Noisy Chain Hypothesis, see Schmalz et al., 2021). In any case, it is relevant that neither of the two skills that, in theory, would be influenced by the implicit learning of orthographic regularities (i.e., MA and GOK) has significantly predicted reading fluency.

No evidence for the involvement of a statistical learning mechanism

Finally, we tested the role of SL in the tested model. To do so, we included an additional parameter in the model that reflected the correlation between MA and GOK, variables susceptible to being influenced by statistical learning. We also hypothesized that if SL is relevant for both skills, a model with such a parameter should better fit the data than a model without it (H6). In contrast to what we hypothesized, the results obtained using SEM models showed that the model including the additional correlation parameter did not fit the data better in either of the two data sets used. In addition, the correlation between MA and GOK was below .10 in both datasets, and neither significantly predicted reading fluency.

The absence of a significant relationship between MA and GOK, as well as the lack of improvement in model fit following the inclusion of such a correlation, contrasts with previous literature according to which the acquisition and development of both skills would be explained by an SL mechanism (Pacton et al., 2005; Deacon et al., 2008; Deacon &

Leung, 2013). Moreover, neither of the two variables most closely linked to SL in our model (i.e., MA and GOK) relevantly predicted fluency, which departs from results reported in studies linking SL to reading performance (Arciuli & Simpson, 2012; Frost et al., 2013). In any case, our results align with more critical literature on the involvement of SL in developing literacy skills (Schmalz et al., 2019; Schmalz et al., 2021).

Our results around SL can be interpreted in more than one way. The first line of interpretation is theoretical and suggests that SL would serve a limited role as a direct predictor of reading performance in the absence of other intermediate processes, in line with the “noisy chain hypothesis” (see Schmalz et al., 2021). However, also from this perspective, one would expect a more significant role for SL after introducing intermediate processes (i.e., MA and GOK) linked, on the one hand, to SL and, on the other hand, to reading fluency into the model. This was not the case. One possibility is that the relationship of both processes to reading fluency is epiphenomenal (see Schmalz et al., 2021) and not causal. For example, MA and GOK might play a secondary role in fluency performance once the reader has mastered other skills. From this standpoint, the reader would draw less and less on their morphological and sub-lexical knowledge (i.e., the statistical aspects of their spelling) when reading fluently and rely more on these other skills. One potential candidate is lexical processing. The results of our model point in this direction: lexical processing was the most robust predictor of fluency, while neither MA nor GOK was significantly associated with this literacy skill. Moreover, the relationship between MA and GOK did not provide any improvement to the model tested.

The second line of interpretation is methodological: it is possible that including a correlation parameter between MA and GOK did not fully capture the role of an SL mechanism. Closely linked to this line of interpretation is the impossibility of attributing any potential correlation between the two constructs to a particular third variable, such as SL

Indeed, not having measured FS directly, it is challenging to establish the intervention of this mechanism in the model. This is the main limitation of our study. However, we consider MA and GOK to be two apt candidates to serve as proxy variables of such a mechanism (Pacton et al., 2005; Deacon et al., 2008; Deacon & Leung, 2013), even more so in the absence of valid and reliable measurements of SL (Siegelman, 2017). In any case, beyond the methodological limitations pointed out, the results of the present study provide solid evidence that, in the German language, and after controlling for the effect of lexical processing, neither MA nor GOK (both variables reflecting statistical regularities of written language) is a relevant predictor of reading fluency. Finally, the relationship between the two variables is small, and their statistical significance is due to chance.

Conclusion, limitations, and further research

Here, we assessed the extent to which various language skills (i.e., MA, lexical processing, and GOK) and age predict reading fluency. In addition, we tested the role of a domain-general mechanism (i.e., statistical learning) in the relationship between two of these skills: MA and GOK. First, we found that only lexical processing and age significantly predicted fluency. Also, lexical processing was a more robust predictor than GOK. None of the SL-related skills (i.e., MA and GOK) significantly predicted fluency, and the model that included a correlation between MA and GOK did not show a better fit to the data, suggesting that SL does not play a relevant role in the predictive model of fluency. In addition, we obtained psychometric evidence that the task used to measure MA has a multidimensional structure. At the methodological level, the study's main limitations concern the MA measurement's low reliability and the difficulty of ruling out the involvement of an SL mechanism because we did not measure this construct directly. However, we consider that how we tested the role of the SL mechanism constitutes a valid and novel alternative. In the future, it will be essential to construct instruments that validly and reliably measure both MA

and SL. Despite these limitations, the present research emphasizes the importance of word-specific orthographic knowledge in the process of fluent reading.

References

- Apel, K., Henbest, V. S., & Masterson, J. (2019). Orthographic knowledge: Clarifications, challenges, and future directions. *Reading and Writing*, 32(4), 873–889.
<https://doi.org/10.1007/s11145-018-9895-9>
- Arciuli, J., & Simpson, I. C. (2012). Statistical Learning Is Related to Reading Ability in Children and Adults: Cognitive Science. *Cognitive Science*, 36(2), 286–304.
<https://doi.org/10.1111/j.1551-6709.2011.01200.x>
- Barker, T. A., Torgesen, J. K., & Wagner, R. K. (1992). The Role of Orthographic Processing Skills on Five Different Reading Tasks. *Reading Research Quarterly*, 27(4), 334.
<https://doi.org/10.2307/747673>
- Birkel, P. (2007a). WRT 3: *Weingartener grundwortschatz, rechtschreib-test für dritte und vierte Klassen* [‘Weingarten spelling test for basic vocabulary’]. Hogrefe.
- Birkel, P. (2007b). WRT 4+: *Weingartener grundwortschatz rechtschreib-test für vierte und fünfte Klassen* [‘Weingarten spelling test for basic vocabulary’]. Hogrefe.
- Bekebrede, J., van der Leij, A., & Share, D. L. (2009). Dutch dyslexic adolescents: Phonological-core variable-orthographic differences. *Reading and Writing*, 22(2), 133–165.
<https://doi.org/10.1007/s11145-007-9105-7>
- Bratlie, S. S., Brinchmann, E. I., Melby-Lervåg, M., & Torkildsen, J. von K. (2022). Morphology—A Gateway to Advanced Language: Meta-Analysis of Morphological Knowledge in Language-Minority Children. *Review of Educational Research*, 92(4), 614–650. <https://doi.org/10.3102/00346543211073186>
- Byrne, B. (1996). The learnability of the alphabetic principle: Children’s initial hypotheses about how print represents spoken language. *Applied Psycholinguistics*, 17(4), 401–426.
<https://doi.org/10.1017/S0142716400008171>

- Carlisle, J.F. (1995). Morphological awareness and early reading achievement. In L.B. Feldman (Ed.), *Morphological aspects of language processing*, (pp. 189–209). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.
- Carlisle, J. F., McBride-Chang, C., Nagy, W., & Nunes, T. (2010). Effects of Instruction in Morphological Awareness on Literacy Achievement: An Integrative Review. *Reading Research Quarterly*, 45(4), 464–487. <https://doi.org/10.1598/RRQ.45.4.5>
- Carlisle, J. F., & Stone, C. A. (2005). Exploring the role of morphemes in word reading. *Reading Research Quarterly*, 40(4), 428–449. <https://doi.org/10.1598/RRQ.40.4.3>
- Cassar, M., & Treiman, R. (1997). The beginnings of orthographic knowledge: Children's knowledge of double letters in words. *Journal of Educational Psychology*, 89(4), 631–644. <https://doi.org/10.1037/0022-0663.89.4.631>
- 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>
- 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 Ltd. <https://doi.org/10.1002/9780470757642.ch1>
- Conrad, N. J., Harris, N., & Williams, J. (2013). Individual differences in children's literacy development: The contribution of orthographic knowledge. *Reading and Writing*, 26(8), 1223–1239. <https://doi.org/10.1007/s11145-012-9415-2>
- Cunningham, A. E., & Stanovich, K. E. (1993). Children's literacy environments and early word recognition subskills. *Reading and Writing*, 5(2), 193–204. <https://doi.org/10.1007/BF01027484>

- Deacon, S. H., Benere, J., & Castles, A. (2012). Chicken or egg? Untangling the relationship between orthographic processing skill and reading accuracy. *Cognition*, 122(1), 110–117. <https://doi.org/10.1016/j.cognition.2011.09.003>
- Deacon, S. H., & Bryant, P. (2005). The strength of children's knowledge of the role of root morphemes in the spelling of derived words. *Journal of Child Language*, 32(2), 375–389. <https://doi.org/10.1017/S0305000904006816>
- Deacon, S. H., Conrad, N., & Pacton, S. (2008). A statistical learning perspective on children's learning about graphotactic and morphological regularities in spelling. *Canadian Psychology / Psychologie Canadienne*, 49(2), 118–124. <https://doi.org/10.1037/0708-5591.49.2.118>
- Deacon, S. H., & Kirby, J. R. (2004). Morphological awareness: Just “more phonological”? The roles of morphological and phonological awareness in reading development. *Applied Psycholinguistics*, 25(2), 223–238. <https://doi.org/10.1017/S0142716404001110>
- Deacon, S. H., & Leung, D. (2013). Testing the statistical learning of spelling patterns by manipulating semantic and orthographic frequency. *Applied Psycholinguistics*, 34(6), 1093–1108. <https://doi.org/10.1017/S0142716412000173>
- Desrochers, A., Manolitsis, G., Gaudreau, P., & Georgiou, G. (2018). Early contribution of morphological awareness to literacy skills across languages varying in orthographic consistency. *Reading and Writing*, 31(8), 1695–1719. <https://doi.org/10.1007/s11145-017-9772-y>
- Diamanti, V., Mouzaki, A., Ralli, A., Antoniou, F., Papaioannou, S., & Protopapas, A. (2017). Preschool Phonological and Morphological Awareness As Longitudinal Predictors of Early Reading and Spelling Development in Greek. *Frontiers in Psychology*, 8, 2039. <https://doi.org/10.3389/fpsyg.2017.02039>
- Epskamp, S. (2022). *semPlot: Path Diagrams and Visual Analysis of Various SEM*

Packages. R package version 1.1.5. Retrieved from <https://CRAN.R-project.org/package=semPlot>.

- Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences*, 19(3), 117–125. <https://doi.org/10.1016/j.tics.2014.12.010>
- Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What Predicts Successful Literacy Acquisition in a Second Language? *Psychological Science*, 24(7), 1243–1252. <https://doi.org/10.1177/0956797612472207>
- Georgiou, G. K., Parrila, R., & Papadopoulos, T. C. (2008). Predictors of word decoding and reading fluency across languages varying in orthographic consistency. *Journal of Educational Psychology*, 100(3), 566–580. <https://doi.org/10.1037/0022-0663.100.3.566>
- Görge, R., De Simone, E., Schulte-Körne, G., & Moll, K. (2021). Predictors of reading and spelling skills in German: The role of morphological awareness. *Journal of Research in Reading*, 44(1), 210–227. <https://doi.org/10.1111/1467-9817.12343>
- Haase, A., & Steinbrink, C. (2022). Associations between morphological awareness and literacy skills in German primary school children: The roles of grade level, phonological processing and vocabulary. *Reading and Writing*, 35(7), 1675–1709. <https://doi.org/10.1007/s11145-021-10247-1>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- 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>

- James, E., Currie, N. K., Tong, S. X., & Cain, K. (2021). The relations between morphological awareness and reading comprehension in beginner readers to young adolescents. *Journal of Research in Reading*, 44(1), 110–130. <https://doi.org/10.1111/1467-9817.12316>
- Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2022). *semTools: Useful tools for structural equation modeling*. R package version 0.5-6. Retrieved from <https://CRAN.R-project.org/package=semTools>
- Kemp, N., & Bryant, P. (2003). Do Beez Buzz? Rule-Based and Frequency-Based Knowledge in Learning to Spell Plural -s. *Child Development*, 74(1), 63–74. <https://doi.org/10.1111/1467-8624.00521>
- König, E., & Gast, V. (2012). *Understanding English-German contrasts* (3rd ed.). *Grundlagen der Anglistik und Amerikanistik: Vol. 29*. Berlin: E. Schmidt.
- Kuhn, M. R., Schwanenflugel, P. J., Meisinger, E. B., Levy, B. A., & Rasinski, T. V. (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](https://doi.org/10.1016/0010-0285(74)90015-2)
- Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To Parcel or Not to Parcel: Exploring the Question, Weighing the Merits. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 151–173. https://doi.org/10.1207/S15328007SEM0902_1
- Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In Search of Golden Rules: Comment on Hypothesis-Testing Approaches to Setting Cutoff Values for Fit Indexes and Dangers in Overgeneralizing Hu and Bentler's (1999) Findings. *Structural Equation Modeling: A Multidisciplinary Journal*, 11(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2

Matsunaga, M. (2008). Item Parceling in Structural Equation Modeling: A Primer.

Communication Methods and Measures, 2(4), 260–293.

<https://doi.org/10.1080/19312450802458935>

McDonald, R. P. (2013). *Test Theory: A Unified Treatment* (0 ed.). Psychology Press.

<https://doi.org/10.4324/9781410601087>

Mousikou, P., Beyersmann, E., Ktori, M., Javourey-Drevet, L., Crepaldi, D., Ziegler, J. C.,

Grainger, J., & Schroeder, S. (2020). Orthographic consistency influences morphological processing in reading aloud: Evidence from a cross-linguistic study. *Developmental Science*, 23(6). <https://doi.org/10.1111/desc.12952>

Nagy, W., Berninger, V., Abbott, R., Vaughan, K., & Vermeulen, K. (2003). Relationship of Morphology and Other Language Skills to Literacy Skills in At-Risk Second-Grade Readers and At-Risk Fourth-Grade Writers. *Journal of Educational Psychology*, 95(4), 730–742.

<https://doi.org/10.1037/0022-0663.95.4.730>

National Institute of Child Health and Human Development. (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). Washington, DC: US Government Printing Office.

National Reading Panel (US), National Institute of Child Health, Human Development (US), National Reading Excellence Initiative, National Institute for Literacy (US), & United States Department of Health. (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: Reports of the subgroups*. National Institute of Child Health and Human Development, National Institutes of Health.

Olson, R., Forsberg, H., Wise, B., & Rack, J. (1994). Measurement of word recognition, orthographic, and phonological skills. In *Frames of reference for the assessment of learning*

disabilities: New views on measurement issues. (pp. 243–277). Paul H Brookes Publishing Co.

Pacton, S., Fayol, M., & Perruchet, P. (2005). Children's implicit learning of graphotactic and morphological regularities. *Child development*, 76(2), 324-339.

Perfetti, C. A. (1984). Reading Acquisition and beyond: Decoding Includes Cognition. *American Journal of Education*, 93(1), 40–60. <https://doi.org/10.1086/443785>

R Core Team. (2022). *R: A language and environment for statistical computing*. R

Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>

Revelle, W. (2022). psych: Procedures for Psychological, Psychometric, and Personality Research. Northwestern University, Evanston, Illinois. R package version 2.2.5, <https://CRAN.R-project.org/package=psych>.

Roman, A. A., Kirby, J. R., Parrila, R. K., Wade-Woolley, L., & Deacon, S. H. (2009). Toward a comprehensive view of the skills involved in word reading in Grades 4, 6, and 8. *Journal of Experimental Child Psychology*, 102(1), 96–113. <https://doi.org/10.1016/j.jecp.2008.01.004>

Ropovik, I. (2015). A cautionary note on testing latent variable models. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.01715>

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2). <https://doi.org/10.18637/jss.v048.i02>

Rothe, J., Darcourt, A., Moll, K., Schulte-Körne, G., & Schmalz, X. (In preparation). Mediation in the relation of orthographic processing on the lexical and sublexical level with reading and spelling skills. A large cross-sectional study in elementary school children in Germany. *Department of Child and Adolescent Psychiatry and Psychotherapy, Faculty of Medicine, LMU*.

- Rothe, J., Schulte-Körne, G., & Ise, E. (2014). Does sensitivity to orthographic regularities influence reading and spelling acquisition? A 1-year prospective study. *Reading and Writing*, 27(7), 1141–1161. <https://doi.org/10.1007/s11145-013-9479-7>
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-Month-Old Infants. *Science*, 274(5294), 1926–1928. <https://doi.org/10.1126/science.274.5294.1926>
- Satorra, A. (2000). Scaled and Adjusted Restricted Tests in Multi-Sample Analysis of Moment Structures. In R. D. H. Heijmans, D. S. G. Pollock, & A. Satorra (Eds.), *Innovations in Multivariate Statistical Analysis* (Vol. 36, pp. 233–247). Springer US. https://doi.org/10.1007/978-1-4615-4603-0_17
- Schmalz, X., Moll, K., Mulatti, C., & Schulte-Körne, G. (2019). Is Statistical Learning Ability Related to Reading Ability, and If So, Why? *Scientific Studies of Reading*, 23(1), 64–76. <https://doi.org/10.1080/10888438.2018.1482304>
- Schmalz, X., Treccani, B., & Mulatti, C. (2021). Developmental Dyslexia, Reading Acquisition, and Statistical Learning: A Sceptic’s Guide. *Brain Sciences*, 11(9), 1143. <https://doi.org/10.3390/brainsci11091143>
- Schneider, W., Blanke, I., Faust, V., & Küspert, P. (2011). WLLP-R Würzburger leise leseprobe [Würzburger Silent Reading Test - Revised]. *Göttingen: Hogrefe*.
- 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., & collaboration with COST Action A8 network. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94(2), 143–174. <https://doi.org/10.1348/000712603321661859>

- Share, D. L. (2008). Orthographic Learning, Phonological Recoding, and Self-Teaching. In *Advances in Child Development and Behavior* (Vol. 36, pp. 31–82). Elsevier.
[https://doi.org/10.1016/S0065-2407\(08\)00002-5](https://doi.org/10.1016/S0065-2407(08)00002-5)
- Siegelman, N., Bogaerts, L., & Frost, R. (2017). Measuring individual differences in statistical learning: Current pitfalls and possible solutions. *Behavior Research Methods*, 49(2), 418–432. <https://doi.org/10.3758/s13428-016-0719-z>
- Spencer, M., Muse, A., Wagner, R. K., Foorman, B., Petscher, Y., Schatschneider, C., Tighe, E. L., & Bishop, M. D. (2015). Examining the underlying dimensions of morphological awareness and vocabulary knowledge. *Reading and Writing*, 28(7), 959–988.
<https://doi.org/10.1007/s11145-015-9557-0>
- Stahl, S. A., & Murray, B. A. (1994). Defining phonological awareness and its relationship to early reading. *Journal of Educational Psychology*, 86(2), 221–234.
<https://doi.org/10.1037/0022-0663.86.2.221>
- Thompson, G. B., Fletcher-Flinn, C. M., & Cottrell, D. S. (1999). Learning correspondences between letters and phonemes without explicit instruction. *Applied Psycholinguistics*, 20(1), 21–50. <https://doi.org/10.1017/S0142716499001022>
- Tighe, E. L., & Binder, K. S. (2015). An investigation of morphological awareness and processing in adults with low literacy. *Applied Psycholinguistics*, 36(2), 245–273.
<https://doi.org/10.1017/S0142716413000222>
- Treiman, R. (1993). *Beginning to spell: A study of first-grade children*. (pp. xiii, 365). Oxford University Press.
- Treiman, R., Cassar, M., & Zukowski, A. (1994). What Types of Linguistic Information Do Children Use in Spelling? The Case of Flaps. *Child Development*, 65(5), 1318.
<https://doi.org/10.2307/1131501>

- Visser, L., Kalmar, J., Linkersdörfer, J., Görgen, R., Rothe, J., Hasselhorn, M., & Schulte-Körne, G. (2020). Comorbidities Between Specific Learning Disorders and Psychopathology in Elementary School Children in Germany. *Frontiers in Psychiatry, 11*, 292. <https://doi.org/10.3389/fpsy.2020.00292>
- Walter, J. (2013). Lernfortschrittsdiagnostik Lesen (LDL) und Verlaufsdiagnostik sinnerfassenden Lesens [Diagnosis of progress in reading comprehension] (VSL): Zwei Verfahren als Instrumente einer formativ orientierten Lesediagnostik. In *Lernverlaufsdiagnostik* (S. 165–201).
- Weiß, R. H., & Weiß, B. (2008). *CFT 20-R mit ws/zf-r: Grundintelligenztest skala 2-revision (cft 20-r) mit wortschatztest und zahlenfolgentest-revision (ws/zf-r)*. Hogrefe.
- Zarić, J., Hasselhorn, M., & Nagler, T. (2021). Orthographic knowledge predicts reading and spelling skills over and above general intelligence and phonological awareness. *European Journal of Psychology of Education, 36*(1), 21–43. <https://doi.org/10.1007/s10212-020-00464-7>

Academic Integrity Statement

I, Alvaro Darcourt, hereby confirm that I completed this master's thesis with the title: “Does knowledge of different orthographic regularities rely on a domain-general statistical learning mechanism? An SEM analysis of a large-scale database” independently, that I have not heretofore presented this paper to another department or university, and I have listed all references used, and have given credit to all additional sources of assistance

Munich,

Alvaro Darcourt

Einverständniserklärung zur Einsichtnahme in die Masterarbeit/

Declaration of Consent: Access to Master Thesis

Bitte die entsprechende Alternative ankreuzen./ *Please tick off the appropriate alternative.*

- ☐ **Ich bin damit einverstanden**, dass für den Fall des Bestehens der Masterarbeit meine Masterarbeit von künftigen Studenten eingesehen werden darf./
I agree that in case of passing the master thesis, future students may access my master thesis.

- ☐ **Ich bin nicht damit einverstanden**, dass für den Fall des Bestehens der Masterarbeit meine Masterarbeit von künftigen Studenten eingesehen werden darf./
I do not agree that in case of passing the master thesis, future students may access my master thesis.

Full Name: ALVARO DAR COURT

Date

Signature