Effectiveness of spelling treatment approaches for learners with dyslexia

– A meta-analysis and systematic review

Authors

Katharina Galuschka^{1*}, Ruth Görgen^{1*}, Julia Kalmar², Stefan Haberstroh¹, Xenia Schmalz¹ and Gerd Schulte-Körne¹

*Joint first authorship

Affiliations

¹ Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital, LMU Munich, Nußbaumstraße 5a, 80336 Munich, Germany

² Center for Psychological Psychotherapy, University of Heidelberg, Bergheimer Straße 58a, 69115 Heidelberg, Germany

Corresponding author

Katharina Galuschka

Phone: 0049 89 4400 56936

Fax: 0049 89 8905 – 6978

ACCEPTED MANUSCRIPT

TO APPEAR IN EDUCATIONAL PSYCHOLOGIST

E-mail addresses

katharina.galuschka@med.uni-muenchen.de

ruth.goergen@med.uni-muenchen.de

julia.kalmar@zpp.uni-hd.de

stefan.haberstroh@med.uni-muenchen.de

xenia.schmalz@med.uni-muenchen.de

gerd.schulte-koerne@med.uni-muenchen.de

Abstract

This systematic review and meta-analysis investigated the efficacy of spelling interventions for the remediation of dyslexia and spelling deficits. Theoretically important moderators, such as the treatment approach as well as orthographic and sample characteristics, were also considered. Thirty-four controlled trials that evaluated spelling interventions in children, adolescents and adults with dyslexia and spelling deficits were included. Results show that treatment approaches using phonics, orthographic (graphotactic or orthographic phonological spelling rules), as well as morphological instruction had a moderate to high impact on spelling performance. A significant influence of interventions that teach memorization strategies to improve spelling could not be confirmed. This work shows that understanding the principles of an orthography is beneficial for learners with dyslexia or spelling deficits and presents key components for effective spelling intervention.

Keywords: Meta-analysis, Dyslexia, Spelling, Writing, Learning disabilities

Learning to spell is a challenging task. In most alphabetic orthographies, spelling skills are more difficult to acquire than reading skills (Bosman & Van Orden, 1997). For a substantial number of children, learning to spell remains challenging across development (Maughan et al., 2009). Difficulties with spelling (e.g., adding, omitting, or substituting letters or graphemes in written words) are a common indicator of specific learning disorders and a core characteristic of dyslexia. The term dyslexia is used to refer to a pattern of learning difficulties characterized by problems with word recognition, decoding and spelling. By definition (Lyon, Shaywitz & Shaywitz, 2003), dyslexia typically results from an underlying phonological deficit that is not due to cognitive impairment or ineffective classroom instruction. As the term dyslexia implies, spelling deficits often co-occur with reading problems.

Despite the frequent co-occurrence, spelling and reading difficulties can also be observed in isolation (Moll, Kunze, Neuhoﬀ, Bruder, & Schulte-Körne, 2014). Frith (1980) proposed that this dissociation arises because children rely on a partial cue reading strategy. Mental representations of the word's orthographic form (orthographic representations), its link to phonology and semantics, are required for reading and spelling. For reading, relying on a partial cue, or weakly specified (incomplete) orthographic representations, is sufficient to quickly recognize a printed word. For spelling, high-quality orthographic representations are necessary, because the readers need detailed information about all the letters within this word (Perfetti & Hart, 2002). Moll and Landerl (2009) hypothesized that problems in building up orthographic representations might encourage some learners with dyslexia to develop highly efficient decoding strategies as a compensatory mechanism. In those children, reading is in the normal range while spelling remains below average.

This study extends the current knowledge of effective spelling treatment approaches for learners with dyslexia by giving a comprehensive, up-to-date overview of spelling interventions. To advance towards a theory of spelling that describes how spelling performance of learners with dyslexia and spelling deficits can be improved, we quantify the effectiveness of different spelling treatment approaches. We also consider the impact of age and severity of reading and spelling disorder on the effectiveness of spelling treatment approaches, to assess whether phonics intervention should precede morphological or orthographic intervention. As a further theoretical contribution, we aim to capture the influence of orthographic consistency on spelling interventions and examine the impact of sample and intervention characteristics on the efficacy of spelling interventions.

Theoretical Perspectives on the Relation between Spelling and Text Production

Spelling is very important for a person's academic and professional achievement. Severe spelling errors can make a text difficult to read and can cause readers to devalue the quality of a written composition (Graham, Harris, & Hebert, 2011; Graham & Hebert, 2011). The consequence of poor spelling skills extends beyond the readability of a text and the appearance of academic incompetence. Spelling requires considerable cognitive effort for poor spellers, so that fewer resources are left for other cognitive processes that are important for writing (e.g., content development). It is hypothesized that spelling deficits (beside handwriting skills) induce cognitive overload and influence text production fluency and quality (Berninger & Swanson, 1994; McCutchen, 1996, 2000).

Continued overload of cognitive resources during spelling may weaken the subsequent development of writing skills (Graham, Berninger, Abbott, Abbott, & Whitaker, 1997; Kellogg, 2008). Therefore, poor spelling is an obstacle for text production in the same way that difficulties in word recognition are a bottleneck for reading comprehension (Perfetti & Hogaboam, 1975). Previous research about the association between spelling and text production has shown that spelling instruction improves compositional fluency (Graham, Harris, & Adkins, 2018; Graham, Harris, & Chorzempa, 2002). However, it is unclear whether spelling ability also has an influence on composition quality (Graham et al., 2018; Graham & Santangelo, 2014). This would support the theoretical view that effective spelling instruction enhances spelling performance which, in turn, frees cognitive resources and paves the way for the acquisition of advanced text production skills. The improvement of compositional quality, however, seems to depend on explicit instruction in composition.

Theoretical Perspectives on Learning to Spell and its Influence on Reading

Spelling not only affects writing and compositional skills, but is also closely linked to reading (Ehri, 1987; Frith, 1985). Most of the linguistic abilities that are important for reading acquisition (e.g., phonological and phoneme awareness, explicit orthographic or morphological knowledge) are also important for spelling ability. In some phases spelling is essential for reading, because it trains phoneme awareness and the alphabetic principle, whereas in other phases, reading boosts spelling, because the reading exposure leads to a richer lexicon of orthographic representations.

The development of spelling and reading skills starts long before children learn to read and write in school. First, preschool children are able to differentiate between writing and drawing; for example, they learn to recall the meaning of some street signs and graphic signals. Later, children acquire phonological awareness skills and become increasingly better

in detecting or manipulating syllables and phonemes in spoken words as well as in segmenting spoken words into syllables and phonemes (Treiman & Bourassa, 2000). Phonological awareness skills and the ability to segment spoken words into phonemes is an important precursor of reading and spelling acquisition (Moll, Ramus, et al., 2014). Thus, it is not surprising that *phonological awareness interventions* (including oral tasks for recognizing phonemes within words, blending phonemes into words, segmenting a word into phonemes, eliminating a phoneme from a word or adding a phoneme to a word) are often implemented to foster spelling and reading skills. It is often assumed that *phonological awareness interventions* in preschool years are beneficial for early reading and spelling development. At later stages, a learner has already gained insight into the phonemic structure, and those interventions may lose their effectiveness in improving reading and spelling skills (Bus & van Ijzendoorn, 1999; Galuschka, Ise, Krick, & Schulte-Körne, 2014).

According to Frith (1985), instruction in spelling and writing trains the ability to segment spoken words into phonemes, and enables the insight into the alphabetic principle: the understanding that graphemes represent phonemes. This insight is then transferred from spelling to reading and allows children to read words they have not encountered before (Frith, 1985). Share (1995) argued that knowledge of phoneme-grapheme and grapheme-phoneme correspondences make up a self-teaching mechanism, which supports the acquisition of orthographic skills and fosters sight word reading (direct retrieval of phonology and semantics associated with a given written word form from the orthographic lexicon). As reading becomes more automatized, fewer cognitive resources are required for decoding graphemes into phonemes. Thus, these resources can be allocated to reading comprehension (Ehri, 2005, 2014).

Individuals demonstrate *phoneme-grapheme knowledge* when they understand the simple mappings between phonemes and graphemes. Due to the strong influence of phoneme-grapheme and grapheme-phoneme correspondence skills on reading and spelling ability, research on reading and spelling instruction and intervention has predominantly focused on the acquisition of this fundamental knowledge. Such treatment approaches are often referred to as *phonics interventions*. A large body of research supports the efficacy of those interventions on reading and spelling performance (Ehri, Nunes, Stahl, & Willows, 2001; Galuschka et al., 2014; McArthur et al., 2012; Weiser & Mathes, 2011). At the same time, representing every phoneme with a reasonable letter is not sufficient for skilled spelling, as can be shown by incorrect spellings such as “wodr” for “water” or the German equivalent “Wasa” for “Wasser”. In many orthographies, one sound can have different spellings in

different words (e.g., the phoneme /i/ in English may be spelled <ee> as in “deep” or <ea> as in “mean”; in German, /f/ may be spelled <sch> as in “Schule” (school) or <s> as in “Stuhl” (chair)). According to Treiman and Kessler (2014), words that cannot be spelled correctly by simple phoneme-grapheme correspondences may follow other patterns: (1) graphotactic (e.g., consonant doubling is less common before /ik/ when it is spelled as <ic> (e.g., “magic”) than when it is spelled as <ick> (e.g., “gimmick”)) (2) phonological-orthographic (e.g., consonant doublets do not occur after long vowels in Dutch, English and German, as in “late” [long a] versus “latte” [short a]) or (3) morphological (e.g., in German vowel change (umlaut) for noun plurals are dependent on the singular “Baum” (tree) to “Bäume” (trees): an understanding of the morphological relationship will allow children to spell the plural form correctly, as opposed to a phonologically plausible but incorrect spelling such as “Beume”). To become a skilled speller, the learner has to increasingly consider those patterns or regularities (Ehri, 2000; Frith, 1985; Treiman, 2018). This is true for all learners, regardless of whether they have a diagnosis of dyslexia or not.

Interventions aiming to help children with dyslexia and spelling deficits deal with the deviations from one-to-one mappings between phonemes and graphemes, and provide explanations for these deviations by explicitly teaching morphological or orthographic knowledge. *Morphological knowledge* refers to the awareness of the smallest meaningful language units (Carlisle, 2003; Nagy, Berninger, Abbott, Vaughan, & Vermeulen, 2003). A large body of research supports the efficacy of morphological instruction (Carlisle, 2003, 2010; Goodwin & Ahn, 2010, 2013; Nagy et al., 2003). *Morphological interventions* for younger children often include practicing of inflections, while morphological interventions for older children often focus on derivations (Kirby et al., 2012). *Orthographic knowledge* refers to the understanding of the orthographic rule system that allows correct writing in terms of rules and patterns of written language. Interventions that foster orthographic skills mainly focus on graphotactic and phonological-orthographic regularities. Children have to learn the constraints and orthographic rules that govern the positions, combinations and resulting pronunciations of letters within words. Squires and Wolter (2016) synthesized intervention studies that evaluated *orthographic interventions* and confirmed their efficacy on spelling skills.

According to the theory of Ehri (2000, 2014), reading supports spelling through better retention of written words in memory. However, the work of Ouellette, Martin-Chang, and Rossi (2017) shows that high-quality orthographic representations for spelling further support sight word reading. These results highlight, once again, how reading and spelling influence

each other. The ideal process to build-up complete orthographic structures is therefore of particular profound importance and has been part of deep theoretical discussions.

Theoretical Discussion: Phonics First or Morphology

A common view of spelling acquisition is that, at the beginning, instruction should focus on phoneme-grapheme correspondences, and orthographic and morphological knowledge should be taught after the basic correspondences are acquired (Treiman & Kessler, 2014). This view is supported by theoretical models of literacy acquisition (Castles, Rastle, & Nation, 2018; Ehri, 2000; Frith, 1985; Jackson & Coltheart, 2013; Sprenger-Charolles, Siegel, & Bonnet, 1998) and a large number of empirical findings (e.g., Castles et al., 2018; Moll, Ramus, et al., 2014; Taylor, Davis, & Rastle, 2017). These relatively simple phoneme-grapheme correspondences are sufficient to decode most of the words which beginning readers encounter in age-appropriate reading materials, and provide a potent self-teaching mechanism to get at least a partial phonological representation of unfamiliar words (Rastle & Taylor, 2018). This view implies that *morphological* or *orthographic interventions* would be less effective for young children in the early grades or for more severely impaired dyslexics, than for older learners in the upper grade levels or less severely impaired dyslexics.

In contrast, Bowers and Bowers (2017) postulate that phonology, morphology and etymology should be taught jointly from the beginning of literacy instruction. They argue that the alphabetic principle does not provide a reasonable description of the words that children are confronted with first, even in age-appropriate materials (Bowers & Bowers, 2018b). The English orthography becomes more transparent when the origin and morphological structure of words are considered. For example, the spelling of the word “nation” is opaque, in the sense that there are different phonologically plausible ways to spell it (e.g., “nayshun”). The spelling, however, becomes more transparent once we consider the Latin origin and the morpheme *-ion*. Thus, Bowers and Bowers (2017) propose that children should learn the logic of the written language system in a meaningful context, and disagree with the assumption that children are not yet able to deal with such complex aspects at the beginning of written language acquisition. This proposal is controversial and the question of the right starting point of phonics, orthographic or morphological instruction resulted in an ongoing theoretical debate in reading research (Bowers & Bowers, 2018a, 2018b; Rastle, 2018; Rastle & Taylor, 2018).

To examine this theoretical debate, and to approach a theory of spelling that accurately describes how learners with dyslexia and spelling deficits should be taught spelling, we aim to estimate the influence of age or grade level, as well as spelling

performance on the effectiveness of phonics and morphological interventions. Irrespective of the great importance that graphemes and phonemes undoubtedly have in alphabetic orthographies, the potential of morphological interventions should not be underestimated especially for learners with reading and spelling deficits. Morphological awareness has been shown to relate to reading and spelling (McCutchen & Logan, 2011; Volkmer, Schulte-Körne, & Galuschka, 2019) and morphological interventions have been shown to foster reading and spelling skills (Carlisle, 2010; Goodwin & Ahn, 2010; McCutchen & Logan, 2011). There are also theoretical arguments implicating that the influence of morphological knowledge extends to writing quality (McCutchen, Stull, Herrera, Lotas, & Evans, 2014; Saddler & Graham, 2005). Thus, morphological interventions appear to be a promising way to support reading, spelling, and text production skills in learners with dyslexia and spelling deficits. This could be even more important in orthographies that adhere to the principle of morpheme consistency (i.e., where the spelling of morphemes is preserved in different word forms even if they have different pronunciations).

Theoretical Perspectives on Cross-linguistic Differences

This contemplation leads us to another important theoretical issue on spelling interventions: the influence of orthography-level differences on the efficacy of different spelling treatment approaches. The general framework of literacy development (as described above) is valid across alphabetic orthographies: it seems that, in European orthographies, alphabetic skill is a predictor of reading and spelling success (Caravolas, 2004). However, all aspects of spelling and reading are connected to the characteristics of a specific orthography. Orthographies vary in the degree to which the spellings reflect the phonology or morphology (Katz & Frost, 1992). In light of the phonics-first versus morphology-first debate, taking these orthography-level differences into account may prove pertinent.

In the previous decades, cross-linguistic studies mainly examined orthographic depth as a source of cross-linguistic differences in literacy acquisition (Serrano et al., 2011; Seymour, Aro, & Erskine, 2003). Orthographic depth is generally conceptualized as orthographic consistency: the presence of more than one pronunciation for a given letter or a cluster of letters (reading direction), or the presence of multiple spellings for one phoneme (spelling direction) (Borgwaldt, Hellwig, & De Groot, 2004; Schmalz, Marinus, Coltheart, & Castles, 2015). Consistency thus measures the extent of adherence to the alphabetic principle, which varies greatly between orthographies (Caravolas, 2005).

There is a general consensus about the approximate classification of several orthographies in terms of their orthographic consistency in reading (feedforward) and spelling (feedback)

direction (Seymour et al., 2003). Finnish has been shown to be on the “consistent” end of the orthographic consistency continuum, both in reading and spelling direction. English is on the opposite end of this continuum. It is a morphophonemic system in which morphemes are often preserved at the cost of phonological consistency (for example, the same grapheme, *c* is used to represent the phonemes /k/ and /j/ in the morphologically related word pair “magic” – “magician”). Historical spellings are also a big source of inconsistency in the English spelling system. Historical spellings represent sounds that were pronounced differently in the past than they are pronounced today (e.g., in English, words beginning with a silent *k*, such as *knight*). Like English, German has been found to be relatively inconsistent in the feedback direction (Borgwaldt et al., 2004; Borgwaldt, Hellwig, & De Groot, 2005). This is mainly caused by morpheme consistency, but there are also some graphotactic and context-dependent phonological features to consider. Vowel length marking is particularly challenging for German children at the beginning of literacy acquisition or children with spelling deficits. However, even in the spelling (feedback) direction, German is more consistent than English.

One method for a quantitative classification that allows for formal ranking of languages on a continuum based on their orthographic consistency was presented by Borgwaldt et al. (2005): namely, onset entropy. Onset entropy focuses on the relationship between the first letter and first phoneme of all words in a given orthography. It quantifies the extent to which words with the same first letter also have the same first phoneme as a measure of feedforward consistency, and vice versa for feedback consistency (for further details, see the method section of the current paper). The advantage of using word onset entropy measures is that the orthographic consistency of languages with varying orthographic structures can be objectively compared without relying on intuition (Borgwaldt et al., 2005; Schmalz et al., 2015).

Alphabetic orthographies contain varying degrees of orthography-specific uncertainties induced by morphological consistency, context-dependent phonological-orthographic features, and etymology. The higher the number and complexity of rules, the number of exceptions, and sources of inconsistency of an orthography, the more difficult it is to impart this knowledge and to use it for the remediation of spelling deficits (Caravolas, 2005). For very complex orthographies like English, it has been argued that interventions concentrating on memorization strategies of whole-word letter sequences are more efficient than interventions focusing on the teaching complex rules and regularities (Dymock & Nicholson, 2017; Hilte & Reitsma, 2011). This implies that the interventional aspects will

vary not only in their importance for a given orthographic system (Bowers & Bowers, 2017), but also in their applicability for spelling remediation. For interventional research on spelling, it is important to know if there are treatment approaches that are less effective in one orthography than in another. This would contribute to an understanding about the interaction between universal cognitive skills and language characteristics. Previous reviews and meta-analyses examined the effects of spelling interventions mainly for the English orthography (Bowers & Bowers, 2017; Graham & Santangelo, 2014; Wanzek et al., 2006; K. J. Williams et al., 2017). Reviews and meta-analyses that include studies from other languages do not allow for a comparison between the efficacy of spelling interventions in different orthographies (Ehri, Nunes, Stahl, et al., 2001; Ehri, Nunes, Willows, et al., 2001; Galuschka et al., 2014; Ise, Engel, & Schulte-Körne, 2012).

Potential Factors Influencing Treatment Efficacy

Previous syntheses on the effectiveness of treatment approaches for learners with dyslexia showed that interventions should focus directly on literacy skills and not on underlying auditory or visual causes to remediate spelling and reading skills (Galuschka et al., 2014). In addition, direct, explicit instruction in spelling has been shown to be more effective for spelling improvement than implicit self-study interventions. Previous studies also report that intervention components such as immediate corrective feedback and various exercise opportunities are beneficial for spelling skills (Wanzek et al., 2006; K. J. Williams, Walker, Vaughn, & Wanzek, 2017). Galuschka et al. (2014) and K. J. Williams et al. (2017) did not find a significant difference between interventions that were implemented individually or in small groups (two to six children) on spelling performance, so spelling deficits may be more effective when conducted in individual sessions and in small groups.

It seems likely that the implementation of interventions by professional personnel (e.g., teachers, psychologists, therapists) is of a higher quality than the implementation of interventions by paraprofessional personnel (personnel with less specific expertise like teacher assistants or students). It is therefore possible, that the professionalism of the implementer influences treatment efficacy. However, previous research on this topic has shown contradictory results (Scammacca, Roberts, Vaughn, & Stuebing, 2015; K. J. Williams et al., 2017). There is also a lack of consistent research findings concerning the influence of age and grade level on treatment efficacy. Ehri, Nunes, Stahl, et al. (2001) and Goodwin and Ahn (2013) showed that early intervention has the greatest impact on reading and spelling performance, compared to interventions performed in older children. A meta-analysis by Graham and Santangelo (2014) did not confirm this finding. Other meta-analyses on the

efficacy of reading and spelling treatment approaches (Galuschka et al., 2014; Wanzek et al., 2006), did not evaluate whether the efficacy of intervention is affected by age or grade level.

A further relevant topic with regard to effective treatment methods is the effect of software usage. Over the past decade, the use of software in educational and therapeutic context has increased (Jamshidifarsania, Garbayab, Limc, Blazevisa, & Ritchie, 2019). This is likely to bring many advantages: First, computerized interventions appear to encourage learning through exceptional motivational aspects (e.g., a rewarding system, an attractive graphic design, or an exciting storyline) (Ecalles, Magnan, Bouchafa, & Gombert, 2009; Kast, Baschera, Gross, Jäncke, & Meyer, 2011; Thalmann, 2014). Second, algorithms can enable the implementation of specialized instructions and adaptive content in computerized interventions (Truong, 2015). To our knowledge, the effectiveness of computerized spelling interventions versus classroom-settings has not yet been systematically tested using a meta-analytic approach. Wanzek et al. (2006) evaluated the influence of assistive software and demonstrated positive effects of software components on speech synthesis, word prediction, and spellchecking on measures of students' spelling accuracy and correction.

To date, it is unknown how the effectiveness of spelling interventions varies between different treatment approaches, and little is known about the impact of sample and intervention characteristics on the efficacy of spelling interventions (e.g., age and grade level of the participants, intensity or length of the intervention, setting characteristics). In addition, there is little empirical work on how the effectiveness of different spelling treatment approaches varies between orthographies.

Purpose Statement

The purpose of our systematic review and meta-analysis is to extend the current knowledge about the effectiveness of spelling treatment approaches on reading and spelling performance of learners with dyslexia and spelling deficits. The specific aims of the current study are fivefold.

First, we aim to present a comprehensive up-to-date overview (in the form of a systematic review and meta-analysis) of controlled trials of all available spelling interventions evaluated on children, adolescents and adults with dyslexia or spelling deficits. Previous syntheses indicated a moderate to high effect size of spelling interventions on spelling outcomes (Ise et al., 2012; K. J. Williams et al., 2017). We expect that, overall, spelling intervention is effective (Hypothesis 1 (H1)). However, we do not have specific predictions about the size of the effect. By including a description of the quality of the literature and using study design (randomized controlled trials, RCT, vs. non-randomized

controlled trials, NRCT), the standardization, and the blind assessment of the outcome measures as moderators, we aim to assess the validity of study findings, and to identify gaps which can be used to guide the design of future studies in the field.

Second, to advance towards a theory of spelling that describes how spelling performance of learners with dyslexia and spelling deficits can be improved, we perform a meta-analysis for different treatment approaches separately (*phonological awareness, phonics, morphological, orthographic, memorization and audio-visual cue interventions* as well as *assisted writing and supportive software* approaches). A detailed description of the interventions is displayed in the methods section. In line with the work of Weiser and Mathes (2011), Squires and Wolter (2016) and Goodwin and Ahn (2010, 2013), we expect effects to be significant for phonics, orthographic or morphological instructions (Hypotheses 2 (H2)). Since reading and spelling influence each other in all phases of literacy development, we evaluate the effects separately for reading and spelling.

The third aim is to investigate interaction effects between treatment approaches and age, as well as severity of reading and spelling disorder. With this analysis, we aim to examine the recent theoretical discussion between Rastle and Taylor (2018) and Bowers and Bowers (2018a), and to determine whether phonics intervention should precede morphological and orthographic intervention. Thus, we are especially interested in whether the efficacy of phonics, morphological or orthographic interventions is affected by age or severity of reading and spelling disorder. Considering the knowledge of phoneme-grapheme (and grapheme-phoneme) correspondences as a fundamental skill that affects higher-order skills, such as the application of orthographic or morphological spelling rules (Castles et al., 2018; Share, 1995), we hypothesize that more severely impaired dyslexics would benefit more from a phonics approach compared to an orthographic or morphological intervention. Furthermore, the efficacy of phonics instructions should decrease with age (Hypotheses 3 (H3)). As research is lacking on the evaluation of morphological and orthographic interventions in first or second grade, we do not have any apriori expectations about the efficacy of those interventions in the early years of literacy acquisition.

Fourth, we aim to contribute to an understanding how orthographic consistency influences the implementation and effectiveness of spelling interventions. Thus, we assess if treatment approaches vary between different orthographies and if the effect sizes for spelling interventions are affected by the orthographies in which they were implemented. We are especially interested in contrasting intervention effects between spelling interventions that were implemented in English and German orthography. Since a higher number of exceptions

and sources of inconsistency of orthography is associated with increased difficulties to remediate spelling deficits (Caravolas, 2005), we expected that for more consistent orthographies, the effect of spelling interventions on reading and spelling would be larger than for inconsistent orthographies. Thus, we assume that intervention effects are larger if the intervention was performed in German than in English orthography (Hypotheses 4 (H4)).

The fifth aim is to investigate whether effect sizes are affected by sample (age, grade level, severity of reading and spelling disorder) and intervention characteristics (amount of the intervention, setting, computer use, implementer), in an explorative way. In order to inform theory and practice, it is important to know whether the effects are consistent across learners and if general principles are relevant for a broad range of settings. Research on this topic is lacking for spelling interventions, therefore we do not make any predictions about which sample and intervention characteristics should influence intervention effects.

Method

We conducted the systematic review according to the recommendations of the Cochrane group (Higgins & Altman, 2008), and follow the PRISMA guidelines (Moher, Liberati, Tetzlaff, Altman, & The-PRISMA-Group, 2009).

Literature Search

Galuschka, Görgen and Kalmar searched for spelling intervention studies published from December of 1989 to December of 2018 in the following databases: ERIC, PsycINFO, Medline and peDOCS. The Literature search began in July 2016 and the following search terms were used in Abstract, Title and Keywords:

(Dyslexia OR spelling disorder OR developmental spelling disorder OR specific spelling disorder OR developmental reading disorder OR reading disability OR reading disorder OR Lesestörung OR Rechtschreibstörung OR Lese-Rechtschreibstörung OR Lese-Rechtschreibschwäche OR Leseschwäche OR Rechtschreibschwäche OR Legasthenie)

AND

(Invented spelling OR spelling OR writing OR writing comprehension OR writing improvement OR writing processes OR writing strategies OR orthographic knowledge OR grammar OR orthographic processing OR word lists OR letter OR orthograph* OR written communication OR diction OR notation OR morphological OR spelling

book OR spelling check OR Orthographie OR orthographisch OR rechtschreib* OR
orthographic rules OR Regeltraining OR Wortlisten OR dictation OR Diktat)

AND

(Intervention OR treatment OR therapy OR therapeutics OR training OR Förderung
OR Therapie OR remediation OR instruction OR exercises OR teaching)

To consider unpublished studies, we searched the website ProQuest using the
aforementioned search terms. We also examined bibliographical references of systematic
reviews (Graham et al., 2018; Graham & Santangelo, 2014; Ise et al., 2012; Squires &
Wolter, 2016; Wanzek et al., 2006; K. J. Williams et al., 2017). In addition, we contacted
experts by sending an e-mail to the members of the Society for the Scientific Studies of
Reading (SSSR) mailing list.

Study Selection Criteria

To be considered for this analysis, studies had to meet the following criteria:

(a) The intervention consisted at least partly of tasks and instructions that aimed to
improve spelling performance. The following contents were classified as tasks and
instructions that aimed to improve spelling performance: phonological and sound-based
exercises, such as verbal segmentation and analysis of words and syllables; exercises in
grapheme-phoneme mapping; memorization of word spellings, such as special pronunciations
of orthographic markers or visualization of written words in memory; explicit instruction and
application of graphotactic and orthographic spelling rules; morphological instructions in
base words, suffixes and inflected, and derived words; and all tasks that were explicitly
described in the original article as developed and used to improve spelling performance.

(b) One or more standardized spelling tests were administered before and after
treatment;

(c) Pre- and post-test results of the spelling tests were reported with sufficient detail to
allow for the calculation of an effect size or if this information could be obtained from the
authors;

(d) Study subjects were children (after starting formal schooling), adolescents, or
adults with spelling deficits or dyslexia whose spelling or reading performance was below the
25th percentile or at least 1 standard deviation (SD), 1 year, or 1 grade below the expected
age or grade level as described in the original study and with an intelligence quotient in the

normal range or, if the results of intelligence tests were not reported, subjects were described as having normal intelligence by the study authors;

(e) Poor spelling (and reading) performance occurred in the subjects' native language and subjects were also taught in their native language;

(f) The study included predominantly subjects without other developmental disorders (such as autism spectrum disorders), comorbid psychiatric disorders (internalizing disorders such as depression or externalizing disorders such as ADHD), neurological diseases (e.g., epilepsy), or chromosomal defects (e.g., trisomy 21)

(g) The study design included a control group: an untrained control group, placebo group, regular school practice control group, or another treatment control group;

(h) The manuscript was written in English or German (due to the authors' language proficiency). Studies not conducted in English or German were included as long as they were published in English or German.

Figure 1 summarizes the process of selecting studies for the meta-analysis.

[Figure 1 near here]

Coding of the Controlled Trials

Following the guidelines of the Cochrane group (Higgins & Altman, 2008), we used coding sheets to record the intervention, methodological, and sample characteristics (description of the characteristics below). To ensure reliability, study characteristics were double coded, which was done by Galuschka, Görgen and Kalmar independently. We determined interrater-reliability for article inclusion or exclusion by calculating Cohen's kappa (Cohen, 1960), and for the coding of the different treatment approaches (description of the treatment approaches below) by percentage of agreement (Cohen, 1960). For the calculations, we used the independent coding results of the raters. Cohen's kappa of the interrater agreement with regard to in- and exclusion was $k = .68$, and percentage of agreement for treatment approaches was 84%, on average. All discrepancies were discussed and resolved. Study details are provided in Supplement 1. Galuschka and Görgen extracted data necessary for effect size calculations (mostly means and SDs of pre- and post-tests). The extracted intervention characteristics from the original studies contained:

(a) The spelling treatment approach (detailed description below);

(b) Total amount of intervention in hours: < 20 hours; 20 to 40 hours; > 40 hours;

(c) Setting: class, group, or individual;

(d) Computer involvement: non-computerized; computer-based, or computer-assisted;

and

(e) Implementer: paraprofessionals, professionals, or a mix of paraprofessionals and professionals.

The spelling treatment approaches were classified into categories based on the descriptions of the intervention in the publication. The category *phonological awareness intervention* included approaches with phonological and sound-based exercises without or with very limited connection to written language. The category *phonics intervention* included approaches gaining basic knowledge of the relationship between phonemes and letters/graphemes as well as exercises combining phonological awareness with written letters. The category *memorization intervention* contained approaches that fostered word-specific memorization of letter sequences and the visual appearance of written words. The category *morphological intervention* included instruction in morpheme and syllable analysis, encouraging the participants to apply knowledge of syllable and morpheme structure to spelling. The category *orthographic intervention* comprised interventions that systematically instructed the application of graphotactic or orthographic-phonological spelling rules. The category *audio-visual cue intervention* comprised multisensory activities combining auditory and visual cues with the aim to support memory strength of graphemes and phonemes. The category *assisted writing* referred to interventions containing assisted writing activities by coaches and tutors. The category *supportive software* consisted of computer-assisted writing approaches.

The extracted methodological characteristics from the original studies included the dependent variable spelling (and reading); the sample size; and the administered spelling (and reading) tests. With regard to the sample, we also extracted (a) the age, (b) the grade level of the participants, (c) the degree of severity of reading and spelling disorder and (d) the language of instruction from the original studies and coded them as sample characteristics. If the authors did not specify the mean age of their participants, we estimated the age from other specifications like grade level or age range (Arnbak & Elbro, 2000; Farber, 2005; Gustafson, Ferreira, & Ronnberg, 2007; Rashotte, MacPhee, & Torgesen, 2001; Reuter-Liehr, 1993). The moderators grade and age level are not independent of each other. Nevertheless, we included both, because grade level refers to years in formal literacy instruction and it is important to know whether the intervention effect of certain treatment approaches changes in the course of the class levels. We used age because we also included studies with adults. Using the mean age of the study population allowed us to check for interaction effects between treatment approach and age (for H3) with a more meaningful analysis. Finally, we considered the following spelling outcomes: phonological spelling, pseudoword spelling and

real word spelling. Relevant reading outcomes were pseudoword reading, nonword reading, and word reading (accuracy, fluency and speed) as well as reading comprehension.

The degree of severity of reading and spelling disorder was determined by converting the reading and spelling standard scores of the pre-tests to z -scores. If more than one standardized reading test was administered and it was not possible to distinguish these by their relevance, the results of the standardized tests were z -transformed and averaged. If more than one standardized spelling test was administered, severity was determined by the standardized spelling test that addresses orthographic writing. Details of our severity assessment are provided in Supplement 2.

If we could not obtain standard scores for calculating z -scores, we used raw scores and grade-equivalent scores (Darch, Eaves, Crowe, Simmons, & Conniff, 2006; Gustafson et al., 2007; Hughes, Phillips, & Reed, 2013). We defined the severity categories in reading and spelling as mild (reading and spelling skills between the 25th and 16th percentile, between 0.67 SD and 1 SD or up to 1 year curricular delay), moderate (reading and spelling skills between the 15th and 7th percentile or 1 SD and 1.5 SD or between 1 and 1.5 years curricular delay) and severe (below the 7th percentile or more than 1.5 SD or above 1.5 year curricular delay).

Effect Size Calculation

To evaluate the efficacy of the interventions, we calculated Hedges' g (Hedges, 1981) and 95 % confidence intervals (CI). The effect size calculations were based on between-group comparisons of pre-post change scores using pooled pre-test SDs, as recommended by Morris (2008). For studies that did not report means and SDs, we calculated effect sizes on the basis of F -test values (Darch et al., 2006; Oakland, Black, Stanford, Nussbaum, & Balise, 1998). If studies included more than one intervention group, we followed the recommendations of Higgins and Green (2008): we combined subgroups if they received the same intervention and if the study subjects were in same age or grade level. If we could not combine subgroups of one study, an overweighting of the effect sizes was counteracted by dividing the sample size of the control group by the number of intervention groups. Some studies used multiple reading and spelling tests to determine treatment efficacy, including standardized measures and non-standardized measures. We preferred standardized measures for effect size calculation. However, if they were not available, non-standardized measures were used (Arnbak & Elbro, 2000; Gustafson et al., 2007; Kast, Meyer, Vogeli, Gross, & Jancke, 2007; Oakland et al., 1998). If a study reported results for several comparable tests (e.g., several standardized tests measuring different aspects of reading or spelling), we used

all tests for calculating the meta-analysis. To avoid errors with inclusion of cluster-randomized controlled trials, we followed the recommendations of Higgins (2011), and reduced the sample size to the number of clusters if statistical methods that account for clustering in the data were not presented in the original paper (Oakland et al., 1998; Reuter-Liehr, 1993; Strehlow et al., 1999; Vadasy, Sanders, & Peyton, 2005; Walter, Schliebe, & Barzen, 2007).

To estimate the average effect size and to account for the statistical dependencies due to multiple effect sizes from the same sample, we used robust variance estimation (RVE) with small sample adjustments (Tipton, 2015). RVE allows for the inclusion and synthesis of dependent effect sizes when correlations between those effect sizes are not reported (Hedges, Tipton, & Johnson, 2010). Instead, we estimated a common mean correlation of $\rho = .7$ between all pairs of effect sizes within each study, and a sensitivity analysis showed robust findings for values of ρ ranging from 0 to 1.0. We marked all mean effect sizes with an apostrophe (g').

If, in a subgroup analysis, synthesized studies reported only one outcome per study, we estimated the average effect sizes with the restricted maximum-likelihood estimator (REML). A random effect model was chosen since we expected the effect size to vary considerably between studies.

All statistical analyses were performed using the free software R version 3.4.0 (R Core Team, 2014). We used the packages *compute.es* (Del Re, 2013), *robumeta* (Fisher & Tipton, 2015), *metafor* (Viechtbauer, 2010), *clubSandwich* (Pustejovsky, 2016) and the Shiny application *weightr* that implements the weight function model published by Vevea and Woods (2005).

Study Quality and Publication Bias

We assessed study quality with the methodology checklist for (randomized) controlled trials published by the Scottish intercollegiate guideline network (Network, Harbour, & Forsyth, 2008). The checklist enables the evaluation of the internal validity of the studies and the possibility of biased results. We further evaluated each study to determine whether checks for treatment fidelity were conducted. We then calculated the percentage of items, for each study, that have either not been met or specified, to represent the risk of bias of each study. We calculated an average value from the resulting risks of bias per study, indicating the overall risk of bias. Lack of randomization, using non-standardized tests, and no blind assessment of the outcomes are likely to lead to overestimations of the true magnitude of an effect size. Thus, we included study design (RCT: yes or no), standardization of the

conducted tests which evaluated the interventions' outcomes (yes or no), and blind assessment of the outcomes (yes, no, or not specified) as moderators in our analysis.

We examined publication bias using the method of Egger, Davey Smith, Schneider, and Minder (1997). In addition, we followed the recommendations of recent publications (Carter, Schönbrodt, Gervais, & Hilgard, 2018, June 4; McShane, Böckenholt, & Hansen, 2016) and performed the weight-function model by Vevea and Woods (2005).

The Egger et al. (1997) method plots a regression line between the precision of studies (independent variable) and the standardized effect (dependent variable), and evaluates if the effect size is related to the size of the study. In the weight-function model of publication bias (Veeva & Woods, 2005), the weights represent the process by which some studies are more likely to be published than others, based on statistical significance (p -values). It is also possible to calculate an adjusted mean effect size estimate that offers a sense of the likely magnitude of the true effect.

Quantifying Orthographic Depth

There is, to date, no consensus about the best metric to quantify orthographic depth. The closest to an objective metric which can be applied to different alphabetic orthographies is onset entropy (Borgwaldt et al., 2004; Borgwaldt et al., 2005). Here, we considered feedback consistency, because the different ways of spelling a given phonological unit should be more relevant to spelling than the different ways of pronouncing a given orthographic unit. Onset entropy focuses on the different spellings of first phonemes: for example, in English, words beginning with the phoneme /dʒ/ can be spelled either with the first letter *g* (*gist*) or *j* (*just*). This means that, for the phoneme /dʒ/, onset entropy is higher than, say, for words beginning with the phoneme /t/, which are generally spelled with the first letter *t*. The entropy for all phoneme-letter associations can be averaged to obtain a single value of orthographic inconsistency for a given orthography. Entropy (H) for each phoneme (x) is calculated with the following formula:

$$H(x) = - \sum_{i=1}^n \left(p_i(x = y_i) \times \log_2(p_i(x = y_i)) \right)$$

For each different spelling of a given phoneme, y_i , we calculated the proportion of times that this spelling occurs relative to all occurrences of this phoneme. This proportion was then multiplied by its logarithm to the second base. The resulting value was summed across all different spellings of a given phoneme. As logarithms of proportions (or, more specifically, numbers between 0 and 1) are negative numbers, the entropy value was then

multiplied by -1 to give a positive number.

Here, we calculated the onset entropy in the phonology-to-orthography direction for English, German, Dutch, Danish, Swedish, and Norwegian (Bokmål). For English, German, and Dutch, the orthographic and phonological word forms were taken from the CELEX database (Baayen, Piepenbrock, & Gulikers, 1995). For Danish, Swedish, and Norwegian, the spoken and written word forms were taken from the NST Lexical Databases (available through Nasjonalbiblioteket, <https://www.nb.no/sprakbanken/>). For all languages, we included all monosyllabic words with greater-than-zero frequency counts. The input files and the Python script which was used to calculate entropy can be downloaded here: <https://osf.io/8r6zf/> (Schmalz, 2018). Unfortunately, we could not find an appropriate database for Brazilian Portuguese, therefore we could not calculate onset entropy for Ferraz et al. (2018).

Results

Results of the Literature Search

The extensive literature search in several electronic databases and systematic reviews revealed 34 studies that met all inclusion criteria and were included in this review.

Four studies included more than one intervention group (Calhoon, 2005; Gustafson et al., 2007; Lange, Mulhern, & Wylie, 2009; Rashotte et al., 2001). We combined the intervention groups, in line with the Cochrane Handbook for Systematic Reviews (Higgins & Green, 2008), if they evaluated similar interventions (the same treatment approach) on participants on the same grade level. We did not combine the two intervention groups (phonological awareness intervention, orthographic intervention) of Gustafson et al. (2007) and the three subgroups (1-2 grade, 3-4 grade, 5-6 grade) of the study by Rashotte et al. (2001), to ensure that we did not lose important information. In total, we included 38 intervention and control group comparisons.

In the 34 studies, the following treatment approaches were implemented: phonics intervention (n = 8), morphological intervention (n = 10), orthographic intervention (n = 7), memorization intervention (n = 5), audiovisual cue intervention (n = 1), supportive software (n = 2), and assisted writing (n = 1). Seventeen studies were conducted in English, eleven studies in German, and two studies were conducted in Swedish. The other studies were each conducted in Norwegian, Danish, Dutch, and Brazilian Portuguese.

Study Quality

Nineteen of the included studies reported a randomized group allocation on the individual level and, and we therefore considered them RCTs. In turn, five trials assigned clusters of patients (from one classrooms, school or clinic) to intervention or control groups (cluster controlled trials (cCT) or cluster RCTs (cRCT)) (Oakland et al., 1998; Reuter-Liehr, 1993; Strehlow et al., 1999; Vadasy et al., 2005; Walter et al., 2007); ten studies did not perform a randomized group allocation, and we therefore considered them NRCTs. The moderator analysis showed that NRCTs or cRCTs found a higher mean effect size ($g' = 0.70$, 95% CI [0.26 – 1.14]) than RCTs ($g' = 0.53$, 95% CI [0.36 – 0.70]), but this difference did not reach statistical significance ($p = 0.510$). Five studies used non-standardized tests to evaluate at least one of the outcome measures (Arnbak & Elbro, 2000; Ferraz et al., 2018; Gustafson et al., 2007; Kast et al., 2007; Oakland et al., 1998). Analyses showed a higher mean effect size for standardized tests ($g' = 0.62$, 95% CI [0.42 – 0.82]) than for non-standardized measures ($g' = 0.29$, 95% CI [-0.00 - 0.58]), but again no significant difference ($p = .072$). Risk of detection bias is high, because only six studies performed a blind assessment of treatment outcomes (Berninger, Lee, Abbott, & Breznitz, 2013; Berninger et al., 2008; Given, Wasserman, Chari, Beattie, & Eden, 2008; Gray, 2015; Gustafson et al., 2007; Hughes et al., 2013). Studies with blind assessment showed a lower mean effect size ($g' = 0.33$) than studies that did not perform a blind assessment ($g' = 0.56$, 95% CI [0.34 – 0.78]), or did not specify if a blind assessment took place ($g' = 0.77$, 95% CI [0.33 – 1.22]). This difference on the descriptive level could not be statistically confirmed ($p = 0.278$).

Eighteen studies provided details on how the authors checked for treatment fidelity. In these eighteen studies, treatment fidelity was mostly ensured through supervising sessions. However, only very few publications reported the percentage of the level of adherence to the treatment script or plan. The vast majority of our study pool did not state whether there were any differences in treatment fidelity across the control and intervention groups, or between pre- and post-test. Details of the quality assessment for each study are displayed in Supplement 3.

Effectiveness of Spelling Interventions

We conducted a meta-analysis which included 28 studies (31 experimental and control group comparisons). The analysis included placebo, waiting list, and untreated control groups, or compared the treatment effects against control groups that received regular school practice or other treatment approaches that did not correspond to treatment approaches of the experimental groups in other included studies. Table 1 displays the results of the meta-

analysis separately for reading and spelling performance. In regards to our first hypothesis (H1), we found, overall, a moderate effect size for spelling treatments on spelling performance and a small to moderate effect size for reading performance.

[Table 1 near here]

With the high I^2 values, there was evidence of substantial heterogeneity. This indicates that the analysis combined intervention effects which are likely drawn from populations with different true effect sizes. To address this issue, and to explore the general effectiveness of different spelling treatments on reading and spelling performance (H2), we performed subgroup analyses. We calculated separate mean effect sizes for the identified treatment approaches if they comprised a minimum of two studies.

Eight studies were coded as phonics (Blachman et al., 2014; Ehri, Satlow, & Gaskins, 2009; Ferraz et al., 2018; Hughes, Phillips, & Reed, 2013; O'Shaughnessy & Swanson, 2000; Oakland, Black, Stanford, Nussbaum, & Balise, 1998; Rashotte, MacPhee, & Torgesen, 2001; Vadasy, Sanders, & Peyton, 2005), eight studies as morphological (Arnbak & Elbro, 2000; Corvacho del Toro, 2016; Gustafson, Ferreira, & Ronnberg, 2007; Kirk & Gillon, 2009; Reuter-Liehr, 1993; Tijms, 2011; Walter, Schliebe, & Barzen, 2007; Weber, Marx, & Schneider, 2007), six studies as orthographic (Calhoon, Sandow, & Hunter, 2010; Darch, Eaves, Crowe, Simmons, & Conniff, 2006; Faber, 2005; Given, Wasserman, Chari, Beattie, & Eden, 2008; Groth, Hasko, Bruder, Kunze, & Schulte-Körne, 2013; Ise & Schulte-Körne, 2010), two studies as memorization intervention (Spencer, Snart, & Das, 1989; Thaler, Landerl, & Reitsma, 2008) and two studies as supportive software (Fasting & Halaas Lyster, 2005; Lange, Mulhern, & Wylie, 2009).

As can be seen in Table 2, *phonics*, *morphological* and *orthographic interventions* showed significant effect sizes for reading and spelling outcomes. Effect sizes for spelling outcomes were substantially larger than for reading outcomes.

[Table 2 near here]

Studies that evaluated *memorization interventions* reported pre- and post-test results only for spelling tests, which made it impossible to estimate the efficacy of this treatment approach on reading. Memorization interventions did not show a significant effect on spelling. However, as there were only two studies in this category, the results have to be interpreted with caution.

A large and significant mean effect size for spelling outcomes was found for studies evaluating *supportive software*. The effect on spelling can be considered large and homogeneous. The effectiveness of *supportive software* on reading performance could not be

statistically confirmed. Again, there are only two studies in this category, results of this subgroup analysis have to be interpreted with caution. The mean effect sizes of the treatment approaches *phonics*, *morphological*, *orthographic*, *memorization interventions* and *supportive software* did not differ significantly from each other (see Table 5). Only one study evaluated an *audio-visual-cue intervention* (Kast et al., 2007). In this study, a non-standardized spelling test was used. Analyses showed an effect of $g = 0.29$ on spelling of untrained words, and an effect of $g = 0.52$ on trained word material. *Phonological awareness intervention* was also evaluated from only one study (Gustafson et al., 2007) and did not significantly enhance reading and spelling skills of the children (spelling: $g = -0.03$; text reading: $g = 0.03$; reading comprehension: $g = -0.12$). Jensen, Lindgren, Andersson, Ingvar, and Levander (2000) evaluated a treatment that assisted the participating Swedish adults with reading and spelling disorder in their writing and reading activities with the aim to encourage and motivate them for literacy activities. Effect sizes for reading and spelling outcomes of all individual studies and outcomes are displayed in Supplement 4.

Qualitative Evaluation of Studies Comparing Different Treatment Approaches.

In six studies, treatment approaches which included components aiming to support spelling skills were compared directly (Abbott & Berninger, 1999; Berninger et al., 2013; Berninger et al., 2008; Gray, 2015; Schulte-Körne, Deimel, Hülsmann, Seidler, & Remschmidt, 2001; Strehlow et al., 1999). In other words, the treatment approaches chosen for the control groups corresponded to our coded spelling intervention categories, which served as an exclusion criterion for the current meta-analysis (see methods section). However, as they contain important information, we synthesized them narratively. In addition, we calculated effect sizes based on between-group comparisons of the pre-post change scores.

The study of Abbott and Berninger (1999) compared a phonics intervention with a combined approach of phonics and morphological intervention (coded as morphological intervention due to explicit instruction in syllable and morpheme analysis), conducted over a period of four months. The results indicated that spelling performance of children between 9 and 13 years tends to benefit more from individual instruction in syllabic and morphemic structure than from pure letter-sound relationship instruction. The combined group showed better spelling performance after the intervention than the pure phonics group ($g = 0.25$, 95% CI [-0.66, 1.14]; *Wide Range Achievement Test*, WRAT (Wilkinson, 1993); $g = 0.06$, 95% CI [-0.84, 0.96], *Wechsler Individual Achievement test*, WIAT (Wechsler, 1992)). The results of the reading tests (decoding and reading fluency) did not reveal any differences between the

two groups (g ranging between -0.02 and 0.12), but individual growth curves showed a trend towards greater individual treatment response for the group receiving combined treatment of phonics and morphological instruction.

Berninger et al. (2013) compared the effects of two treatment approaches on groups of dyslexic children aged between 9 and 14 years by stepwise introducing different spelling components and strategies over a period of five months: First, they compared phonics instruction for spelling and reading to phonics in the reading direction only. They found no distinct advantage of one strategy over the other. Adding spelling memorization strategies (mentally photographing a word to identify letters in their correct word position and imagining a word in the mind's eye before writing) and comparing it with phonics components indicated that memorization strategies are less beneficial for improving phonological spelling ($g = -0.55$, 95% CI [-1.36, 0.3]), though it may be effective for improving word decoding skills ($g = 0.45$, 95% CI [-0.39, 1.26]). In a subsequent step, they added morphological instruction and compared to memorization strategies. They observed no significant differences in reading and spelling after this step. The authors noted that the duration of morphological instruction may have not been sufficiently long to identify further improvement in reading and spelling.

Three further studies compared morphological intervention and memorization intervention (Berninger et al., 2008; Gray, 2015; Strehlow et al., 1999). The results of these three studies were largely consistent with each other. Strehlow et al. (1999) evaluated a memorization intervention that aimed to improve spelling and learning transfer to untrained words by memorizing the correct spelling of highly frequent regular and irregular German words. The control intervention was coded as a morphological intervention. Pretests revealed differences in intelligence between the two groups of third grade children (in favor of the memorization intervention group). After 1 year of individual treatment, spelling improvement was significantly greater for children receiving memorization intervention than for children receiving morphological intervention ($g = 0.93$, 95% CI [0.04, 1.76]). Considering the differences between groups on the cognitive level, effects could not solely be attributed to the interventions but could also be related to the differences in intelligence.

Gray (2015) conducted a study with adults. After 4 weeks of individual intervention, participants of the morphological group showed greater gains in reading (effect sizes of reading tests between $g = 0.21$ -0.40), and reading comprehension ($g = 0.30$, 95% CI [-0.39, 0.98]), than participants receiving memorization intervention. In reading, in contrast to the memorization group, the morphological group showed transfer effects from taught to

untaught word material. No clear progress in spelling was revealed for either treatment ($g = -0.03 - 0.10$).

Berninger et al. (2008) evaluated a memorization intervention (referred to as “orthographic spelling treatment” in the original study) and a morphological intervention (“morphological spelling treatment”), conducted in a group setting with adolescents in Grades 4-9. The memorization intervention aimed to support precise, high-quality orthographic representations. The goal of the morphological intervention was to help children to establish representations of base-words and affixes and to improve their knowledge of morphological spelling rules. After a three-week period, the results revealed no clear differences between both groups in orthographic spelling ($g = 0.16$, 95% CI [-0.48, 0.79]). However, the morphological group seemed to improve more in phonological spelling ($g = 0.48$, 95% CI [-0.17, 1.12]) than the memorization group. However, the memorization group showed better performance in phonological decoding compared to the morphological group ($g = 0.45$, 95% CI [-0.2, 1.08]).

Schulte-Körne et al. (2001) compared the effectiveness of a German orthographic intervention with a phonological awareness intervention in group setting, including children between the ages of 9 and 11 years, with moderate reading and severe spelling impairment. In comparison to the control group, the experimental group improved in spelling performance ($g = 0.78$, 95% CI [-0.19, 1.68]) and reading ($g = 0.66$, 95% CI [-0.29, 1.56]). In a pre-post comparison of the phonological awareness group, spelling performance of the control group decreased. Schulte-Körne et al. (2001) concluded that an ineffective treatment, similar to no treatment, leads to degradation in performance compared to children of the same age.

In sum, morphological interventions were applied mainly to children from the upper elementary grades (Abbott & Berninger, 1999; Berninger et al., 2013; Berninger et al., 2008) and young adults (Gray, 2015). As pointed out by Gray (2015), morphological interventions could enable transfer effects from familiar to unfamiliar word material. This appears to be especially important for reading fluency, comprehension and spelling of unknown words (Berninger et al., 2008; Gray, 2015). Benefits of memorization interventions, in comparison to phonics and morphological interventions, could not be confirmed. The results of Schulte-Körne et al. (2001) underline the relevance of effective spelling treatment approaches and support the effectiveness of orthographic intervention to improve orthographic spelling for German children. Detailed information of the narratively synthesized studies is also provided in Supplement 1.

Influence of Age and Severity of Reading and Spelling Deficit on the Effectiveness of Different Spelling Treatment Approaches

To address the third aim of our systematic review, we performed meta-regressions. For meta-regressions, all outcomes (reading and spelling skills) were combined. All outcomes allowed us to use all information and to increase statistical power. Studies that did not include or did not specify a certain variable were excluded from the meta-regression in question. We did not find a significant main effect of age or severity on the magnitude of intervention effects (see Table 3).

To test hypothesis three (H3), we performed meta-regression to evaluate if the mean effect sizes of phonics, orthographic and morphological interventions would be affected by age and severity of reading and spelling disorder. The effectiveness of phonics interventions tended to decrease with age, and the effectiveness of morphological interventions tended to increase with age, but this could only be observed on a descriptive level. Against our hypotheses, the efficacy of phonics interventions decreased with increasing severity, while the efficacy of orthographic and morphological interventions increased with increasing severity (see Figure 2). However, due to the small dataset, main effects and interactions of the meta-regression models did not allow for clear conclusions (see Table 4).

[Table 4 near here]

Cross-linguistic Differences

To address the fourth aim (and H4), we performed meta-regressions and a moderator analysis (see Table 3 and Table 5) to explore if the mean effect sizes of studies vary depending on orthographic consistency. All outcomes (reading and spelling skills) were combined. Seventeen of our included studies were performed in English-speaking countries (Onset entropy = 0.286) and showed a moderate mean effect size. Ten studies came from German-speaking countries (Onset entropy: 0.123) and showed a large mean effect size. The other studies came from Norway (Onset entropy: 0.128), Denmark (Onset entropy: 0.309), Sweden (Onset entropy: 0.124), the Netherlands (Onset entropy: 0.124) and Brazil (onset entropy could not be calculated) (each orthography was represented by one study). Performing a moderator analysis, interventions in German descriptively showed larger effect sizes than interventions conducted in English, but this difference did not reach statistical significance.

We also evaluated if the effectiveness of spelling interventions varies as a function of onset entropy. Again, this could not be statistically confirmed. Phonics interventions were mainly evaluated for English-speaking children, with one exception, and orthographic and

memorization interventions only in German and English. Morphological intervention components seem to be in widespread use in different orthographies like German, Danish, Dutch and English.

Influence of Intervention and Sample Characteristics

To investigate our fifth research aim, we performed meta-regressions and moderator analyses (see Table 3 and Table 5). None of the moderator analyses indicated a significant difference. We observed a trend regarding the setting of an intervention ($p = .061$).

Interventions that were conducted in the classroom tended to show a smaller mean effect size than interventions that were implemented in group or individual setting. For severity, age, and amount of intervention as continuous variables we performed meta-regressions.

However, no significant link between the amount of intervention and intervention effects was revealed.

[Table 3 near here]

[Table 5 near here]

Publication Bias

We evaluated publication bias in 28 studies (31 experimental and control group comparisons), that we included in the meta-analysis for reading and spelling performance. Based on Egger et al. (1997)'s regression method, we found significant publication bias (standard errors significantly predicted the effect sizes, $p = .035$). To illustrate this bias, we performed Vevea and Woods (2005) weight function model. We applied the weight functions under consideration of the summary of them. The exact weights we used are presented in Supplement 5.

Assuming a moderate one-tailed selection, our weight function revealed a transformed mean of $g' = 0.43$. The weight function model is based on the assumption that the study's p -value affects its probability of publication. It is difficult, if not impossible, to model this adequately. Furthermore, the weight function model was performed on dependent mean effect sizes, which might have an influence on the estimation of the adjusted mean effect size. Thus, the adjusted mean effect size can only be an indication of the true effect size.

Discussion

The purpose of our systematic review and meta-analysis was to extend the current knowledge about the effectiveness of spelling treatment approaches on reading and spelling performance of learners with dyslexia and spelling deficits. Testing our first hypothesis (H1), we found that spelling interventions are effective: Learners with dyslexia or spelling deficits

who take part in a spelling intervention show better reading and spelling performance, compared to children who received regular school practice, or no spelling instruction. This finding is consistent with previous syntheses evaluating the effectiveness of spelling interventions on children with reading or spelling disabilities or dyslexia (Squires & Wolter, 2016; Wanzek et al., 2006; Weiser & Mathes, 2011; K. J. Williams et al., 2017).

Effectiveness of Spelling Interventions

As we hypothesized (H2), mean effect sizes for spelling instructions involving phonics, orthographic and morphological interventions are significantly greater than zero. Phonics, morphological, and orthographic interventions support children by making the spoken and written language system more transparent. This can help build and automate spoken and written language structures and, in turn, reduce cognitive load. As expected, the present meta-analysis is consistent with earlier work by Weiser and Mathes (2011), Squires and Wolter (2016) and Goodwin and Ahn (2010, 2013). However, Goodwin and Ahn (2010, 2013) showed only a small effect size for morphological trainings on spelling (Goodwin & Ahn, 2010, 2013), whereas the mean effect size of morphological interventions on spelling performance in our analyses can be considered large. This discrepancy in mean effect sizes can be explained by two major differences: First, some of the morphological interventions in our synthesis also included syllable analysis and blending exercises, which influences phonological spelling (e.g., Berninger et al., 2008). Secondly, instruction of morphemes was always related to their impact on spelling. In order to foster spelling with morphological interventions, it is especially important to explicitly display the effects of morphological regularities on spelling (Bowers & Bowers, 2017). In addition to spelling, morphological interventions seem to have a positive influence on children with different impairments. They also seem to be very promising for children with additional reading deficits (Goodwin & Ahn, 2010, 2013), specific language impairment (Berninger et al., 2013) and problems in composition or written expression (Graham & Santangelo, 2014).

Using qualitative and quantitative methods, we did not find a significant influence of memorization intervention on spelling performance. However, we found only two studies evaluating memorization interventions which could be included in the meta-analysis (Spencer, Snart, & Das, 1989; Thaler, Landerl, & Reitsma, 2008). Thus, we may have had insufficient power to detect a small but true improvement. Nevertheless, they do not seem to be effective for remediating problems with phonological spelling (Berninger et al., 2013), and they do not support the acquisition of strategies which lead to transfer effects (Berninger et al., 2008; Gray, 2015).

Phonics First or Morphology

We did not find that phonics instruction is more effective than morphological interventions in the early years of formal literacy instruction or for more severe spelling deficits. According to our meta-regression analyses, it seems likely that phonics, morphological and orthographic interventions are applicable across a wide age span. Our qualitative synthesis indicates that instruction in morphemic structure might be especially promising for children in the upper elementary and secondary grades (Abbott & Berninger, 1999). However, meta-analyses of morphological interventions demonstrated significantly larger effects for preschool and early elementary students than for older students in the upper elementary grades and high school (Goodwin & Ahn, 2010, 2013).

Phonics treatments aim to help students crack the alphabetic code and are therefore widely used in treating children with dyslexia and spelling deficits, especially during the early stages of reading and spelling acquisition (Berninger et al., 2013; Berninger et al., 2008). The knowledge of phoneme-grapheme (and grapheme-phoneme) correspondences is a fundamental skill that affects the application of orthographic or morphological spelling rules (Castles et al., 2018; Share, 1995). However, simple grapheme-phoneme and phoneme-grapheme correspondences are not sufficient for skilled spelling. Orthographic and morphological interventions showed high to moderate effect sizes in our analysis. Thus, we argue that graphotactic and orthographic-phonological spelling rules, as well as morphological instruction, should be provided as soon as children master the basic phoneme-grapheme and grapheme-phoneme mappings and are confronted with text material for which the basic phoneme-grapheme and grapheme-phoneme mappings would lead to spelling and reading mistakes. In order to resolve the question if, at the beginning of literacy acquisition, phonics instruction alone leads to better literacy outcomes than a combination of phonics and morphological instruction, more primary studies are needed.

Cross-linguistic Differences

In line with our fourth hypothesis, the present meta-analysis showed a higher mean effect size for spelling interventions implemented in German than for spelling interventions implemented in English. However, this difference was not statistically significant. Phonics was the most investigated treatment approach for spelling interventions in English. In German, more orthographic and morphological interventions were evaluated. The few studies evaluating orthographic or morphological interventions in English indicated moderate to large effect sizes on spelling skills (Berninger et al., 2013; Calhoon, Sandow, & Hunter, 2010; Darch et al., 2006; Kirk & Gillon, 2009). Even in English, the spelling of many words

can be explained through morphological and orthographic rules, and this resource should not be ignored in spelling intervention (Bowers & Bowers, 2017). Thus, in the future, more English studies should be conducted which would assess treatment approaches dealing with orthographic and morphological regularities.

Besides the studies with German and English interventions, our meta-analysis comprised only five studies from other alphabetic orthographies (Dutch, Danish, Swedish, Norwegian, and Brazilian Portuguese). The inclusion of studies from all languages and the use of onset entropy values allowed us to assess the effect of orthography on the effectiveness of intervention along a continuum. We did not find support for our hypothesis, that with increasing orthographic consistency, the intervention effects would increase. As is typical in meta-analysis, the search procedure was limited to studies published in the languages in which the authors were sufficiently proficient to conduct the search and to extract the relevant information. While searching for papers in a larger variety of languages would undoubtedly present a fuller picture, the current analysis offers the first step to capture the influence of orthographic consistency on spelling intervention and also serves as a model of how this can be done in the future.

Nevertheless, our review showed that, across orthographies and even in inconsistent orthographies such as English and Danish, learners with dyslexia and spelling deficits benefit from interventions which are proximal contributors to spelling acquisition and focus on the logic underlying the orthographic system. In orthographies varying in consistency, morphological and orthographic interventions are effective components for spelling remediation.

Factors Influencing Treatment Efficacy

Another aim of the present meta-analysis was to explore several factors influencing the effectiveness of interventions. Regarding the use of technology in spelling intervention, we found a large and statistically significant effect of supportive software on spelling skills. This indicates a potential for supportive software as an additional aid for patients with dyslexia and spelling deficits. Unfortunately, the limited study pool in this subgroup analyses precludes clear conclusions. Further research should tackle this topic more systematically to gain more evidence of how software can be used to help persons with dyslexia or spelling problems to cope with their spelling deficits in professional and daily life. A current meta-analysis on 5 to 16 year old students has already shown that technological multimedia-based tools for writing can enhance the motivation and the users composing skills (C. Williams & Beam, 2019). Our subgroup analyses also touched on the topic of computer involvement. We

found that traditional individual instruction cannot be substituted by computer-based approaches as it was also stated by C. Williams and Beam (2019). Nevertheless, the significant mean effect sizes on reading and spelling performance with regard to computer-based approaches are worth noting. Computer-based programs can be helpful for individuals with mild reading and spelling deficits, or they can be useful for bridging the gap until individual treatment can be provided. Individual treatment showed the largest mean effect size, which makes it very likely that computer-based approaches alone are not sufficient, especially for severely affected individuals.

Comparing different settings, interventions that were conducted in the classroom showed the smallest effects on spelling performance. Individual settings appeared to show the largest effect size, suggesting that spelling skills benefit particularly from one-to-one instructions. In classroom settings, instruction may not be specific enough and the work pace not sufficiently adapted to the individual. Our analysis revealed moderate effect sizes on spelling performance for interventions that were implemented in group settings, indicating that intervention in small groups could be an efficient alternative to individual instruction. This finding is similar to research on reading performance, which showed that one-to-one and small group instruction tend to be particularly effective (Wanzek et al., 2016; Wanzek & Vaughn, 2008).

We could not determine the effect of age or grade level on the efficacy of an intervention. Though the age of the participants in our study pool ranged from grade 1 through adulthood, only a few studies evaluated spelling interventions on children in first grade or in adults. Early intervention seems to be important for an effective remediation (Graham & Santangelo, 2014) and dyslexia and spelling deficits last a lifetime (Maughan et al., 2009). Therefore, it is particularly important that future studies represent learners with dyslexia and spelling deficits at all ages.

Influence of Study Quality and Publication Bias

To ensure high methodological quality of the present meta-analysis, only controlled trials were considered for analysis. Even though our analyses did not show any significant differences between study design, standardization of the conducted tests and blind assessment of the outcomes, we descriptively saw lower mean effect sizes if a randomized group allocation and blind assessment of the outcome measures was performed. To minimize risk of bias, the magnitude of systematic errors regarding randomization as well as blinding should be considered. It is critical to ensure that an intervention was implemented as planned, and that it was delivered to each study subject in a comparable manner, to attain and demonstrate

treatment fidelity. Of our included studies, 47% did not control or did not mention how they controlled for treatment fidelity. Negligence of this aspect was also criticized by previous work (Graham & Santangelo, 2014).

It is well-known that publication bias and questionable research practices inflate effect size estimates of meta-analyses (Gelman & Carlin, 2014; Gelman & Loken, 2014; Schmidt, 1992). Our attempts to control for publication bias showed that the effects might be overestimated in our meta-analysis. It is likely that, if all relevant studies were included, the effect size might shift, but the key finding would probably remain unchanged. The adjusted mean effect size is an indication of the true unbiased effect size, but it should be interpreted with caution. The weight function model was performed on dependent mean effect sizes, which might have an influence on the estimation of the adjusted mean effect size. While we are unable to estimate exactly what effect publication bias has on the results of the current study, it is an important conclusion in itself that there is publication bias in the literature. This calls for pre-registered studies to assess the impact of spelling or reading treatments in the future.

Limitations

A limitation of our meta-analysis and systematic review is a lack of an extensive cross-linguistic comparison of spelling interventions. The controlled trials included in the current study mainly focused on spelling interventions in the German and English orthography, due to the language proficiency of the authors. Our conclusions therefore are strongest in relation to the English and German orthographies.

We also cannot make any claims as to whether spelling instruction leads to improvements in different aspects of spelling (e.g., spelling word lists, spelling in context, phonological spelling) or other writing skills (e.g., compositional fluency or quality), because the studies we included in our meta-analysis assessed spelling mainly by word dictation and none of our included studies assessed the output of their intervention on writing measures. To resolve this question more primary studies are needed.

The meta-analysis showed substantial heterogeneity in our study pool. It is a common criticism that meta-analysis comprise studies that are not compatible with each other. To prevent this and to perform a meaningful analysis, we only included studies with a minimum required quality (controlled trials), applied up-to-date methodological techniques, and performed meta-regressions and subgroup analyses to identify factors that could influence the efficacy of interventions. In addition, we included a qualitative synthesis.

Nevertheless, some of the spelling interventions were part of a broad literacy

treatment while others focused on one aspect of spelling instruction. Our coding scheme had the main objective of accurately categorizing the interventions, but we cannot rule out that there are other factors that may have influenced the intervention effects. By focusing on learners with dyslexia or spelling deficits, the results of the current meta-analysis can only be transferred to this population. To assess the effectiveness of the intervention programs on other groups, evaluating the programs with other samples of participants beside those with dyslexia is mandatory.

Conclusion

The present meta-analysis shows that phonics, orthographic and morphological interventions are effective components in treating spelling of learners with dyslexia and spelling deficits. Morphological interventions are researched across various orthographies, while pure phonics interventions were studied in English-speaking countries only. Findings from our analysis identify a gap in the literature on early implementation of spelling interventions. Given that classroom interventions do not seem sufficiently specific to an individual's needs, the implementation of small group interventions in educational systems is essential. Of particular importance are interventions in individual settings, as shown by large effect sizes in our analysis. Computer-based and computer-assisted approaches can be a valuable alternative, or they can be used to complement individual or small-group interventions. It is striking that the support of adults with reading and spelling deficits is often neglected. Considering that dyslexia is a lifelong condition, future research should also focus on interventions addressing the question of how adults with dyslexia can be assisted. We further conclude that, to date, across orthographies an evidenced-based implementation of spelling interventions involves the application of phonics, the explicit instruction of graphotactic or orthographic-phonological spelling rules, and morphological instruction.

References

- Abbott, S. P., & Berninger, V. W. (1999). It's never too late to remediate: Teaching word recognition to students with reading disabilities in grades 4–7. *Annals of Dyslexia*, 49, 223-250. doi:10.1007/s11881-999-0025-x
- *Arnbak, E., & Elbro, C. (2000). The effects of morphological awareness training on the reading and spelling skills of young dyslexics. *Scandinavian Journal of Educational Research*, 44(3), 229-251. doi:10.1080/00313830050154485
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX Lexical Database. Release 2 (CD-ROM): Linguistic Data Consortium, University of Pennsylvania.
- Berninger, V. W., Lee, Y.-L., Abbott, R. D., & Breznitz, Z. (2013). Teaching children with dyslexia to spell in a reading-writers' workshop. *Annals of Dyslexia*, 63(1), 1-24. doi:10.1007/s11881-011-0054-0
- Berninger, V. W., & Swanson, H. L. (1994). Modifying Hayes and Flower's model of skilled writing to explain beginning and developing writing. *Advances in cognition and educational practice*, 2, 57-81.
- Berninger, V. W., Winn, W. D., Stock, P., Abbott, R. D., Eschen, K., Lin, S.-J., Nagy, W. (2008). Tier 3 specialized writing instruction for students with dyslexia. *Reading and Writing*, 21(1-2), 95-129. doi:10.1007/s11145-007-9066-x
- *Blachman, B. A., Schatschneider, C., Fletcher, J. M., Murray, M. S., Munger, K. A., & Vaughn, M. G. (2014). Intensive reading remediation in grade 2 or 3: Are there effects a decade later? *Journal of Educational Psychology*, 106(1), 46. doi:10.1037/a0033663
- Borgwaldt, S. R., Hellwig, F. M., & De Groot, A. (2004). Word-initial entropy in five languages: Letter to sound, and sound to letter. *Written Language & Literacy*, 7(2), 165-184. doi:10.1075/wll.7.2.03bor

- Borgwaldt, S. R., Hellwig, F. M., & De Groot, A. M. (2005). Onset entropy matters—Letter-to-phoneme mappings in seven languages. *Reading and Writing, 18*(3), 211-229.
doi:10.1007/s11145-005-3001-9
- Bosman, A. M., & Van Orden, G. C. (1997). Why spelling is more difficult than reading. *Learning to spell: Research, theory, and practice across languages, 10*, 173-194.
- Bowers, J. S., & Bowers, P. N. (2017). Beyond phonics: the case for teaching children the logic of the English spelling system. *Educational Psychologist, 52*(2), 124-141.
doi:10.1080/00461520.2017.1288571
- Bowers, J. S., & Bowers, P. N. (2018a). The importance of correctly characterising the English spelling system when devising and evaluating methods of reading instruction: Comment on Taylor, Davis, and Rastle (2017). *Quarterly Journal of Experimental Psychology, 174*7021818759477. doi:10.1177/1747021818759477
- Bowers, J. S., & Bowers, P. N. (2018b). *There is no evidence to support the hypothesis that systematic phonics should precede morphological instruction: Response to Rastle and colleagues*. Retrieved from <https://doi.org/10.31234/osf.io/mvdqr>
- Bus, A. G., & van Ijzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology, 91*(3), 403-414. doi:10.1037/0022-0663.91.3.403
- Calhoon, M. B. (2005). Effects of a Peer-Mediated Phonological Skill and Reading Comprehension Program on Reading Skill Acquisition for Middle School Students with Reading Disabilities. *Journal of Learning Disabilities, 38*(5), 424-433.
doi:10.1177/00222194050380050501
- *Calhoon, M. B., Sandow, A., & Hunter, C. V. (2010). Reorganizing the instructional reading components: could there be a better way to design remedial reading programs to

maximize middle school students with reading disabilities' response to treatment?

Annals of Dyslexia, 60(1), 57-85. doi:10.1007/s11881-009-0033-x

Caravolas, M. (2004). Spelling development in alphabetic writing systems: A cross-linguistic perspective. *European Psychologist*, 9(1), 3-14. doi:10.1027/1016-9040.9.1.3

Caravolas, M. (2005). Learning to spell in different languages: How orthographic variables might affect early literacy. In R. M. Joshi & P. G. Aaron (Eds.), *Handbook of Orthography and Literacy* (pp. 497-511). Mahwah, NJ: Lawrence Erlbaum Associates.

Carlisle, J. F. (2003). Morphology matters in learning to read: A commentary. *Reading Psychology*, 24(3-4), 291-322. doi:10.1080/02702710390227369

Carlisle, J. F. (2010). Effects of instruction in morphological awareness on literacy achievement: An integrative review. *Reading Research Quarterly*, 45(4), 464-487. doi:10.1598/RRQ.45.4.5

Carter, E., Schönbrodt, F., Gervais, W. M., & Hilgard, J. (2018, June 4). Correcting for bias in psychology: A comparison of meta-analytic methods. Retrieved from <https://osf.io/rf3ys/>.

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

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37-46. doi:10.1177/001316446002000104

*Corvacho del Toro, I. (2016). Zur qualitativen Rechtschreibfehleranalyse und einer schriftsystematischen lernförderlichen Behandlung der Rechtschreibstörung. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, 44(5), 1-13. doi: 10.1024/1422-4917/a000457

- *Darch, C., Eaves, R. C., Crowe, D. A., Simmons, K., & Conniff, A. (2006). Teaching spelling to students with learning disabilities: A comparison of rule-based strategies versus traditional instruction. *Journal of Direct Instruction*, 6(1), 1-16.
doi:10.2307/1511340
- Del Re, A. (2013). compute. es: Compute effect sizes. *R package version*, 0.2-2.
- Dymock, S., & Nicholson, T. (2017). To what extent does children's spelling improve as a result of learning words with the look, say, cover, write, check, fix strategy compared with phonological spelling strategies? *Australian Journal of Learning Difficulties*, 22(2), 171-187. doi:10.1080/19404158.2017.1398766
- Ecalte, J., Magnan, A., Bouchafa, H., & Gombert, J. E. (2009). Computer-based training with ortho-phonological units in dyslexic children: New investigations. *Dyslexia: An International Journal of Research and Practice*, 15(3), 218-238.
doi:10.1002/dys.v15:310.1002/dys.373
- Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, 315(7109), 629-634.
doi:10.1136/bmj.315.7109.629
- Ehri, L. C. (1987). Learning to read and spell words. *Journal of Literacy Research*, 19(1), 5-31. doi:10.1080/10862968709547585
- Ehri, L. C. (2000). Learning to read and learning to spell: Two sides of a coin. *Topics in Language Disorders*, 20(3), 19-36. doi:10.1097/00011363-200020030-00005
- Ehri, L. C. (2005). Development of sight word reading: Phases and findings.
doi:10.1002/9780470757642.ch8
- Ehri, L. C. (2014). Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning. *Scientific Studies of Reading*, 18(1), 5-21.
doi:10.1080/10888438.2013.819356

- Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research*, 71(3), 393-447.
doi:10.3102/00346543071003393
- Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly*, 36(3), 250-287. doi:10.1598/rrq.36.3.2
- *Ehri, L. C., Satlow, E., Gaskins, I. (2009) Grapho-Phonemic Enrichment Strengthens Keyword Analogy Instruction for Struggling Young Readers. *Reading & Writing Quarterly*, 25(2-3), 162-191. doi: 10.1080/10573560802683549
- *Faber (2005). Systematische Rechtschreibförderung mit algorithmischen und selbstinstruktiven Lernhilfen: Die Leistungen von konzeptgemäß trainierten und untrainierten Schülern im Vergleich. *Heilpädagogische Forschung*(31), 129-144.
- *Fasting, R. B., & Halaas Lyster, S.-A. (2005). The effects of computer technology in assisting the development of literacy in young struggling readers and spellers. *European Journal of Special Needs Education*, 20(1), 21-40.
doi:10.1080/0885625042000319061
- *Ferraz, E., dos Santos Gonçalves, T., Freire, T., Mattar, T. d. L. F., Lamônica, D. A. C., Maximino, L. P., & Crenitte, P. A. P. (2018). Effects of a Phonological Reading and Writing Remediation Program in Students with Dyslexia: Intervention for Specific Learning Disabilities. *Folia Phoniatrica et Logopaedica*, 70, 59-73.
doi:10.1159/000489091
- Fisher, Z., & Tipton, E. (2015). Robumeta: Robust variance meta-regression. R package version 1.6.

- Frith, U. (1980). Unexpected spelling problems. In U. Frith (Ed.), *Cognitive processes in spelling* (pp. 495-515). London: Academic Press.
- Frith, U. (1985). Beneath the surface of developmental dyslexia. In K. Patterson, J. Marshall, & M. Coltheart (Eds.), *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading* (pp. 301–330). London: Erlbaum.
- Galuschka, K., Ise, E., Krick, K., & Schulte-Körne, G. (2014). Effectiveness of Treatment Approaches for Children and Adolescents with Reading Disabilities: A Meta-Analysis of Randomized Controlled Trials. *PLOS ONE*, 9(2), e89900.
doi:10.1371/journal.pone.0089900
- Gelman, A., & Carlin, J. (2014). Beyond Power Calculations Assessing Type S (Sign) and Type M (Magnitude) Errors. *Perspectives on Psychological Science*, 9(6), 641-651.
- Gelman, A., & Loken, E. (2014). The statistical crisis in science: Data-dependent analysis - a "garden of forking paths" - explains why many statistically significant comparisons don't hold up. *American Scientist*, 102, 460-465.
- *Given, B. K., Wasserman, J. D., Chari, S. A., Beattie, K., & Eden, G. F. (2008). A randomized, controlled study of computer-based intervention in middle school struggling readers. *Brain and language*, 106(2), 83-97.
doi:10.1016/j.bandl.2007.12.001
- Goodwin, A., & Ahn, S. (2010). A meta-analysis of morphological interventions: Effects on literacy achievement of children with literacy difficulties. *Annals of Dyslexia*, 60(2), 183-208. doi:10.1007/s11881-010-0041-x
- Goodwin, A., & Ahn, S. (2013). A meta-analysis of morphological interventions in English: Effects on literacy outcomes for school-age children. *Scientific Studies of Reading*, 17(4), 257-285. doi:10.1080/10888438.2012.689791

- Graham, S., Berninger, V. W., Abbott, R. D., Abbott, S. P., & Whitaker, D. (1997). Role of mechanics in composing of elementary school students: A new methodological approach. *Journal of Educational Psychology*, 89(1), 170. doi:10.1037/0022-0663.89.1.170
- Graham, S., Harris, K. R., & Adkins, M. (2018). The impact of supplemental handwriting and spelling instruction with first grade students who do not acquire transcription skills as rapidly as peers: a randomized control trial. *Reading and Writing*, 31(6), 1273-1294. doi:10.1007/s11145-018-9822-0
- Graham, S., Harris, K. R., & Chorzempa, B. F. (2002). Contribution of spelling instruction to the spelling, writing, and reading of poor spellers. *Journal of Educational Psychology*, 94(4), 669. doi:10.1037/0022-0663.94.4.669
- Graham, S., Harris, K. R., & Hebert, M. (2011). It is more than just the message: Presentation effects in scoring writing. *Focus on Exceptional Children*, 44(4), 1. doi:10.17161/fec.v44i4.6687
- Graham, S., & Hebert, M. (2011). Writing to read: A meta-analysis of the impact of writing and writing instruction on reading. *Harvard Educational Review*, 81(4), 710-744. doi:10.3102/0034654317746927
- Graham, S., & Santangelo, T. (2014). Does spelling instruction make students better spellers, readers, and writers? A meta-analytic review. *Reading and Writing*, 27(9), 1703-1743. doi:10.1007/s11145-014-9517-0
- Gray, S. (2015). *The effects of morpho-phonemic and whole word instruction on the literacy skills of adult struggling readers*. New York: City University of New York.
- *Groth, K., Hasko, S., Bruder, J., Kunze, S., & Schulte-Körne, G. (2013). Interventionseffekte bei Lese-Rechtschreibstörung: Evaluation von zwei

- Förderkonzepten unter besonderer Betrachtung methodischer Aspekte. *Lernen und Lernstörungen*, 2(3), 161-175. doi:10.1024/2235-0977/a000038
- *Gustafson, S., Ferreira, J., & Ronnberg, J. (2007). Phonological or Orthographic Training for Children with Phonological or Orthographic Decoding Deficits. *Dyslexia*, 13(3), 211-229. doi:10.1002/dys.339
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *J Educ Behav Stat*, 6(2), 107-128. doi:10.2307/1164588
- Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research synthesis methods*, 1(1), 39-65. doi:10.1002/jrsm.5
- Higgins, J. (2011). Cochrane handbook for systematic reviews of interventions. Version 5.1.0 [updated March 2011]. The Cochrane Collaboration. www.cochrane-handbook.org.
- Higgins, J., & Altman, D. (2008). Assessing risk of bias in included studies. *Cochrane Handbook for Systematic Reviews of Interventions: Cochrane Book Series*, 187-241. doi:10.1002/9780470712184.ch8
- Higgins, J., & Green, S. (2008). *Cochrane handbook for systematic reviews of interventions* (Vol. 5): Wiley Online Library.
- Hilte, M., & Reitsma, P. (2011). Effects of explicit rules in learning to spell open-and closed-syllable words. *Learning and instruction*, 21(1), 34-45. doi:10.1016/j.learninstruc.2009.10.002
- *Hughes, J. A., Phillips, G., & Reed, P. (2013). Brief exposure to a self-paced computer-based reading programme and how it impacts reading ability and behaviour problems. *PLOS ONE*, 8(11), e77867. doi:10.1371/journal.pone.0077867

- Ise, E., Engel, R. R., & Schulte-Körne, G. (2012). Effective treatment of dyslexia: A meta-analysis on the effectiveness of German-language intervention approaches. *Kindheit und Entwicklung, 21*(2), 122-136. doi:10.1026/0942-5403/a000077
- *Ise, E., Schulte-Körne, G. (2010) Spelling deficits in dyslexia: evaluation of an orthographic spelling training. *Annals of Dyslexia, 60*, 18-39. doi: 10.1007/s11881-010-0035-8
- Jackson, N. E., & Coltheart, M. (2013). *Routes to reading success and failure: Toward an integrated cognitive psychology of atypical reading*: Psychology Press.
- Jamshidifarsania, H., Garbayab, S., Limc, T., Blazevisa, P., & Ritchie, J. M. (2019). Technology-based reading intervention programs for elementary grades: An analytical review. *Computers & Education, 128*, 427-451. doi:10.1016/j.compedu.2018.10.003
- *Jensen, J., Lindgren, M., Andersson, K., Ingvar, D. H., & Levander, S. (2000). Cognitive intervention in unemployed individuals with reading and writing disabilities. *Applied neuropsychology, 7*(4), 223-236. doi:10.1207/S15324826AN0704_4
- Kast, M., Baschera, G.-M., Gross, M., Jäncke, L., & Meyer, M. (2011). Computer-based learning of spelling skills in children with and without dyslexia. *Annals of Dyslexia, 61*(2), 177-200. doi:10.1007/s11881-011-0052-2
- *Kast, M., Meyer, M., Vogeli, C., Gross, M., & Jancke, L. (2007). Computer-based multisensory learning in children with developmental dyslexia. *Restorative Neurology and Neuroscience, 25*(3), 355-370.
- Katz, L., & Frost, R. (1992). The reading process is different for different orthographies: The orthographic depth hypothesis. *Advances in Psychology, 94*, 67-67. doi:10.1016/S0166-4115(08)62789-2
- Kellogg, R. T. (2008). Training writing skills: A cognitive developmental perspective. *Journal of writing research, 1*(1), 1-26. doi:10.17239/jowr-2008.01.01.1

- Kirby, J. R., Deacon, S. H., Bowers, P. N., Izenberg, L., Wade-Woolley, L., & Parrila, R. (2012). Children's morphological awareness and reading ability. *Reading and Writing*, 25(2), 389-410. doi:10.1007/s11145-010-9276-5
- *Kirk, C., & Gillon, G. T. (2009). Integrated morphological awareness intervention as a tool for improving literacy. *Language, Speech, and Hearing Services in Schools*, 40(3), 341-351. doi:10.1044/0161-1461(2008/08-0009)
- *Lange, A. A., Mulhern, G., & Wylie, J. (2009). Proofreading using an assistive software homophone tool. *Journal of Learning Disabilities*, 42(4), 322-335. doi:10.1177/0022219408331035
- Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A definition of dyslexia. *Annals of Dyslexia*, 53, 1-14. doi:10.1007/s11881-003-0001-9
- Maughan, B., Messer, J., Collishaw, S., Pickles, A., Snowling, M., Yule, W., & Rutter, M. (2009). Persistence of literacy problems: spelling in adolescence and at mid-life. *Journal of child psychology and psychiatry, and allied disciplines*, 50(8), 893-901. doi:10.1111/j.1469-7610.2009.02079.x
- McArthur, G., Eve, P. M., Jones, K., Banales, E., Kohnen, S., Anandakumar, T., . . . Castles, A. (2012). Phonics training for English-speaking poor readers. *Cochrane Database Syst Rev*, 12, CD009115. doi:10.1002/14651858.CD009115.pub2
- McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. *Educational psychology review*, 8(3), 299-325. doi:10.1007/BF01464076
- McCutchen, D. (2000). Knowledge, processing, and working memory: Implications for a theory of writing. *Educational Psychologist*, 35(1), 13-23. doi:10.1207/S15326985EP3501_3

- McCutchen, D., & Logan, B. (2011). Inside incidental word learning: Children's strategic use of morphological information to infer word meanings. *Reading Research Quarterly*, 46(4), 334-349. doi:10.1002/RRQ.003
- McCutchen, D., Stull, S., Herrera, B. L., Lotas, S., & Evans, S. (2014). Putting Words to Work: Effects of Morphological Instruction on Children's Writing. *Journal of Learning Disabilities*, 47(1), 86-97. doi:10.1177/0022219413509969
- McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for publication bias in meta-analysis: An evaluation of selection methods and some cautionary notes. *Perspectives on Psychological Science*, 11(5), 730-749. doi:10.1177/1745691616662243
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The-PRISMA-Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med*, 6(7), e1000097. doi:10.1371/journal.pmed.1000097
- Moll, K., Kunze, S., Neuhoﬀ, N., Bruder, J., & Schulte-Körne, G. (2014). Specific Learning Disorder: Prevalence and Gender Differences. *PLOS ONE*, 9(7), e103537.
- Moll, K., & Landerl, K. (2009). Double dissociation between reading and spelling deficits. *Scientific Studies of Reading*, 13(5), 359-382. doi:10.1080/10888430903162878
- Moll, K., Ramus, F., Bartling, J., Bruder, J., Kunze, S., Neuhoﬀ, N., . . . Lohvansuu, K. (2014). Cognitive mechanisms underlying reading and spelling development in five European orthographies. *Learning and instruction*, 29, 65-77. doi:10.1016/j.learninstruc.2013.09.003
- Morris, S. (2008). Estimating Effect Sizes From Pretest-Posttest-Control Group Designs. *Organizational Research Methods*, 11(2), 364-386. doi:1094428106291059
- Nagy, W., Berninger, V. W., Abbott, R. D., Vaughan, K., & Vermeulen, K. (2003). Relationship of Morphology and Other Language Skills in At-Risk Second-Grade

- Readers and At-Risk Fourth Grade Writers. *Journal of Educational Psychology*, 95(4), 730-742. doi:10.1037/0022-0663.95.4.730
- Network, S. I. G., Harbour, R. T., & Forsyth, L. (2008). *SIGN 50: a guideline developer's handbook*: Scottish Intercollegiate Guidelines Network.
- *O'Shaughnessy, T. E., & Swanson, H. L. (2000). A comparison of two reading interventions for children with reading disabilities. *Journal of Learning Disabilities*, (3), 257-277. Retrieved from <http://www.mrw.interscience.wiley.com/cochrane/clcentral/articles/419/CN-00492419/frame.html> doi:10.1177/002221940003300304
- *Oakland, T., Black, J. L., Stanford, G., Nussbaum, N. L., & Balise, R. R. (1998). An evaluation of the dyslexia training program: A multisensory method for promoting reading in students with reading disabilities. *Journal of Learning Disabilities*, 31(2), 140-147. doi:10.1177/002221949803100204
- Ouellette, G., Martin-Chang, S., & Rossi, M. (2017). Learning from our mistakes: Improvements in spelling lead to gains in reading speed. *Scientific Studies of Reading*, 21(4), 350-357. doi:10.1080/10888438.2017.1306064
- Perfetti, C. A., & Hogaboam, T. (1975). Relationship between single word decoding and reading comprehension skill. *Journal of Educational Psychology*, 67(4), 461. doi:10.1037/h0077013
- Perfetti, C., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189-213). Amsterdam: John Benjamin.
- Pustejovsky, J. (2016). clubSandwich: Cluster-Robust (Sandwich) Variance Estimators with Small-Sample Corrections. R package version 0.2. 1. [Computer software]. <https://cran.rproject.org/package=clubSandwich>.

- R Core Team. (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2014. Retrieved from <https://www.r-project.org/>.
- *Rashotte, C. A., MacPhee, K., & Torgesen, J. K. (2001). The effectiveness of a group reading instruction program with poor readers in multiple grades. *Learning Disability Quarterly*, 24(2), 119-134. doi:10.2307/1511068
- Rastle, K. (2018). The place of morphology in learning to read in English. *Cortex*, 116, 45-54. doi:10.1016/j.cortex.2018.02.008
- Rastle, K., & Taylor, J. (2018). Print-sound regularities are more important than print-meaning regularities in the initial stages of learning to read: Response to Bowers & Bowers (2018). *Quarterly Journal of Experimental Psychology*, 71(7), 1501-1505. doi:10.1177/1747021818775053
- *Reuter-Liehr, C. (1993). Behandlung der Lese-Rechtschreibschwäche nach der Grundschulzeit: Anwendung und Überprüfung eines Konzeptes. *Zeitschrift für Kinder-und Jugendpsychiatrie*, 21(3), 135-147.
- Saddler, B., & Graham, S. (2005). The effects of peer-assisted sentence-combining instruction on the writing performance of more and less skilled young writers. *Journal of Educational Psychology*, 97(1), 43. doi:10.1037/0022-0663.97.1.43
- Scammacca, N., Roberts, G., Vaughn, S., & Stuebing, K. (2015). A Meta-Analysis of Interventions for Struggling Readers in Grades 4-12: 1980-2011. *Journal of Learning Disabilities*, 48(4), 369-390. doi:10.1177/0022219413504995
- Schmalz, X. (Producer). (2018). Python Script for calculating onset entropy. Retrieved from <https://osf.io/8r6zf/>

- Schmalz, X., Marinus, E., Coltheart, M., & Castles, A. (2015). Getting to the bottom of orthographic depth. *Psychon Bull Rev*, 22(6), 1614-1629. doi:10.3758/s13423-015-0835-2
- Schmidt, F. L. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist*, 47(10), 1173.
- Schulte-Körne, G., Deimel, W., Hülsmann, J., Seidler, T., & Remschmidt, H. (2001). Das Marburger Rechtschreib-Training–Ergebnisse einer Kurzzeit-Intervention. *Zeitschrift für Kinder-und Jugendpsychiatrie und Psychotherapie*, 29(1), 7-15. doi:10.1024//1422-4917.29.1.7.
- Serrano, F., Genard, N., Sucena, A., Defior, S., Alegria, J., Mousty, P., . . . Seymour, P. H. (2011). Variations in reading and spelling acquisition in Portuguese, French and Spanish: A cross-linguistic comparison. *Journal of Portuguese Linguistics*, 10(1), 183-204. doi:10.5334/jpl.106
- Seymour, P. H., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *Br J Psychol*, 94(Pt 2), 143-174. doi:10.1348/000712603321661859
- Share, D. L. (1995). Phonological recoding and self-teaching: < i> sine qua non</i> of reading acquisition. *Cognition*, 55(2), 151-218. doi:10.1016/0010-0277(94)00645-2
- *Spencer, F., Snart, F., & Das, J. (1989). A process-based approach to the remediation of spelling in students with reading disabilities. *Alberta Journal of Educational Research*, Vol 35(4), 269-282.
- Sprenger-Charolles, L., Siegel, L. S., & Bonnet, P. (1998). Reading and spelling acquisition in French: the role of phonological mediation and orthographic factors. *Journal of Experimental Child Psychology*, 68(2), 134-165. doi:10.1006/jecp.1997.2422

- Squires, K. E., & Wolter, J. A. (2016). The effects of orthographic pattern intervention on spelling performance of students with reading disabilities: A best evidence synthesis. *Remedial and Special Education, 37*(6), 357-369. doi:10.1177/0741932516631115
- Strehlow, U., Haffner, J., Busch, G., Pfüller, U., Rellum, T., & Zerahn-Hartung, C. (1999). An Schwächen üben oder durch Stärken ausgleichen? Vergleich zweier Strategien in der Förderung von Kindern mit einer umschriebenen Lese-Rechtschreib-Schwäche. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie, 27*(2), 103-113. doi:10.1024//1422-4917.27.2.103
- Taylor, J., Davis, M. H., & Rastle, K. (2017). Comparing and validating methods of reading instruction using behavioural and neural findings in an artificial orthography. *Journal of Experimental Psychology: General, 146*(6), 826. doi:10.1037/xge0000301
- *Thaler, V., Landerl, K., & Reitsma, P. (2008). An evaluation of spelling pronunciations as a means of improving spelling of orthographic markers. *European Journal of Psychology of Education, 23*(1), 3-23. doi:10.1007/BF03173137
- Thalmann, S. (2014). Adaption Criteria for the Personalized Delivery of Learning Materials: A Multi-Stage Empirical Investigation. *Australasian Journal of Educational Technology, 30*(1), 45-60. doi:10.14742/ajet.235
- *Tijms, J. (2011) Effectiveness of computer-based treatment for dyslexia in a clinical care setting: Outcomes and moderators. *Educational Psychology, 31*(7), 873-896. doi: 10.1080/01443410.2011.621403
- Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods, 20*(3), 375. doi: 10.1037/met0000011
- Treiman, R. (2018). Teaching and Learning Spelling. *Child Development Perspectives, 0*(0), 1-5. doi:10.1111/cdep.12292

- Treiman, R., & Bourassa, D. C. (2000). The development of spelling skill. *Topics in Language Disorders*, 20(3), 1-18. doi:10.1097/00011363-200020030-00004
- Treiman, R., & Kessler, B. (2014). How children learn to spell words: New York, NY: Oxford University Press.
- Truong, H. M. (2015). Integrating learning styles and adaptive e-learning system: Current developments, problems and opportunities. *Computers in Human Behavior*, 55, 1185-1193. doi:10.1016/j.chb.2015.02.014
- *Vadasy, P., Sanders, E., & Peyton, J. (2005). Relative effectiveness of reading practice or word-level instruction in supplemental tutoring: how text matters. *J Learn Disabil.*, 38(4), 364-380. doi:10.1177/00222194050380041401
- Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: sensitivity analysis using a priori weight functions. *Psychological Methods*, 10(4), 428. doi:10.1037/1082-989X.10.4.428
- Viechtbauer, W. (2010). Metafor: meta-analysis package for R. *R package version*, 2010, 1-0.
- Volkmer, S., Schulte-Körne, G., & Galuschka, K. (2019). Die Rolle der morphologischen Bewusstheit bei Lese-und Rechtschreibleistungen. *Zeitschrift für Kinder-und Jugendpsychiatrie und Psychotherapie*, 47(4), 334-344. doi:10.1024/1422-4917/a000652
- *Walter, J., Schliebe, L., & Barzen, S. (2007). Evaluation eines morphemorientiert-strategischen Rechtschreibtrainings in schulischen Fördergruppen mit Grundschulern der 3. *Heilpädagogische Forschung*, 13(14), 3-153.
- Wanzek, J., Vaughn, S., Scammacca, N., Gatlin, B., Walker, M. A., & Capin, P. (2016). Meta-analyses of the effects of tier 2 type reading interventions in grades K-3. *Educational psychology review*, 28(3), 551-576. doi:10.1007/s10648-015-9321-7

- Wanzek, J., Vaughn, S., Wexler, J., Swanson, E. A., Edmonds, M., & Kim, A.-H. (2006). A synthesis of spelling and reading interventions and their effects on the spelling outcomes of students with LD. *Journal of Learning Disabilities*, 39(6), 528-543. doi: 10.1177/00222194060390060501
- Wanzek, J., & Voughn, S. (2008). Response to varying amounts of time in reading intervention for students with low response to intervention. *J Learn Disabil.*, 41(2), 126-142. doi:10.1177/0022219407313426
- *Weber, J.-M., Marx, P., & Schneider, W. (2002). Profitieren Legastheniker und allgemein lese-rechtschreibschwache Kinder in unterschiedlichem Ausmaß von einem Rechtschreibtraining? *Psychologie in Erziehung und Unterricht*, 49(1), 56-70.
- Wechsler, D. (1992). *Wechsler Individual Achievement Test: Manual*. San Antonio: TX: Psychological Corportation.
- Weiser, B., & Mathes, P. (2011). Using encoding instruction to improve the reading and spelling performances of elementary students at risk for literacy difficulties: A best-evidence synthesis. *Review of Educational Research*, 81(2), 170-200. doi:10.3102/0034654310396719
- Wilkinson, G. (1993). *The Wide Range Achievement Test* (3rd ed.). Wilmington: Wide Range, Inc.
- Williams, C., & Beam, S. (2019). Technology and writing: Review of research. *Computers & education*, 128, 227-242. doi: 10.1016/j.compedu.2018.09.024
- Williams, K. J., Walker, M. A., Vaughn, S., & Wanzek, J. (2017). A Synthesis of Reading and Spelling Interventions and Their Effects on Spelling Outcomes for Students With Learning Disabilities. *Journal of Learning Disabilities*, 50(3), 286-297. doi:10.1177/0022219415619753

List of Tables

Table 1. Mean effect sizes for spelling interventions on reading and spelling performance.

Table 2. Efficacy of treatment approaches on reading and spelling outcomes.

Table 3. Meta-Regressions for Onset Entropy, Severity, Age and Amount of the Intervention for the whole study pool combined for reading and spelling outcomes.

Table 4. Main Effects and Interactions of Age and Severity for phonics, orthographic and morphological interventions

Table 5. Exploratory moderator analyses combined for reading and spelling outcomes.

List of figures

Figure 1. Flow chart of study selection process.

Figure 2. Scatterplot for effect sizes vs. age and severity grouped by treatment approach.

List of supplements

Supplement 1. Study details.

Supplement 2. Severity assessment of reading and spelling disorder.

Supplement 3. Quality assessment.

Supplement 4. Forest-Plot.

Supplement 5. Weight function for publication bias.

Tables

Table 1

Mean effect sizes for spelling interventions on reading and spelling performance.

	g', 95% CI	SE	p	I²	n/k
Reading/spelling combined	0.58; [0.40, 0.76]	0.09	< .001	67 %	31 / 107
Spelling	0.68; [0.45, 0.92]	0.12	< .001	75 %	31 / 35
Reading	0.42; [0.27, 0.57]	0.07	< .001	54 %	23 / 72

Note. n = number of experimental and control group comparisons; k = number of outcomes.

Table 2

Efficacy of treatment approaches on reading and spelling outcomes.

Treatment approach	Outcome	g' , 95% CI	SE	p	I^2	n/k
Phonics	Spelling	0.68; [0.15, 1.21]	0.23	.017	86%	10 / 14
	Reading	0.62; [0.37, 0.86]	0.10	< .001	67%	10 / 44
Morphological intervention	Spelling	0.80; [0.39, 1.20]	0.21	< .001	51 %	8 / 8
	Reading	0.30; [-0.18, 0.77]	0.16	.150	27 %	5 / 10
Orthographic intervention	Spelling	0.67; [0.05, 1.28]	0.31	.034	83 %	6 / 6
	Reading	0.19; [0.09, 0.29]	0.02	.014	0 %	4 / 12
Memorization intervention	Spelling	0.22; [-0.39, 0.82]	0.31	.484	0 %	2 / 2
Supportive software	Spelling	1.25; [0.83, 1.66]	0.21	< .001	0 %	2 / 2
	Reading	0.52; [-5.1, 6.14]	0.44	.449	77 %	2 / 3

Note. n = number of experimental and control group comparisons; k = number of outcomes.

Table 3

Meta-Regressions for Onset Entropy, Severity, Age and Amount of the Intervention for the whole study pool combined for reading and spelling outcomes.

Moderator	<i>B</i>	SE	<i>p</i>	n/k
Severity	-0.004	0.16	.983	27/95
Age	-0.01	0.01	.502	31/107
Onset Entropy	-1.234	1.329	.363	30/101
Amount	0.00	0.00	.984	30/101

Note. n = number of experimental and control group comparisons; k = number of outcomes.

Table 4

Main Effects and Interactions of Age and Severity for phonics, orthographic and morphological interventions

	Model Age				Model Severity			
	<i>B</i>	SE	<i>df</i>	<i>p</i>	<i>B</i>	SE	<i>df</i>	<i>p</i>
Intercept	1.59	0.73	5.54	.077	1.19	0.34	3.73	.029
Morphological intervention	-1.15	1.03	4.10	.327	-1.12	0.70	5.01	.169
Orthographic intervention	-0.57	2.54	3.35	.836	-0.90	0.49	3.87	.147
Age	-0.11	0.08	4.72	.210				
Morphological intervention x Age	0.13	0.10	4.52	.260				
Orthographic intervention x Age	0.07	0.22	4.37	.768				
Severity					0.41	0.19	1.79	.171
Morphological intervention x Severity					-0.90	0.50	4.41	.141
Orthographic intervention x Severity					-0.63	0.38	3.16	.189

Table 5

Exploratory moderator analyses combined for reading and spelling outcomes.

Moderator	Level	g' ; 95 % CI	B	F	df	p	n/k
Treatment approach	Phonics (intercept)	0.63; [0.33, 0.92]	0.65	0.54	2.1	.730	10 / 58
	Morphological intervention	0.62; [0.22, 1.02]	0.01				8 / 18
	Orthographic intervention	0.61; [-0.24, 1.46]	-0.05				6 / 18
	Memorization intervention	0.22; [-0.39, 0.82]	-0.44				2 / 2
	Supportive software	0.86; [-1.63, 3.36]	0.21				2 / 5
Language	English (intercept)	0.56; [0.36, 0.76]	0.56	0.46	14.7	.510	15 / 69
	German	0.81; [0.13, 1.49]	0.21				9 / 14
	Other	0.48; [0.13, 0.82]	-0.10				7 / 24
Setting	Class (intercept)	0.28; [0.02, 0.58]	0.28	3.45	16.2	.061	8 / 26
	Group	0.65; [0.43, 0.86]	0.37				13 / 50
	Individual	0.81; [0.29, 1.33]	0.52				9 / 25
Computer	Computer-assisted (intercept)	0.54; [-1.44, 2.52]	0.61	0.67	2.14	.60	2 / 3

Implementer	Computer-based	0.44; [0.15, 0.73]	-0.19				9 / 23
	no computer	0.68; [0.40, 0.95]	0.05				20 / 81
	Mixed (intercept)	0.49; [0.22, 0.78]	0.50	0.30	13.9	.743	7 / 38
	Paraprofessionals	0.56; [0.30, 0.82]	0.12				8 / 17
	Professionals	0.63; [0.27, 0.98]	0.12				15 / 46
Grade	Grade 1-2	0.75; [-0.17, 1.33]	0.71	0.96	8.43	.422	4 / 21
	Grade 2-4	0.44; [0.21, 0.66]	-0.27				12 / 37
	> grade 4	0.75; [0.26, 1.24]	0.03				11 / 35

Note. Table comprises *F*, df, and *p*-values from a Wald-type tests of linear contrasts from a fitted linear regression model, using a sandwich estimator for the variance-covariance matrix and a small sample correction for the *p*-value.

n = number of experimental and control group comparisons; k = number of outcomes.

Figures

Figure 1. Flowchart of study selection process.

Figure 2. Scatterplot for effect sizes vs. age and severity grouped by treatment approach.