Spelling performance in children with Developmental Language Disorder: A meta-analysis across European languages

Nelly Joye*[1], Lucie Broc[2], Thierry Olive[3] & Julie Dockrell[1]

[1] UCL Institute of Education, UK

[2] University Côte d'Azur & CNRS, France

[3] CNRS & University of Poitiers, France

*Corresponding author, nelly.joye.14@ucl.ac.uk, Centre for Language, Literacy and Numeracy: Research and Practice (LLNR&P), Psychology and Human Development, UCL Institute of Education, 25 Woburn Square, London WC1H 0AA, UK

Abstract

Spelling difficulties often occur in children with Developmental Language Disorder (DLD), the extent and drivers of these difficulties are underspecified. Meta-analyses were conducted to address this gap.

We identified child-based factors (age, language spoken, profile of difficulties) and task-based factors (text production or dictation) as potential moderators of the impact of DLD on spelling performance.

32 studies were analyzed, including 984 children with DLD. Large differences between the spelling of children with DLD and their age-matched peers were found. The average mean effect size was -1.42, (95%CI [-1.60, -1.24]). Heterogeneity was large. Effect size estimates were larger when participants had additional phonological or reading difficulties than when they did not. No differences were found between participants with DLD and their language-matched peers.

The results are discussed with respect to the underlying processes that impact on spelling across languages. Specific recommendations for future studies are made.

Keywords: Developmental Language Disorder, spelling, meta-analysis

Introduction

Spelling difficulties are often associated to Developmental Language Disorder (DLD), and are argued to provide a window into residual language deficits (Bishop & Clarkson, 2003). It is currently unclear whether all children with DLD encounter such difficulties (McCarthy, Hogan & Catts, 2012), whether reported difficulties reflect children's language levels (Mackie & Dockrell, 2004), are influenced by the language spoken (Broc et al., 2013) and the extent spelling difficulties are explained by other co-occurring problems or the spelling task completed (Dyslexia, McCarthy et al., 2012; Phonology, Bishop & Clarkson, 2003). There is a complex relationship between oral and written language difficulties, which may change over development, and is not easily captured by a single deficit model (Brizzolara et al., 2011). However, our understanding is challenged by inconsistencies across studies and the potential impacts of co-occurring difficulties. To address these limitations, we conducted a meta-analysis to test and extend conclusions drawn from studies with small numbers of participants, differing language profiles and diverse sets of co-occurring problems. Meta-analyses provide a theory-neutral framework to examine the relevant research studies. To our knowledge, this is the first attempt to systematically review and assess this literature using a meta-analysis. We aim to provide a rigorous basis to inform research protocols and develop our conceptualization of spelling problems in children with DLD.

Children with Developmental Language Disorders (DLD) experience problems with the structural dimensions of language, including grammar and vocabulary. In the present meta-analysis, we followed the terminology reached in the Delphi study of Bishop et al. (2017) and identified studies where participants met the criteria for DLD, although the diagnostic labels in the studies varied (Bishop, 2014; Reilly, Bishop & Tomblin, 2014). There are a number of reasons to predict that children with DLD will experience problems with spelling. Strong links exist between oral language skills and spelling ability in typically developing children (Kim, 2010; Muter & Snowling, 1998; Nagy, Berninger, & Abbott, 2006; Ouellette & Sénéchal, 2008; Tong, McBride-Chang, Shu, & Wong, 2009). Spelling draws on both the phonological and morphological features of the language (Nagy et al., 2006), skills at risk in DLD, although the relative importance of these linguistic factors may differ across development and languages (Caravolas, 2004).

Children with DLD often experience co-occurring difficulties with literacy, including decoding, reading comprehension and the production of written text (Botting, Simkin, & Conti-Ramsden, 2006; Catts, Bridges, Little, & Tomblin, 2008; Catts, Fey, Tomblin, & Zhang, 2002; Dockrell, Lindsay, & Connelly, 2009; Dockrell, Lindsay, Connelly, & Mackie, 2007; Scott & Windsor, 2000; Windsor, Scott, & Street, 2000). A specific area of weakness is in spelling performance, where students are reported to perform poorly relative to age-matched peers, produce atypical errors, and experience continued difficulties into adolescence (Bishop & Clarkson, 2003; Connelly, Dockrell, Walter, & Critten, 2012; Mackie & Dockrell, 2004). Despite an increasing number of studies exploring spelling in children with DLD, the extent and locus of their difficulties remain underspecified.

There are several challenges in interpreting the results of current studies. The magnitude of the difference in scores between children with DLD and age-matched typically-developing (TD) peers is rarely reported and the studies typically rely on small samples. Furthermore, there is a lack of consistency in how participants are identified, limiting generalizations. Differences are evident in the age of participants, the language in which their spelling is assessed, the spelling task used and the extent to which co-occurring problems are reported. Meta-analyses provide the possibility of synthesizing findings across studies (Hedges & Cooper, 1994), giving a more accurate indication of effect size (Button et al., 2013).

There is also uncertainty as to whether spelling difficulties are to be expected given the children's language skills or whether spelling errors are atypical, reflecting specific features of DLD (Bishop & Clarkson, 2003; Broc et al., 2014; Critten, Connelly, Dockrell, & Walter, 2014; Nauclér, 2004; Soriano-Ferrer & Contreras-González, 2012; Windsor et al., 2000). Some studies have compared children with DLD to younger children matched on language levels, suggesting that children with DLD spell in a manner commensurate with younger language-matched peers (Mackie & Dockrell, 2004; Mackie, Dockrell, & Lindsay, 2013). Other studies have highlighted greater problems with morphological spelling, arguably reflecting the children's specific difficulties with inflectional morphology (Critten et al., 2014). Studies that compare the spelling of children with DLD and language-matched peers allow for a more nuanced understanding of the nature of the children's spelling difficulties and their causal mechanisms.

The spelling performance of children with DLD likely varies with age. To date, few studies have reported longitudinal data, so developmental trajectories of spelling skills remain underexplored. In English, Dockrell et al. (2007; 2009) followed a cohort of children with DLD from eight to 16, assessing the impact of spelling abilities on their written text. Longitudinal performance in spelling was not reported. Broc et al. (2013) compared the spelling performance of French children with DLD to age-matched TD peers in two groups of participants aged 7-11 and 12-18. In the younger, but not the older group, there was a significant difference in spelling performance between the DLD and TD groups. The authors suggested that spelling difficulties may reduce with age, however these conclusions were drawn from cross-sectional data. To our knowledge, no other study has compared the spelling performance of different age groups of participants with DLD and no longitudinal studies have been reported. This developmental perspective is important as the factors which underpin spelling performance change as children become more competent spellers (Treiman, 2017).

There is also evidence that spelling difficulties may differ across languages. Studies of reading acquisition demonstrate that rates vary between children learning deep and shallow orthographies. In English, considered to be a deep orthography, the rate of development is twice as slow for reading as in more shallow orthographies (Seymour, Aro, & Erskine, 2003). Similarly, Wimmer and Landerl (1997) observed faster spelling development rates in German, considered a shallower orthography than English. Furthermore, the inconsistency of orthographies is usually stronger from phonology to spelling than from spelling to phonology (see Ziegler, Jacobs, & Stone, 1996 and Ziegler, Stone, & Jacobs, 1997). In a recent comparison of the spelling accuracy of English- and Italian-speaking TD children in grades 2-5, Marinelli, Romani, Burani and Zoccolotti (2015) demonstrated both faster rates of spelling development in Italian and more persistent cross-linguistic gaps in spelling than in reading accuracy. As such, phonology to spelling depth is likely to affect spelling performance of children with DLD.

Variation in the tasks used to assess spelling further challenges our understanding of the links between DLD and spelling. McCarthy et al. (2012) suggested that children whose spelling was assessed in a task of written text production, where they choose the words they spell, might put them at an advantage, compared to a task of word dictation, where word choice is constrained. Broc et al. (2013) assessed this hypothesis by comparing the performance of children with DLD in a task of written text production and in a task of word dictation, using the same metrics (a proportion of spelling errors per word). There were significantly larger differences between children with DLD and their peers in dictation than in written text production. These results suggest that task differences across studies may influence spelling performance (see Sumner, 2013, p.100 for a similar argument for dyslexia).

The occurrence of expressive phonological difficulties is common in DLD (Bird, Bishop, & Freeman, 1995; Bishop & Clarkson, 2003; Brizzolara et al., 2011; Lewis, O'Donnell, Freebairn, & Taylor, 1998). Expressive phonological difficulties are characterized by errors in the production of speech sounds, that persist beyond the expected age, and/or that are not typically observed during development. Phonological skills affect spelling. The relationship found between phonological awareness and spelling is relatively uncontested and may mediate the relationship between oral language and spelling across languages (Moll et al., 2014). Difficulties with expressive phonology play a role in the development of phonological awareness (Bird & Bishop, 1992; Bird et al., 1995; Bishop & Snowling, 2004; Fraser, Goswami, & Conti-Ramsden, 2010). However, the evidence establishing links between expressive phonological difficulties and literacy difficulties is mixed. The co-occurrence of expressive phonological difficulties and literacy difficulties ranges between 25-30% (Peterson, Pennington, Shriberg, & Boada, 2009). However, to date, studies have focused on reading, but not spelling. Arguably, phonological skills will be more important in applying the alphabetic principle when it comes to producing rather than recognizing a written word (Bird et al., 1995).

A further challenge in understanding the association between DLD and spelling is to isolate the differential effects of language and reading on spelling performance. Reading difficulties often co-occur with language disorder (Catts et al., 2002; Catts, 1993). Given the close developmental relationship between reading and spelling (Swanson, Trainin, Necoechea, & Hammill, 2003; Zutell & Rasinski, 1989) this is likely to affect children with DLD. Large correlations (> .7) between concurrent measures of reading and spelling are consistently reported in the literature (Dockrell et al., 2009; Vandewalle et al., 2012; Wolter, Self, & Apel, 2011). Recent studies have suggested that it is reading skills, not oral language, that predict spelling in children with DLD (Mackie et al., 2013; McCarthy et al., 2012).

To develop a more comprehensive understanding of the spelling difficulties experienced by children with DLD and the mechanisms that underpin these problems, a systematic search of the literature was completed and a meta-analysis was used to pool results from studies and compute an overall effect size. We examined both child- and task- related factors that have been reported to influence children's spelling performance. We addressed two questions. The first focused on developmental patterns of spelling performance and the second on the impact of phonological and reading profiles. To examine the first research question we explored differences in spelling performance between children with DLD and chronological age-matched peers to establish whether differences in spelling performance were moderated by child factors (age and language spoken) or task factors. We then examined whether the same differences of performance were evident when comparisons were made with language-matched peers. To address the second question, we examined whether difficulties with phonology and reading influenced spelling performance. We considered both phonology and reading skills within the profiles of children with DLD.

We reasoned that both task and language would have significant effects on the spelling differences of children with DLD in comparison to both chronological age and language-matched peers. We further predicted that expressive phonological difficulties alone would affect spelling. Finally, we predicted that co-occurring difficulties with reading would affect the severity of the children's spelling difficulties.

Method

The present meta-analysis was conducted and reported following the guidance of the PRISMA statement (Moher, Liberati, Tetzlaff, Altman, & the PRISMA group, 2009) and of the EPPI-centre (Gough, Oliver, & Thomas, 2012; 2013).

Selection criteria

To address the research questions, we selected empirical studies meeting the following criteria:

1. Children with DLD were included and DLD was assessed by a standardized language test relevant to the children's context, or previously ascertained by a clinician/professional responsible for diagnosing DLD in the country of the study. Because the terminology for DLD varies across countries and contexts, a wide range of diagnostic terms was accepted (see Appendix A).
2. The spelling performance of children with DLD was compared to that of a group of children with typical language development (TD), matched on either chronological age or language levels.
3. Participants aged between 5 and 18.
4. Sufficient data for effect size calculation were reported. Studies without sufficient data were excluded, after checking if missing data were available from the authors.
5. Children included were assessed in European languages, defined broadly as the languages spoken in the 37 states of the European Cooperation in Science and Technology (COST, 2017), and represented by a morphophonemic alphabetic system. Non-alphabetic systems were excluded for comparability issues and there were too few studies on non-European alphabetic to be included.

6. The language in which children were assessed was the language of instruction.

7. The language in which the study was reported could be any of the European languages included in the study.

Location and selection of studies

The selection process is summarized in Figure 1 below and described below.

Please insert figure 1

Databases. Studies were identified using psychological, educational, medical and general databases identified in previous meta-analyses (Torgerson & Elbourne, 2002; Vugs, Cuperus, Hendriks, & Verhoeven, 2013), and by our library services. In order to recruit grey literature or theses, the database Opengrey was also searched. All relevant national databases were also searched, using the terminology in the corresponding language. Databases searched are listed in Appendix A.

Boolean search phrase. A Boolean search was used in each database, in order to identify relevant records. The Boolean terms included a range of terms used to refer to DLD and spelling, with the aim of capturing the widest range of studies possible. It excluded terms that identified studies with adults, acquired language difficulties, or other known syndromes affecting language. The search terms are presented in Appendix A.

Selection of studies. Our inclusion criteria were applied at the three stages of study selection: screening of titles, abstracts, and full texts. The bibliography of the included studies was searched for other references.

Reliability of study selection. The first two authors independently screened 10% of the titles (121 titles out of the 1188) in order to ensure the reliability of study selection. Agreement was reached in 94% of cases. The remaining seven titles were given to the fourth author for a third rating and discussion until agreement was reached on inclusion or exclusion. The first two authors conducted the screening of abstracts and full texts jointly. The fourth author completed a consistency check.

Data extraction and coding

Data extraction. Data from the 53 studies that were initially selected were entered into a spreadsheet, which followed guidance from the PRISMA statement (Moher et al., 2009) and included general information about each study (authors, date, language and country of study), identified the study aims, characteristics of the population recruited, methodological features for the spelling measure and summarized the results. The data necessary for the computation of effect sizes were extracted at this stage: the mean, standard deviation and sample size of at least one group of children with DLD and a group of controls. Studies were subdivided when they contained more than one set of such data. It was also at this stage that missing data were identified and 21 studies were further excluded. This data extraction yielded a total of 64 research findings from 32 studies. Appendix B summarizes the characteristics of each of the studies retained for analysis.

Coding of the data. The research findings were further coded for the language of study, the type of spelling task (Word Dictation -WD- or Text Production -TP), the age of participants in the DLD group, the matching measure for the control group (age, language level), and individual characteristics of the groups (presence or absence of phonological impairment - PI - or reading impairment - RI). Details are provided in Appendix C.

Quality appraisal. The data extraction grid also comprised a quality appraisal section, based on the Critical Appraisal Skills Programme Checklist for cohort studies (Critical Appraisal Skills Programme, 2013). The appraisal assessed the relevance of the methods for the research questions (including type of task, items spelled and scoring used). Potential biases such as recruitment (including target setting, screening procedure for determining DLD group inclusion - if any) and matching biases (age-matching, ability-matching, from same setting or not) (Gough et al., 2012) were also examined. The quality appraisal grid and study specifications/limitations can be found in Appendix D.

Computation of the effect sizes. Standardized mean differences and their variance were calculated for each instance where a spelling score was reported for children with DLD and a control group, using the formulae for Hedges' *g* (see Borenstein et al., 2009, pp.27-28; Lipsey & Wilson, 2001), including a correction for small sample size (Borenstein et al., 2009; Lipsey & Wilson, 2001). In the current meta-analysis, the mean of the group with DLD was always subtracted to the mean of the TD group, resulting in negative *g*-values. *g*-values furthest from zero indicated the largest group differences. For consistency in the meta-analysis, when scores were expressed in number of errors rather than number of correct responses, the mean scores were inverted (e.g. $M = 1$ becoming $M = -1$), as recommended by Lipsey and Wilson (2001).

Data analysis

The data were analysed using the metafor package (Viechtbauer, 2010) for R (R Core Team, 2017). The analysis consisted of a series of individual meta-analyses to address the research questions, as detailed below.

Developmental patterns. To measure the magnitude of the effect of DLD on spelling performance, results from studies comparing the spelling of children with DLD and age-matched TD peers were computed. When a study reported the results of more than one group of children with DLD (e.g. DLD+RI and DLDnoRI), their mean scores were combined to be compared to the TD group (see Borenstein et al., 2009, pp.221-222). Similarly, when studies reported both

WD and TP for the same samples of children, these were combined to obtain a mean effect size (see Borenstein et al., 2009, pp.227-228). The resulting effect sizes were entered into a random-effects model. The moderating effects of age and language were then explored using two meta-regressions: one with age as a moderator, and one with language as a moderator. Secondly, in an attempt to identify task-related variability, a subgroup analysis was conducted. Overall effect sizes were computed separately for the studies assessing spelling on word dictation and those assessing spelling on text production. The effect sizes obtained were compared using a *Z*-test, as described in Borenstein et al. (2009, p.168).

We further assessed differences in spelling scores between children with DLD and younger TD children matched on language. A random-effects model was applied to these five studies. All the studies in this sample assessed children's spelling in English, and all children with DLD were aged 9-10, so it was not possible to examine effects of language or age. In a second step, a summary effect size was computed separately for the two studies assessing children's spelling in WD and for the three studies assessing children's spelling in TP. These summary effect sizes were compared using a *Z*-test, to assess the effect of the task on difference in performance.

Co-occurring difficulties. To address our second research we first examined the impact of phonological impairment and, in separate analyses the impact of reading difficulties. For both analyses, age and language were considered as potential moderators of the effects. Where studies assessed both TP and WD, they were initially combined and subsequently examined separately to ascertain task effects.

For phonological difficulties, the spelling performance of children with an isolated phonological impairment (PInoDLD) and their age-matched TD peers was compared. Effect size estimates for studies comparing the spelling performance of children with phonological and other linguistic difficulties (DLD+PI) to those of children with PInoDLD was computed.

For reading difficulties, the spelling performance of children with an isolated DLD and no reading difficulties (DLDnoRI) and their age-matched TD peers was computed followed by analyses comparing the spelling performance of children with both DLD and reading difficulties (DLD+RI) and children with DLDnoRI.

Management of dependence. Following Lipsey and Wilson (2001), a definition of independence of results based on the population sample was chosen. Research findings were considered independent as long as two different experimental groups were assessed, within the same study or across studies. When necessary, a "shifting unit of analysis" approach (Borenstein et al., 2009) was used in order to manage dependence between research findings, whereby research findings were grouped by research questions, as described in the analysis section above.

Quantification and management of heterogeneity. Heterogeneity is the presence of variation in effect sizes in a sample of studies. Following Higgins (2008), I^2 was used as an indicator of heterogeneity, in addition to the Q statistic, its p-value, and tau-squared. It is expressed in percentage of total variability attributable to heterogeneity. Indicative I^2 values of 25%, 50% and 75% were used to benchmark the studies' heterogeneity as low, moderate or high (Higgins, Thompson, Deeks, & Altman, 2003).

Choice of a meta-analysis model. A random-effects model with a DerSimonian-Laird estimator was chosen for all analyses, with the assumption that the effect tested across studies was not unique. When data were available we examined the effect of age or language in the model, using a meta-regression. When possible, studies were also grouped by task and summary effects were compared between tasks using a Z-test by using separate estimates of the heterogeneity for each task (Borenstein et al., 2009, pp. 167-168). This procedure ensured dependent results across tasks were computed separately.

Outliers, small study size, and publication bias. For all analyses presented, a sensitivity analysis was conducted to examine the impact of potential outliers on the effect size estimates and heterogeneity using the “leav1out” function of the metafor package (Viechtbauer, 2010). Each study was removed from the model and the model was applied to all other studies. This was done iteratively for all studies in the model. We also accounted for publication bias in the analysis when possible. Specifically, when the meta-analysis involved more than five studies, small-study effects were checked using funnel plots with an Egger’s test for plot asymmetry (see Rothstein, Sutton and Borenstein, 2005, pp.75, 90-91). Each study’s effect size was plotted against their standard error. The “funnel” and “regtest” functions of the metafor package were used (Viechtbauer, 2010). When funnel plots and Egger’s test suggested asymmetry, effect size estimates were adjusted using the PEESE selection methods correction described by Carter, Schönbrodt, Gervais and Hilgard (pre-print). This correction model was chosen for its performance in meta-analyses with statistically-significant results and high between-study heterogeneity.

Results

Spelling difficulties in DLD and the impact of age, language and task

Thirty-one research findings (two studies combine due to same sample) were included to examine the magnitude of the effect of DLD on spelling performance (see table 1 in supplemental material). The random-effects model yielded a significant ($p < .0001$) and large effect size in favor of the TD group ($g = -1.42$, 95%CI [-1.60, -1.24]). Figure 2 presents the effect sizes and summary estimate of the model comparing the spelling scores of children with DLD and age-matched TD children across tasks and subgroups.

Please insert figure 2

However, heterogeneity was large (I^2 = 66.27%, $Q(30)$ = 88.94, $p < .0001$), indicating unidentified sources of variability. Neither age ($Q_{mod}(1)$ = .21, p = .65, β = -.01, I^2_{res} = 66.18%) nor language ($Q_{mod}(5)$ = 7.98, p = .16, β = [-.99; .53][1], I^2_{res} = 65.24%) accounted for a significant amount of heterogeneity. Analysis by language revealed very large and significant effect sizes across all languages except Russian, where only one study was found: Dutch (g = -1.20, 95% CI [-1.76, -.63], $p < .0001$), English (g = -1.41, 95% CI [-1.62, -1.20, $p < .0001$]), French (g = -1.40, 95% CI [-2.16, -.65], $p < .0001$), Italian (g = -2.40, 95% CI [-3.19, -1.61], $p < .0001$), Spanish (g = -1.56, 95% CI [-2.54, -.58], $p < .001$) and Russian (g = -0.87, 95% CI [-1.40, -.33]). Following the sensitivity analysis, the effect sizes (g = -1.39 to g = -1.45) and heterogeneity (I^2 = 51.10 to I^2 = 67.39) remained large and significant. No research finding had significant impact on the result. Visual inspection of the funnel plots and Egger's test indicated no asymmetry for smaller studies (z = -1.78, p = .07). Funnel plots are presented in supplemental material, as well as all graphical representations for the subsequent sub-analyses.

Effect of task. In this analysis 36 research findings were used, 26 in WD and 10 in TP (see table 2 in supplemental material). Summary estimates for WD and TP were not significantly different (Z_{diff} = -.82, 95% CI [-1,29, .53], p = .41). For studies assessing spelling with a WD task, the random-effects model yielded a significant ($p < .0001$) and large effect size in favor of

---

[1] *Given the anglocentricity of the sample, English was always defined as the constant/intercept (g = -1.41 [-1.61; -1.20], p < .0001). β-values ranged from -.99 for Italian ([-1.81; -0.18], z = -2.38, p = .02) to .53 for Russian ([-0.41; 1.49], z = 1.12, p = .26), with β-values of -.15 for Spanish ([-1.16; 0.85], z = -.30, p = .76), .004 for French ([-0.78; 0.79], z = .01, p = .99) and .21 for Dutch ([-0.39; 0.81], z = .69, p = .49).*

the TD group (*g* = -1.53, 95% CI [-1.75, -1.32]). Heterogeneity of the sample was large (I^2 = 72.90%) and the test for heterogeneity was significant ($Q(25) = 92.25$, $p < .0001$). For studies assessing spelling in a text production task, there was also a significant ($p < .001$) and large effect size in favor of the TD group (*g* = -1.15, 95% CI [-1.36, -.95]). In this case heterogeneity was small (I^2 = 0%) and the test for heterogeneity was not significant ($Q(9) = 5.42$, $p = .80$).

Across tasks, children with DLD experienced significant difficulties with spelling. Age and language were not significant moderators of the effect sizes observed for WD ($Q_{mod\text{-}age}(1) = .97$, $\beta = -.04$, $p = .32$ and $Q_{mod\text{-}lang}(4) = 6.77$, β = [-.85; .68][2], $p = .15$) or TP ($Q_{mod\text{-}age}(1) = .76$, $\beta = .04$, $p = .38$ and $Q_{mod\text{-}lang}(2) = 2.24$, β = [-.48; -.19][3], $p = .32$). The sensitivity analysis demonstrated that no research finding significantly affected the significance and size of the effect and the heterogeneity of the sample (WD effect size range *g* = -1.46 to *g* = -1.58, heterogeneity range I^2 = 62.86% to I^2 = 73.93%; TP effect size range *g* = -1.11 to *g* = -1.21, heterogeneity I^2 = 0%). Visual examinations of funnel plots and Egger's tests showed no asymmetry for TP ($z = .41$, $p = .68$). However, there was asymmetry in WD ($z = -2.40$, $p = .02$). After applying the PEESE correction for publication bias, effect sizes remained significant ($p < .001$) and large overall (*g* = -1.08, 95% CI [-1.49; -.067]).

Spelling performance differences in DLD children and language-matched controls

Four research findings were computed from studies comparing the spelling performance of children with DLD and younger language matched peers (see table 3 in supplemental

---

[2] *With English as the intercept (g = -1.55 [-1.80; -1.30], p < .0001), β-values ranged from -.85 in Italian ([-1.74; 0.04], z = -1.88, p = .06) to .68 in Russian, ([-0.37; 1.74], z = 1.27, p = .20) with .01 for French ([-0.84; 0.88], z = .04, p = .97) and .35 for Dutch ([-0.32; 1.02], z = 1.06, p = .30).*

[3] *With English as the intercept (g = -1.07 [-1.31; -.84], z = -8.94, p < .0001), β-values were -.48 in Spanish (-1.16; 0.18], z = -1.42, p = .16) and -.19 in French ([-0.77; 0.38], z = -.66, p = -1.42).*

material). The random-effects model was not significant (p = .27) and the effect size very small (g = -.20, 95% CI [-.54, .15]). Children with DLD performed at a level similar to their language-matched peers. The heterogeneity was moderate (I^2 = 37.88%) and the test for heterogeneity was not significant (Q(3) = 4.83, p = .18). No further analyses were computed as all studies were in English and sampled the same age range (9 to 11).

In order to assess task effects, difference in scores for the two research findings assessing children in WD and for the four studies assessing children in TP were computed separately (see table 4 in supplemental material). There was no significant difference (Z_{diff} = -.23, 95% CI [-1.73, 1.35], p = .82) in the summary effect sizes estimated separately for WD (g = -.24, 95% CI [-1.02, .55]) or TP (g = -.06, 95% CI [-.47, .36]). Separate summary estimates were not significant (WD p = .55, TP p = .78), confirming that across tasks, children with DLD were commensurate with language-matched peers in spelling.

The impact of the phonological impairment

To assess the unique impact of phonological difficulties on spelling performance, we computed results from four studies comparing children with an isolated phonological impairment and their age-matched peers (see table 5 in supplemental material). The random-effects model yielded a significant (p < .001) and moderate effect size (g = -.61, 95% CI [-1.18, -.16]). TD children had significantly higher spelling scores than children with an isolated PI and no difficulties in other language domains. The heterogeneity in this sample was moderate (I^2 = 48.88%), and the test for heterogeneity was not significant (Q(3) = 5.87, p = .12. When entered into the meta-regression, neither age (Q_{mod}(1) = .49, β = -.07, p = .48, I^2_{res} = 60.44%) nor language (Q_{mod}(1) = .94, $g_{English}$ = -.46, $\beta_{Italian}$ = -.69, p = .33, I^2_{res} = 53.27%) accounted for a significant amount of heterogeneity.

To assess potential task effects, the only study assessing children in TP was removed and a random-effects model was applied to the remaining research findings. Results were consistent to those obtained when including the TP finding. The summary effect size was still significant ($p < .001$) and moderate ($g = -.74$, 95% CI [-1.30, -.19]). Heterogeneity was slightly reduced (from $I^2 = 48.88\%$ to $I^2 = 46.57\%$) and the test for heterogeneity was still not significant ($Q(3) = 5.62$, $p = .13$). When entered into the meta-regression, age ($Q_{mod}(1) = .36$, $\beta = -.06$, $p = .55$, $I^2_{res} = 60.34\%$) and language ($Q_{mod}(1) = .52$, $g_{English} = -.61$, $\beta_{Italian} = -.54$, $p = .47$ $I^2_{res} = 58.22\%$) did not account for any significant amount of heterogeneity.

The impact of an isolated phonological impairment (PInoDLD) was contrasted to children with DLD who also experienced a phonological impairment (DLD+PI) (see table 6 in supplemental material). The random-effects model yielded a significant ($p < .01$) and large effect size estimate ($g = -1.18$, 95%CI [-1.81, -.54]). The heterogeneity in the studies' effect sizes was moderate ($I^2 = 51.71\%$) and the test for heterogeneity was not significant ($Q(3) = 6.21$, $p = .10$). Children with DLD+PI performed significantly worse on spelling tasks than children with PI alone. The tests for the moderators age ($Q_{mod}(1) = .02$, $\beta = .02$) and language ($Q_{mod}(1) = .04$, $g_{English} = -1.25$, $\beta_{Italian} = .18$) were not significant ($p = .90$ and $p = .84$ respectively) and the addition of these moderators in the models increased heterogeneity by 16% in both cases.

To assess potential task effects, the only study assessing children in TP was further removed and a random-effects model was applied to the remaining four effect sizes. The model was consistent with the results including the TP finding. The summary effect size was significant ($p < .0001$) and large ($g = -1.14$, 95% CI [-1.80, -.49]). Heterogeneity was increased (from $I^2 = 51.71\%$ to $I^2 = 54.37\%$) although the test for heterogeneity was still not significant ($Q(3) = 6.58$, $p = .09$). When entered into the meta-regression, age ($Q_{mod}(1) = .01$, $\beta = .01$, $p = .92$, $I^2_{res} =$

69.48%) and language ($Q_{mod}(1) = .02$, $g_{English} = -1.21$, $\beta_{Italian} = .14$, $p = .88$, $I^2_{res} = 69.55\%$) did not account for significant heterogeneity.

The impact of co-occurring reading difficulties

Five studies using WD compared performance between children with DLD and no reading difficulties with CA matched peers (see table 7 in supplemental material). The random-effects model was significant ($p < .001$) with a moderate effect size estimate ($g = -.65$, 95% CI [-1.13, .16]), in favor of age-matched TD controls. However, heterogeneity in the sample was very high ($I^2 = 76.58\%$, $Q(4) = 17.08$, $p = .002$). Neither age ($Q_{mod}(1) = 1.78$, $\beta = -0.32$, $p = .18$, $I^2_{res} = 73.40\%$), nor language ($Q_{mod}(1) = .49$, $g_{English} = -0.57$, $\beta_{Dutch} = -0.49$, $p = .48$, $I^2_{res} = 80.32\%$) were significant moderators. Because of the large heterogeneity of this sample and the suspected presence of outliers, each individual research finding was removed to see if it affected the overall effect size and heterogeneity. Excluding the study by Ramus et al. (2013) removed the majority of the heterogeneity in the sample ($I^2 = 4.44\%$, $Q(3) = 3.14$, $p = .37$), and affected the size but not the significance of the overall effect ($g = -.35$, 95% CI [-.59, -.11], $p < .01$). The removal of the other studies did not reduce heterogeneity.

The impact of reading impairment on the spelling performance of children with DLD

In the final analysis, the performance of children with a reading impairment and DLD was compared to children with a DLD but no reading impairment in five research findings using WD (see table 8 in supplemental material). The random-effects model was significant ($p < .001$) with a large summary estimate ($g = -1.72$, 95% CI [-2.28, -1.16]). Children with an isolated language disorder (but no reading impairment) performed significantly better than their peers with reading and language disorder on WD. The heterogeneity of the studies sample was large ($I^2 = 72.48\%$, $Q(4) = 14.54$, $p < .01$). Neither age ($Q_{mod}(1) = 2.90$, $\beta = -0.49$, $p = .08$, $I^2_{res} =$

64.92%), nor language ($Q_{mod}(1) = .15$, $g_{English} = -1.69$, $\beta_{Dutch} = -0.33$, $p = .70$, $I^2_{res} = 78.17\%$) accounted for a significant amount of the heterogeneity in the sample. Given the large heterogeneity of this sample and the presence of outliers, each individual research finding was removed to see if it affected the overall effect size and heterogeneity. Excluding the study by Ramus et al. (2013) eliminated all heterogeneity from the sample ($I^2 = 0\%$, $Q(3) = 2.00$, $p = .57$), and the effect size remained significant ($p < .001$) and large after removing this study ($g = -1.37$, 95% CI [-1.65, -1.08]).The removal of the other studies did not affect heterogeneity.

Discussion

Meta-analyses were used to evaluate evidence regarding the impact of child- and task-based factors on the spelling of children with DLD. As predicted, we found large effect sizes demonstrating poorer performance in children with DLD in comparison to age-matched peers but not language-matched peers. Heterogeneity was typically large and contrary to predictions, differences were not moderated by the language spoken, the children's age or the task completed. As predicted, expressive phonological difficulties alone resulted in significantly poorer performance, although spelling difficulties were more severe when other language skills were also affected. Reading difficulties also added to the severity of spelling difficulties in children with DLD, although spelling difficulties existed in children with DLD-alone. These results raise questions about the cognitive underpinnings of spelling difficulties in DLD.

Developmental patterns of spelling development in children with DLD

Children with DLD performed significantly poorer in spelling than their TD age-matched peers with very large ($g > 1.3$), despite significant heterogeneity. Our results point to large lags in spelling development for children with DLD and additionally highlight the significant variation in performance. This heterogeneity in performance demonstrates the importance of further explicating the problems that underpin spelling performance for these children.

By contrast, the spelling performance of children with DLD was commensurate with that of language-matched TD peers, suggesting a delay in spelling development in line with language development. Data for this analysis were only available for children aged between nine and 11, in a small number of English studies, raising questions about the generalizability of the findings. The spelling performance of children in these studies was equivalent to that of children aged around seven years, which, for English, reflects spelling primarily underpinned by phonology. At this point in development children apply their knowledge of sound-to-letters correspondence to represent a plausible spelling of the word (Treiman, 2017). When children are required to exploit other types of regularities, such as morphological or orthographic segments, to represent words accurately (Apel & Masterson, 2001) differences may be more evident. The absence of data from more transparent orthographic systems with a richer morphology than English is problematic in understanding the loci of the children's problems. It is likely that in transparent orthographies children rely earlier on morphological and orthographic segments. Systematic examination of spelling with language-matched samples at different points in development and in orthographies contrasted for morphological richness and orthographic transparency is needed. There is also an indication from the research literature that specific errors are made by children with DLD, which may not be produced by younger children matched on either spelling or language, and which are not made by aged-matched TD peers. For example, difficulties in spelling morphemes marking regular past tense, plural and progressive present are consistently reported in English (Mackie & Dockrell, 2004; Mackie et al., 2013; Silliman, Bahr & Peters, 2006; Windsor et al., 2000). Differences in the rate of phonologically inaccurate (Critten et al., 2014) and orthographically inaccurate (Mackie et al., 2013) spelling errors are also reported, although not consistently (Bishop & Clarkson, 2003; Mackie & Dockrell, 2004; Dockrell & Connelly, 2015; Silliman et al., 2006). Multicomponent frameworks for the analysis of spelling errors, such as the Phonological, Orthographic, and Morphological assessment of spelling (POMAS, Silliman, Bahr, & Peters, 2006; POMplexity, Quick & Erickson, 2018), or the coding system proposed by Apel and Masterson (2001) may prove useful tools for such qualitative analysis. Such systems go beyond phonological based error analysis, whose ability to pick up on specific patterns of spelling difficulties may be limited across development (Treiman, Kessler, Pollo, Byrne & Olson, 2016). Multicomponent frameworks allow for a finer-grained analysis to identify specific errors which may be produced by children with DLD. Such frameworks may be adapted and used in languages other than English, where data on the specific errors made by children with DLD are missing. Evaluation of those systems on larger samples of typical and atypical spellers over time and across languages are also needed before potential markers of atypical development can be identified (see Treiman et al., 2016, for an evaluation of several spelling error analysis schemes).

Sources of variations in spelling performance

One of the striking results in our initial analysis was the heterogeneity of the effect sizes observed across studies. To address this variation, the effects of language, age and task on effect sizes were examined. We further considered the phonological and reading profiles of a subgroup of children with DLD. We will firstly discuss the impact of reading and phonological profiles on the spelling performance of children with DLD, and then consider other potential sources of variations.

Our results suggest that phonological difficulties alone are sufficient to account for differences in spelling scores as compared to age-matched TD peers, in children as young as eight years old (English speakers) and as old as 16 years-old (Italian speakers). Thus, independent of age and orthographic transparency, children rely heavily on phonological skills to spell words in dictation. This is consistent with previous regression analysis conducted with children in grades 2-7, across French, English, German, Hungarian and Finnish languages (Moll et al., 2014). It is also consistent with the hypothesis that phonological recoding allows for the formation of accurate orthographic representations and later spelling performance (Shahar-Yames & Share, 2008). Isolated phonological difficulties are enough to drive spelling difficulties but as the results show, more pervasive language difficulties increase severity. Spelling involves representing sounds in words, but also the words themselves and links between words (Apel & Masterson, 2001; Treiman, 2017).

Our results also indicate the unique role of language skills in spelling. In the studies where the reading performance of children with DLD was assessed, children without reading difficulties, but with DLD still experienced difficulties with spelling as compared the age-matched TD peers. Decoding skills may not be enough to form accurate orthographic representations (Angellelli, Marinelli, & Zoccolotti, 2010). Whether these difficulties are driven only by phonological difficulties visible in spelling performance (e.g. Brizzolara et al., 2011; Stothard et al., 1998) or by different profiles of language difficulties remains unclear. Further studies with a differentiated profile of language and reading difficulties and a qualitative account of spelling errors could address this empirically.

Nonetheless, the addition of reading difficulties to the profiles of children with DLD had a significant impact on their spelling scores, as compared to those of children with DLDnoRI.

This confirms previous evidence of high correlations between reading and spelling skills (Dockrell et al., 2009; Vandewalle et al., 2012; Wolter et al., 2011) and suggests that poor reading further compromises the development of spelling abilities (Swanson et al., 2003; Zutell & Rasinski, 1989). Again, whether poor reading skills act as a proxy of poor phonological representations or as a driver for the formation of inaccurate orthographic representation remains unanswered. Experimental or intervention data may help address these questions (Angellelli et al., 2010).

Clinical thresholds, number and type of tasks used for language assessment, screening of the whole population or of a particular sample, non-verbal abilities and co-morbid disorders and age of the child at the time of language assessment are all paramount in defining what kind of language difficulties children are likely to experience. Different profiles of language difficulties may have consequences on the types of difficulties children experience in other literacy domains. For example, there is evidence of the differentiated roles of vocabulary and phonology in reading comprehension and written text generation on the one hand, and in word decoding and word spelling on the other hand (Dockrell & Connelly, 2015; Muter, Hulme, Snowling & Stevenson, 2004; Nation, Cocksey, Taylor & Bishop, 2010). All of these factors varied widely in the sample of studies that we analyzed. While representative of the DLD literature (Bishop et al., 2016), these variations raise challenges for establishing the links between specific oral language difficulties and spelling across development, and as such, developing explanatory models.

Similarly, measurement differences may appear because of differences in how spelling was assessed across studies, beyond the dichotomy between word dictation and text production that we used in our analyses. Heterogeneity was particularly high in the WD sample, reflecting variations in the choice of words children were given to spell (regular or irregular words, shorter or longer words), time constraints (with or without a time limit), and the context of presentation of the word material (in isolation or in sentence context). In spelling as in reading, different processes are assessed with different sets of words. Examining the spelling of irregular words may be particularly appropriate when it comes to assessing whole orthographic representations, whereas spelling regular words, pseudo words, or morphologically-complex words might allow for the assessment of sound- and meaning-to-letter correspondences and rules (Kohnen, Nickels, & Castles, 2009). The presence of homophones or morphologically-inflected words in the word sample, as in most contextualized spelling lists, might also call on a wider range of language-related skills (Apel & Masterson, 2001). Lack of details of the word lists in some of the studies did not allow us to differentiate these factors systematically. Appendix B captures these methodological differences and provides a description of the spelling task and of the recruitment criteria, as reported in each study. By contrast, text production tasks likely reflect the words participants felt confident to spell, thereby reducing heterogeneity in the results. However direct comparisons across tasks are missing and the number of studies using TP are small.

Surprisingly, our results did not reproduce the task differences observed in the individual studies by Broc et al. (2013), on the spelling of children with DLD, and by Sumner (2013) on the spelling performance of children with dyslexia. This result should be interpreted cautiously given the heterogeneity of word dictation tasks discussed above. More systematic comparisons of these two tasks, with comparable sets of words and measures of spelling errors are needed to elucidate which task may be more appropriate to assess both functional aspects and specific processes in the spelling performance of children with DLD.

We also hypothesized that orthographic characteristics of languages would play a role in the spelling performance of children with DLD, with larger effect sizes in less consistent languages such as French and English, and smaller effect sizes in more consistent languages such as Italian or Spanish. This hypothesis was made on the basis of current evidence on the slower rates of reading (Seymour et al., 2008) and spelling development (Wimmer & Landerl, 1997) in less consistent languages such as English. However, no differences between languages were found. The Anglocentricity (Share, 2008) of our sample may skew the results. Twenty-two of the 31 included studies assessed English spelling, compared to nine in other languages. Previous results from Marinelli et al. (2015) suggest that spelling accuracy, but not reading accuracy, is moderated by orthographic consistency. They compared the spelling performance of English and Italian children in 2nd and 5th grade on a set of words controlled for regularity, frequency and length. To our knowledge, this study is the only direct evidence of an effect of orthographic consistency on typical spelling performance. More studies are needed to account for language differences in the spelling profiles of children with DLD.

Limitations

Although the present study draws from the largest sample of studies available to-date on the spelling of children with DLD, it is limited by a number of methodological difficulties. First, qualitative differences in the spelling of children with DLD and TD peers could not be assessed using a meta-analysis. Narrative reviews may be appropriate to assess the potential qualitative differences that may affect the spelling of children with DLD (Bishop & Clarkson, 2003; Broc et al., 2014; Critten et al., 2014; Nauclér, 2004; Soriano-Ferrer & Contreras-González, 2012; Windsor et al., 2000). Secondly, the definition of spelling, phonological skills, language disorder and reading were limited to those presented in the studies included, and did not account for the multiple facets of these skills. Similarly, the profiles of the children with DLD were underspecified. Different profiles of children with DLD may lead to different spelling outcomes.

Attempts were made to reflect recruitment biases in Appendix B but were limited by the amount of information reported in the studies. Finally, conclusions from the present study are limited by the publication bias inherent in most meta-analyses, where there is a greater likelihood of publishing significant findings. Funnel plots, Egger's tests and selection methods adjustments provide an imperfect correction for such bias (Carter, Schönbrodt, Gervais, & Hilgard, pre-print).

Recommendations for future studies

Rigorous protocols need to be developed to clarify the relationships between oral and written language skills in children with DLD , to inform models of atypical spelling development and spelling assessment more generally. The following recommendations for research are derived from the analyses:

- Studies should provide a detailed description of how the sample with DLD was identified including phonological and reading profiles and co-morbidity with other disorders.
- Younger language- or spelling-matched peers should be included and a greater age range of children with DLD should be assessed, to provide insight into the patterns of spelling development in children with DLD.
- Different spelling tasks need to be described and compared within protocols including the use of qualitative and quantitative grids of analysis and well-controlled spelling lists.
- Studies should be conducted in languages other than English, ideally in languages that tap into phonological, orthographic and morphological processes. Direct cross-linguistic comparisons are needed to assess these processes.

Conclusion

A meta-analysis of the literature available on the spelling performance of children with DLD as compared to age- and language-matched peers was conducted. Results identified a large and significant difference in spelling scores between children with DLD and age-matched, but not language-matched peers. We confirmed the impact of phonological and reading skills on the spelling profiles of children with DLD, although difficulties in non-phonological skills may have a differential impact on spelling. Spelling provides a platform for assessing non-phonological language-literacy links that may be hindered in atypical development, beyond the initial stages of literacy development. The recommendations we raised from this meta-analysis should inform the future research agenda.

Acknowledgments

The preparation of this work was supported by the European COST Action IS1401 European Literacy Network. We would also like to thank the authors of the primary studies who supported the current work with supplementary information, as well as Professor Chloë Marshall, who commented on the final version of this paper.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Angelelli, P., Marinelli, C., & Zoccolotti, P. (2010). Single or dual orthographic representations for reading and spelling? A study of Italian dyslexic–dysgraphic and normal children. *Cognitive Neuropsychology*, *27*(4), 305–333. http://dx.doi.org/10.1080/02643294.2010.543539

Apel, K., & Masterson, J. (2001). Theory-guided spelling assessment and intervention: A case study. *Language, Speech, and Hearing Services in Schools*, *32*(3), 182–95. https://doi.org/10.1044/0161-1461(2001/017)

Bird, J., & Bishop, D. V. (1992). Perception and awareness of phonemes in phonologically impaired children. *European Journal of Disorders of Communication: The Journal of the College of Speech and Language Therapists, London*, *27*(4), 289–311. https://doi.org/10.3109/13682829209012042

*Bird, J., Bishop, D. V., & Freeman, N. H. (1995). Phonological awareness and literacy development in children with expressive phonological impairments. *Journal of Speech, Language, and Hearing Research*, *38*(2), 446–462. https://doi.org/10.1044/jshr.3802.446

Bishop, D., & Adams, C. (1990). A prospective study of the relationship between specific language impairment, phonological disorders and reading retardation. *Journal of Child Psychology and Psychiatry*, *31*(7), 1027–1050. https://doi.org/10.1111/j.1469-7610.1990.tb00844.x

*Bishop, D. V. M., & Clarkson, B. (2003). Written language as a window into residual language deficits: a study of children with persistent and residual speech and language impairments. *Cortex*, *39*(2), 215–237. https://doi.org/10.1016/S0010-9452(08)70106-0

*Bishop, D. V., McDonald, D., Bird, S., & Hayiou-Thomas, M. E. (2009). Children who read words accurately despite language impairment: Who are they and how do they do it? *Child Development*, *80*(2), 593–605. https://doi.org/10.1111/j.1467-8624.2009.01281.x

Bishop, D. V. M., & Snowling, M. J. (2004). Developmental dyslexia and specific language impairment: Same or different? *Psychological Bulletin, 130*(6), 858–886. https://doi.org/10.1037/0033-2909.130.6.858

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & the CATALISE consortium (2016). CATALISE: A multinational and multidisciplinary Delphi consensus study. Identifying language impairments in children. *PLOS ONE, 11*(7): e0158753. https://doi.org/10.1371/journal.pone.0158753

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T. & the CATALISE-2 consortium (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*. https://doi.org/10.1111/jcpp.12721

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Chichester, UK: John Wiley & Sons, Ltd. Retrieved from http://doi.wiley.com/10.1002/9780470743386

Botting, N., Simkin, Z., & Conti-Ramsden, G. (2006). Associated reading skills in children with a history of specific language impairment (SLI). *Reading and Writing*, *19*(1), 77–98. https://doi.org/10.1007/s11145-005-4322-4

*Brizzolara, D., Gasperini, F., Pfanner, L., Cristofani, P., Casalini, C., & Chilosi, A. M. (2011). Long-term reading and spelling outcome in Italian adolescents with a history of specific language impairment. *Cortex*, *47*(8), 955–973. https://doi.org/10.1016/j.cortex.2011.02.009

Broc, L., Bernicot, J., Olive, T., Favart, M., Reilly, J., Quémart, P., Catheline, N., Gicquel, L., & Jaafari, N. (2014). Évaluation de l’orthographe des élèves dysphasiques en situation de narration communicative : Variations selon le type d’orthographe, lexicale versus morphologique. *Revue Européenne de Psychologie Appliquée*, *64*(6), 307–321. https://doi.org/10.1016/j.erap.2014.09.004

*Broc, L., Bernicot, J., Olive, T., Favart, M., Reilly, J., Quémart, P., & Uzé, J. (2013). Lexical spelling in children and adolescents with specific language impairment: Variations with the writing situation. *Research in Developmental Disabilities*, *34*(10), 3253–3266. https://doi.org/10.1016/j.ridd.2013.06.025

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*, 365–376. https://doi.org/10.1038/nrn3475

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

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (pre-print). Correcting for bias in psychology: A comparison of meta-analytic methods. Retrieved 2018, May 19, from http://doi.org/10.17605/OSF.IO/9H3NU

Catts, H. W. (1993). The relationship between speech-language impairments and reading disabilities. *Journal of Speech, Language, and Hearing Research*, *36*, 948–958. https://doi.org/doi:10.1044/jshr.3605.948

Catts, H. W., Bridges, M. S., Little, T. D., & Tomblin, J. B. (2008). Reading achievement growth in children with language impairments. *Journal of Speech, Language, and Hearing Research*, *51*(6), 1569–1579. https://doi.org/10.1044/1092-4388(2008/07-0259)

Catts, H. W., Fey, M. E., Tomblin, J. B., & Zhang, X. (2002). A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language, and Hearing Research*, *45*(6), 1142–1157. https://doi.org/10.1044/1092-4388(2002/093)

*Connelly, V., Dockrell, J. E., Walter, K., & Critten, S. (2012). Predicting the quality of composition and written language bursts from oral language, spelling, and handwriting skills in children with and without specific language impairment. *Written Communication*, *29*(3), 278–302. https://doi.org/10.1177/0741088312451109

*Cordewener, K. A. H., Bosman, A. M. T., & Verhoeven, L. (2012). Specific language impairment affects the early spelling process quantitatively but not qualitatively. *Research in Developmental Disabilities*, *33*(4), 1041–1047. https://doi.org/10.1016/j.ridd.2012.01.011

COST European Cooperation in Science and Technology (2017). COST member states [Online]. Retrieved from http://www.cost.eu/about_cost/cost_member_states, June 22.

Critical Appraisal Skills Programme (2017). *CASP Cohort study checklist* [Online]. Retrieved from http://www.casp-uk.net/checklists, June 22.

*Critten, S., Connelly, V., Dockrell, J. E., & Walter, K. (2014). Inflectional and derivational morphological spelling abilities of children with Specific Language Impairment. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00948

*Deacon, S. H., Cleave, P. L., Baylis, J., Fraser, J., Ingram, E., & Perlmutter, S. (2014). The representation of roots in the spelling of children with Specific Language Impairment.

Journal of Learning Disabilities, *47*(1), 13–21. https://doi.org/10.1177/0022219413509965

*de Bree, E., Wijnen, F., & Gerrits, E. (2010). Non-word repetition and literacy in Dutch children at-risk of dyslexia and children with SLI: Results of the follow-up study. *Dyslexia*, *16*(1), 36–44. https://doi.org/10.1002/dys.395

*Dockrell, J. E., & Connelly, V. (2015). The role of oral language in underpinning the text generation difficulties in children with specific language impairment: Text generation in children with specific language impairments. *Journal of Research in Reading*, *38*(1), 18–34. https://doi.org/10.1111/j.1467-9817.2012.01550.x

*Dockrell, J. E., Lindsay, G., & Connelly, V. (2009). The impact of specific language impairment on adolescents' written text. *Exceptional Children*, *75*(4), 427–446. https://doi.org/10.1177/001440290907500403

*Dockrell, J. E., Lindsay, G., Connelly, V., & Mackie, C. (2007). Constraints in the production of written text in children with specific language impairments. *Exceptional Children*, *73*(2), 147–164. https://doi.org/10.1177/001440290707300202

*Eisenmajer, N., Ross, N., & Pratt, C. (2005). Specificity and characteristics of learning disabilities. *Journal of Child Psychology and Psychiatry*, *46*(10), 1108–1115. https://doi.org/10.1111/j.1469-7610.2004.00394.x

Fraser, J., Goswami, U., & Conti-Ramsden, G. (2010). Dyslexia and specific language impairment: The role of phonology and auditory processing. *Scientific Studies of Reading*, *14*(1), 8–29. http://dx.doi.org/10.1080/10888430903242068

Gough, D., Oliver, S., & Thomas, J. (Eds.). (2012). *An introduction to systematic reviews*. London: SAGE.

Gough, D., Oliver, S., Thomas, J. (2013). *Learning from research: Systematic reviews for informing policy decisions: A quick guide* [Online]. London: Nesta. Retrieved from http://www.alliance4usefulevidence.org/assets/Alliance-FUE-reviews-booklet-3.pdf

*Goulandris, N. K., Snowling, M. J., & Walker, I. (2000). Is dyslexia a form of specific language impairment? A comparison of dyslexic and language impaired children as adolescents. *Annals of Dyslexia*, *50*(1), 103–120. https://doi.org/10.1007/s11881-000-0019-1

Higgins, J. P. T. (2008). Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified. *International Journal of Epidemiology*, *37*(5), 1158–1160. https://doi.org/10.1093/ije/dyn204

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

Kim, Y. (2010). Componential skills in early spelling development in Korean. *Scientific Studies of Reading*, *14*(2), 137-158. http://dx.doi.org/10.1080/10888430903034812

Kohnen, S., Nickels, L., & Castles, A. (2009). Assessing spelling skills and strategies: A critique of available resources. *Australian Journal of Learning Difficulties*, 14(1), 113–150. https://doi.org/10.1080/19404150902783450

*Larkin, R. F., & Snowling, M. J. (2008). Comparing phonological skills and spelling abilities in children with reading and language impairments. *International Journal of Language & Communication Disorders*, *43*(1), 111–124. https://doi.org/10.1080/13682820601178584

*Larkin, R. F., Williams, G. J., & Blaggan, S. (2013). Delay or deficit? Spelling processes in children with specific language impairment. *Journal of Communication Disorders*, *46*(5–6), 401–412. https://doi.org/10.1016/j.jcomdis.2013.07.003

*Lewis, B. A., O'Donnell, B., Freebairn, L. A., & Taylor, H. G. (1998). Spoken language and written expression—Interplay of delays. *American Journal of Speech-Language Pathology*, *7*(3), 77–84. https://doi.org/10.1044/1058-0360.0703.77

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. London: Sage Publications.

*Mackie, C., & Dockrell, J. E. (2004). The nature of written language deficits in children with SLI. *Journal of Speech, Language, and Hearing Research*, *47*(6), 1469–1483. https://doi.org/10.1044/1092-4388(2004/109)

*Mackie, C. J., Dockrell, J., & Lindsay, G. (2013). An evaluation of the written texts of children with SLI: The contributions of oral language, reading and phonological short-term memory. *Reading and Writing*, *26*(6), 865–888. https://doi.org/10.1007/s11145-012-9396-1

Marinelli, C. V., Romani, C., Burani, C., & Zoccolotti, P. (2015). Spelling acquisition in English and Italian: A cross-linguistic study. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.01843

*McCarthy, J. H., Hogan, T. P., & Catts, H. W. (2012). Is weak oral language associated with poor spelling in school-age children with specific language impairment, dyslexia or both? *Clinical Linguistics & Phonetics*, *26*(9), 791–805. https://doi.org/10.3109/02699206.2012.702185

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 Medicine, 6*(7): e1000097. https://doi.org/10.1371/journal.pmed.1000097

Moll, K., Ramus, F., Bartling, J., Bruder, J., Kunze, S., Neuhoff, N., Streiftau, S., Lyytinen, H., Leppänen, P. H. T., Lohvansuu, K., Tóth, D., Honbolygó, F., Csépe, V., Bogliotti, C., Iannuzzi, S., Démonet, J-F., Longeras, E., Valdois, S., George, S, Soares-Boucaud, I., Le Heuzey, M-F., Billard, C., O'Donovan, M., Hill, G., Williams, J., Brandeis, D., Maurer, U., Schulz, E., van der Mark, S., Müller-Myhsok, B., Schulte-Körne, G., Landerl, K., (2014). Cognitive mechanisms underlying reading and spelling development in five European orthographies. *Learning and Instruction*, *29*, 65–77. https://doi.org/10.1016/j.learninstruc.2013.09.003

Muter, V., Hulme, C., Snowling, M.J. & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology*, 40(5), 665–681. https://doi.org/doi:10.1037/0012-1649.40.5.665

Muter, V., & Snowling, M. (1998). Concurrent and longitudinal predictors of reading: The role of metalinguistic and short-term memory skills. *Reading Research Quarterly, 33*(3), 320. https://doi.org/10.1598/RRQ.33.3.4

Nagy, W., Berninger, V., & Abbott, R. (2006). Contributions of morphology beyond phonology to literacy outcomes of upper elementary and middle school students. *Journal of Educational Psychology, 98*(1), 134–147. https://doi.org/10.1037/0022-0663.98.1.134

Nation, K., Cocksey, J., Taylor, J.S.H. & Bishop, D.V.M. (2010). A longitudinal investigation of early reading and language skills in children with poor reading comprehension. *Journal of Child Psychology and Psychiatry*, *51*(9), 1031–1039. https://doi.org/doi:10.1111/j.1469-7610.2010.02254.x

Nauclér, K. (2004). Spelling development in Swedish children with and without language impairment. *Journal of Multilingual Communication Disorders*, *2*(3), 207–215. https://doi.org/10.1080/14769670400018315

Ouellette, G. P., & Sénéchal, M. (2008). A window into early literacy: Exploring the cognitive and linguistic underpinnings of invented spelling. *Scientific Studies of Reading*, *12*(2), 195–219. http://dx.doi.org/10.1080/10888430801917324

Peterson, R. L., Pennington, B. F., Shriberg, L. D., & Boada, R. (2009). What influences literacy outcome in children with speech sound disorder? *Journal of Speech, Language, and Hearing Research*, *52*(5), 1175–1188. https://doi.org/10.1044/1092-4388(2009/08-0024)

*Puranik, C. S., Lombardino, L. J., & Altmann, L. J. (2006). Writing through retellings: An exploratory study of language-impaired and dyslexic populations. *Reading and Writing*, *20*(3), 251–272. https://doi.org/10.1007/s11145-006-9030-1

Quick, N. & Erickson, K. (2018). A Multilinguistic Approach to Evaluating Student Spelling in Writing Samples. *Language, Speech, and Hearing Services in Schools*, [Advance online publication], 1-15. https://doi.org/10.1044/2018_LSHSS-17-0095

R Core Team. (2017). *R: A language and environment for statistical computing* [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/, June, 22

*Rakhlin, N., Cardoso-Martins, C., Kornilov, S. A., & Grigorenko, E. L. (2013). Spelling well despite developmental language disorder: What makes it possible? *Annals of Dyslexia*, *63*(3–4), 253–273. https://doi.org/10.1007/s11881-013-0084-x

*Ramus, F., Marshall, C. R., Rosen, S., & van der Lely, H. K. J. (2013). Phonological deficits in specific language impairment and developmental dyslexia: towards a multidimensional model. *Brain*, *136*(2), 630–645. https://doi.org/10.1093/brain/aws356

Reilly, S., Bishop, D., & Tomblin, B. (2014). Terminological debate over language impairment in children: Forward movement and sticking points. *International Journal of Language & Communication Disorders, 49*(4), 452–462. https://doi.org/10.1111/1460-6984.12111

*Rose, J. C., Lincoln, A. J., & Allen, M. H. (1992). Ability profiles of developmental language disordered and learning disabled children: A comparative analysis. *Developmental Neuropsychology*, *8*(4), 413–426. https://doi.org/10.1080/87565649209540535

Rothstein, H., Sutton, A. J, & Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Chichester, England; Hoboken, NJ: Wiley.

Scott, C. M., & Windsor, J. (2000). General language performance measures in spoken and written narrative and expository discourse of school-age children with language learning disabilities. *Journal of Speech, Language, and Hearing Research*, *43*(2), 324–39. https://doi.org/10.1044/jslhr.4302.324

*Scuccimarra, G., Cutolo, L., Fiorillo, P., Lembo, C., Pirone, T., & Cossu, G. (2008). Is there a distinct form of developmental dyslexia in children with specific language impairment? Findings from an orthographically regular language. *Cognitive and Behavioral Neurology, 21*(4), 221–226. https://doi.org/10.1097/WNN.0b013e31818a5caf

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

Shahar-Yames, D., & Share, D. (2008). Spelling as a self-teaching mechanism in orthographic learning. *Journal of Research in Reading, 31*(1), 22–39. https://doi.org/10.1111/j.1467-9817.2007.00359.x

Share, D. L. (2008). On the anglocentricities of current reading research and practice: The perils of overreliance on an "outlier" orthography. *Psychological Bulletin*, *134*(4), 584–615. https://doi.org/10.1037/0033-2909.134.4.584

Silliman, E. R., Bahr, R. H., & Peters, M. L. (2006). Spelling patterns in preadolescents with atypical language skills: Phonological, morphological, and orthographic factors. *Developmental Neuropsychology*, *29*(1), 93–123. http://doi.org/10.1207/s15326942dn2901_6

*Soriano-Ferrer, M., & Contreras-González, M. (2012). Narraciones escritas en niños con Trastorno Específico del Lenguaje (TEL). *Universitas Psychologica*, *11*(4), 1341–1351.

Stothard, S. E., Snowling, M. J., Bishop, D. V. M., Chipchase, B. B., & Kaplan, C. A. (1998). Language-impaired preschoolers: A follow-up into adolescence. *Journal of Speech, Language and Hearing Research*, *41*(2), 407. https://doi.org/10.1044/jslhr.4102.407

Sumner, E. J. (2013) *Children with dyslexia: Spelling as a constraint on writing development.* PhD Thesis. Oxford Brookes University.

Swanson, H. L., Trainin, G., Necoechea, D. M., & Hammill, D. D. (2003). Rapid naming, phonological awareness, and reading: A meta-analysis of the correlation evidence. *Review of Educational Research*, *73*(4), 407–440. https://doi.org/10.3102/00346543073004407

Tong, X., McBride-Chang, C., Shu, H., & Wong, A. M-Y. (2009). Morphological awareness, orthographic knowledge, and spelling errors: Keys to understanding early Chinese literacy acquisition. *Scientific Studies of Reading*, *13*(5), 426–452. http://dx.doi.org/10.1080/10888430903162910

Torgerson, C. J., & Elbourne, D. (2002). A systematic review and meta-analysis of the effectiveness of information and communication technology (ICT) on the teaching of spelling. *Journal of Research in Reading*, *25*(2), 129–143. https://doi.org/10.1111/1467-9817.00164

Treiman, R. (2017). Learning to Spell Words: Findings, Theories, and Issues. *Scientific Studies of Reading*, *21*(4), 265–276. https://dx.doi.org/10.1080/10888438.2017.1296449

Treiman, R., Kessler, B., Pollo, T., Byrne, B., & Olson, R. (2016). Measures of Kindergarten Spelling and Their Relations to Later Spelling Performance. *Scientific Studies of Reading*, *20*(5), 349-362. https://doi/10.1080/10888438.2016.1186168

*Vandewalle, E., Boets, B., Ghesquière, P., & Zink, I. (2012). Auditory processing and speech perception in children with specific language impairment: Relations with oral language and literacy skills. *Research in Developmental Disabilities*, *33*(2), 635–644. https://doi.org/10.1016/j.ridd.2011.11.005

Viechtbauer, W. (2010). Conducting meta-analysis in R with the metafor package. *Journal of Statistical Software*, *363*, 1–48. https://doi.org/10.18637/jss.v036.i03

Vugs, B., Cuperus, J., Hendriks, M., & Verhoeven, L. (2013). Visuospatial working memory in specific language impairment: A meta-analysis. *Research in Developmental Disabilities*, *34*(9), 2586–2597. https://doi.org/10.1016/j.ridd.2013.05.014

*Werfel, K. L., & Krimm, H. (2015). Utility of the Spelling Sensitivity Score to analyze spellings of children with specific language impairment. *Australian Journal of Learning Difficulties*, *20*(1), 39–53. https://doi.org/10.1080/19404158.2015.1047871

*Williams, G. J., Larkin, R. F., & Blaggan, S. (2013). Written language skills in children with specific language impairment: Writing and SLI. *International Journal of Language & Communication Disorders*, *48*(2), 160–171. https://doi.org/10.1111/1460-6984.12010

Wimmer, H., & Landerl, K. (1997). How learning to spell German differs from learning to spell English. In C.A. Perfetti, L. Rieben, & M. Fayol (Eds.), *Learning to spell: Research, theory and practice across languages.* Mahwah, NJ: Erlbaum.

Windsor, J., Scott, C. M., & Street, C. K. (2000). Verb and noun morphology in the spoken and written language of children with language learning disabilities. *Journal of Speech, Language, and Hearing Research*, *43*(6), 1322–36. https://doi.org/10.1044/jslhr.4306.1322

*Wolter, J. A., Self, T., & Apel, K. (2011). Initial mental graphemic representation acquisition and later literacy achievement in children with language impairment: A longitudinal study. *Journal of Learning Disabilities*, *44*(6), 543–555. https://doi.org/10.1177/0022219410392042

Ziegler, J. C., Jacobs, A. M., & Stone, G. O. (1996). Statistical analysis of the bidirectional inconsistency of spelling and sound in French. *Behavior Research Methods, Instruments, & Computers*, *28*(4), 504–515.

Ziegler, J. C., Stone, G. O., & Jacobs, A. M. (1997). What is the pronunciation for-ough and the spelling for/u/? A database for computing feedforward and feedback consistency in English. *Behavior Research Methods*, *29*(4), 600–618.

Zutell, J. & Rasinski, T. (1989). Reading and spelling connections in third and fifth grade students. *Reading Psychology*, *10*(2), 137–55. http://dx.doi.org/10.1080/0270271890100203