What spelling errors can tell us about the development of processes involved in children's spelling

Georgia Niolaki[1*], Alexandra Negoita[2], Janet Vousden[3], Aris Terzopoulos[4], Laura Taylor[5], Jackie Masterson[6]

[1]Bath Spa University, United Kingdom, [2]Coventry University, United Kingdom, [3]Nottingham Trent University, United Kingdom, [4]Birmingham City University, United Kingdom, [5]University of Northampton, United Kingdom, [6]Development Education Research Centre, Institute of Education, University College London, United Kingdom

Conflict of interest statement

Author contribution statement

G. Niolaki, A. Negoita, J. Vousden, A. Terzopoulos, J. Masterson: These authors contributed equally to this work and share the first authorship

L. Taylor These authors share last authorship

Keywords

Abstract

Word count: 224

Spelling is an essential foundation for reading and writing. However, many children leave school with spelling difficulties. By understanding the processes children use when they spell, we can intervene with appropriate instruction tailored to their needs. Our study aimed to identify key processes (lexical-semantic and phonological) by using a spelling assessment that distinguishes different printed letter strings/word types (regular and irregular words, and pseudowords). Misspellings in the test from 641 pupils in Reception Year to Year 6 were scored using alternatives to binary correct versus incorrect scoring systems. The measures looked at phonological plausibility, phoneme representations and letter distance. These have been used successfully in the past but not with a spelling test that distinguishes irregularly spelled words from regular words and pseudowords. The findings suggest that children in primary school rely on both lexical-semantic and phonological processes to spell all types of letter string, but this varies depending on the level of spelling experience (younger Foundation/Key stage 1 and older Key stage 2). Although children in younger year groups seemed to rely more on phonics, based on the strongest correlation coefficients for all word types, with further spelling experience, lexical processes seemed to be more evident, depending on the type of word examined. The findings have implications for the way we teach and assess spelling and could prove to be valuable for educators.

Contribution to the field

Research in spelling is sparse in comparison to reading, and many children leave school with spelling difficulties. Therefore we focus on misspellings made by 641 pupils in Reception Year to Year 6. Rather than focusing solely on binary (correct vs incorrect) scoring ways, accuracy and phonological plausibility, we also tested alternative non-binary measures; an automatised continuous measure, the Automated Measure of Phoneme Representation (AMPR) and letter distance (LD) in our analyses. In addition, we used a spelling assessment that distinguishes different printed item types (regular and irregular words, and pseudowords). The main findings of this study were that lower year_groups rely more on sublexical processes rather than the higher year_groups. These changes are also related to the different types of words. Furthermore, the non-binary measures captured these processes in more detail than the binary ones. The findings support well-established theories in spelling research, e.g., lexical quality hypothesis, dual route model. In addition, it offers a new approach for assessing spelling using the three different word types and utilising spelling errors to inform developmental progress. Both these can have an impact on designing necessary support for children with spelling difficulties.

Funding information

Ethics statements

Studies involving animal subjects
Generated Statement: No animal studies are presented in this manuscript.

Studies involving human subjects
Generated Statement: The studies involving human participants were reviewed and approved by Bath Spa University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

Inclusion of identifiable human data
Generated Statement: No potentially identifiable human images or data is presented in this study.

Data availability statement

What spelling errors can tell us about the development of processes involved in children's spelling

Every letter counts; What errors can tell us about the children's spelling skill?Spelling in primary school children, moving beyond cognitive elements

Georgia Z. Niolaki[1] [†], Alexandra Negoita[2] [†],Janet I. Vousden[3] [†], Aris R. Terzopoulos[4] [†], Laura Taylor[5] and Jackie Masterson[6] [†]

[1] School of Education, Bath Spa University

[2] School of Psychological, Social and Behavioural Sciences, Coventry University,

[3] Department of Psychology, Nottingham Trent University

[4] School of Psychology, Birmingham City University

[5] School of Psychology, Northampton University

[6] Department of Psychology, UCL, Institute of Education

Author Note

Correspondence concerning this article should be addressed to Georgia Z. Niolaki, School of Education, Newton Park, Bath Spa University, Bath, UK. Tel.:07881594772. Email: g.niolaki@bathspa.ac.uk

†These authors contributed equally to this work and share first authorship

Pages 32 number of words 6,510

ABSTRACT

Spelling is an essential foundation for reading and writing. However, many children leave school with spelling difficulties. By understanding the processes children use when they spell, we can intervene with appropriate instruction tailored to their needs. Our study aimed to identify key processes (lexical-semantic and phonological) by using a spelling assessment that distinguishes different printed letter strings/word types (regular and irregular words, and pseudowords). Misspellings in the test from 641 pupils in Reception Year to Year 6 were scored using alternatives to binary (correct versus incorrect) scoring systems. The measures looked at phonological plausibility, phoneme representations and letter distance. These have been used successfully in the past but not with a spelling test that distinguishes different letter string typesirregularly spelled words frorm regular words and pseudowords. The findings suggest that children in primary school rely on both lexical-semantic and phonological processes to spell all types of letter stringitems (irregular, regular and pseudowords)letter string, but this varies depending on the level of spelling experience (younger Foundation/Key– stage 1 and older Key– stage 2). Although children in younger year groups seemed to rely more on phonics, based on the strongest correlation coefficients for all letter stringword types, with further spelling experience, lexically related variables processes

seemed to become more evidentimportant, depending on the type of word examined. These findings have implications for the way we teach and assess spelling and could prove to be valuable for educators.

Keywords: Spelling; phonological plausibility (PhP); letter distance (LD); Automated Measures of Phonemes Representations (AMPR).

Introduction

Acquiring spellingLearning to spell is a lengthy and challenging process, and it is not surprising that some children struggle and need extra support (Nag et al., 2018). Therefore, it is important to find accurate and sensitive methods for assessing children's spellings (Treiman & Kessler, 2004). One way to achieve this is to move away from the traditional binary correct vs incorrect scoring method. Research has suggested that looking at children's misspellings beyond the dichotomous scoring system might reveal patterns of development in linguistic knowledge (Masterson & Apel, 2010; de Bree & van der Boer, 2019; Treiman et al., 2019; see also for non-English orthographies see also Carvalhais et al., 2020; 2021; Niolaki & Masterson, 2012) and underlying cognitive processes that children use when they spell (Caravolas et al., 2001; Kessler, 2009; Treiman et al., 2016). These processes involve perception, attention, memory and expertise, all of which past research has explored in considerable depth (Georgiou et al., 2012; Niolaki et al., 2022; Rønneberg and Torrance, 2019; Zoccolotti et al., 2020). As such, to date, spelling has been primarily studied so far more in relation to cognitive abilities than to the actual spelling errors made. Recent research has suggested that it is worth looking in more depth at the characteristics of spelling attempts with different types of words (i.e., regular and irregular words and pseudowords) (de Bree & van der Boer, 2019; Niolaki et al., 2020). The present study aimed to investigate whether children rely on different processes for different types of words in their spelling, and explored

whether this reliance might change from Reception year to Year 6 (equivalent to Kindergarten to Grade 6 in the USA).

Dual route (DR) models of spelling have been proposed to explain how information is processed when spelling words (Barry, 1994). Similarly, models, such as the integration of multiple patterns (IMP) model that supports statistical learning (which is considered a powerful strategy for generalising learning from untaught items) agree that both sublexical and lexical/semantic processes exist. The IMP model also recognises that items can be divided into regularly and irregularly spelled words (Treiman, 2018, p. 648). These models suggest that people use sound-to-spelling rules, or *sublexical processing*, as well as stored knowledge of whole-word spellings, or *lexical processing* (Barry, 1994; Martin & Barry, 2012). Children are believed to possess the basis for lexical processing before they start to spell, as a result of spoken language and early exposure to printed text (Barry, 1994). Children are also believed to memorise the orthographic rules and learn to spell correctly as they gain more experience with print. As a result, children perform better on words that occur frequently in books (Kessler et al., 2012). A more pronounced frequency effect has been observed for irregularly spelled words (such as *DEBT, YACHT, MORTGAGE*), indicating that children need to hear, read and write these words more times than regularly spelled words (such as *BARGE*, *SLATE* and *TARGET*) in order to create a correct representation in memory (Koerse et al., 2000).

Empirical evidence suggests that novice spellers rely heavily or even exclusively on the sub-lexical route (Caravolas et al., 2001), and they progressively develop the lexical-semantic route. In the UK, phonics training is the starting fuel for children to understand how phoneme-grapheme correspondences operate. This is important as phonemes in English can have several spelling options (Barry & Seymour, 1988), for example, /k/ can be spelled <c>, <k>, or <ck>. The options vary in probability of occurrence, and this depends in part on

whether context-free or context-sensitive probabilities are considered (context sensitive as in <magic -> magician>) (Spencer, 2009). Overall, spelling encompasses different tiers of linguistic awareness (phonology, orthography, morphology and semantics), so to establish solid lexical representations, the spellers need to be aware of all these and their interrelationships.

To gain insight into the processes children use in spelling, sensitive, non-binary scoring systems have been developed to examine spelling errors (e.g., Bruck & Waters, 1988; Kessler, 2009; Treiman & Kessler, 2004). These systems have also been found to have good discriminatory power for students struggling with spelling (Treiman et al., 2016; for non-English languages see Joye et al., 2020; Niolaki et al., 2014). The non-binary measures rely, for example, on how close the misspelled word is to the correct spelling and can be used as a fine-grained measure to monitor the progress a child is making in developing spelling skills (Masterson & Apel, 2010). Many children, especially in the Foundation Year, produce more errors than correct responses (Carvalhais et al., 2021; Georgiou et al., 2020), and the measures can provide us with a clear picture of the use of lexical-semantic and sublexical processes, and how children's reliance on these processes changes as they progress in spelling skill. Also, error analysis is not affected by floor effects which is an issue when we solely look at accuracy especially for young spellers (Treiman et al., 2019). We next present the measures we included in the current investigation.

Phonologically plausible errors (e.g., *elephant* spelled <elefant>) are considered to reflect use of sublexical spelling processes and are particularly apparent in novice spellers. Caravolas et al. (2001) analysed spellings from 153 children in the three first years in school in the UK. They emphasised the critical predictive role of phonological spelling ability for later reading and spelling accuracy. The researchers argued that children need phonics training (structured instruction that helps children to spell unfamiliar words) to build a solid

sound-letter mapping system and then to become skilled spellers who competently use orthographic rules (Caravolas et al., 2001).

In a longitudinal and cohort study with 95 Portuguese students from two age groups—Grades 4–7 ~~(n = 47)~~ and 6–9, ~~(n = 48), the researchers~~ (Carvalhais et al.~~,~~ (2021) ~~also used~~conducted spelling error analyses and looked at phonological plausibility as one of the critical variables. They found ~~that misspellings decreased from T1 to T2 and~~ that younger children made more misspellings than the older ~~ones~~children and ~~. Similarly, p~~phonologically inappropriate errors ~~decreased from T1 to T2 in both groups, and they~~ were ~~also~~ less in the older group. These findings are consistent with similar observations made in the English orthography (Caravolas et al., 2001; Treiman & Bourassa, 2000). The~~se~~ research~~ers~~ highlights the importance of phoneme-grapheme associations in the earliest stages of spelling in English.

In the current study we employed two separate measures of phonological plausibility - a binary phonological plausibility score (PhP, e.g., spelling *elephant* as <elefant>), and an automatised continuous measure, the Automated Measure of Phoneme Representation (~~(~~AMPR~~) http://spell.psychology.wustl.edu/AMPR/~~). ~~The AMPR calculates a lower score when phonologically implausible errors are made, for example, *life* spelled as *live* would receive a score of .66 (2/3 phonemes correctly represented) while *life* spelled <life> would receive a score of 1 (3/3 phonemes correctly represented, the highest score).~~ The main difference between PhP and AMPR is the first is a binary measure (phonologically plausible error or not) and is hand scored (by the research team), whereas the second provides a score computed across the word. For AMPR, values closer to 0 indicate a lower quality of error, as the target is distant from the response (Treiman & Kessler, 2004), meaning that as children's spelling skill develops the AMPR score should increase.

We also used letter distance (LD)[21] in our analyses to capture the number of letter additions, deletions and substitutions needed to create the correct phonological and orthographic spelling from an error. ~~Kessler (2009) created a computerised tool to calculate this variable.~~

Treiman et al. (2016) investigated a range of scoring measures in a longitudinal study with children from Kindergarten to Grade 2. The findings indicated that children possess some phonological knowledge early in spelling development, however, LD, the lexically-related measure, proved a better predictor of spelling accuracy in beginner spellers than the phoneme distance measure, the sublexically-related one (Treiman et al., 2016). We considered that it would be informative to see, in our study with UK children, and almost 21 years after the implementation of synthetic phonics in schools, whether PhP or LD would be more strongly associated with emergent spelling, and whether the association would differ for regular words, irregular words and pseudowords, as these are assumed to draw on ~~reflect~~ differen~~ce~~t processes, that is, (whole-word, and sublexical~~phonological)~~. For LD, values that are closer to 0 indicate less distance from the correct spelling. Thus, as children develop their spelling knowledge LD values should decrease.

Several studies have reported that more sensitive scoring methods can effectively capture developmental changes in spelling and strong associations with reading and phonological ability (Ritchey et al., 2010; Clemens et al., 2014; Frisby, 2016). As noted above, Treiman et al. (2016) investigated a range of measures in a longitudinal study. Participants were 374 children from kindergarten age to Grade 2 in the USA and Australia. The researchers employed letter-based measures (LD, correctness, letter sequence) and phoneme-based measures (AMPR, phoneme distance, sound-spelling) at two-time points. The findings revealed that letter-based measures accounted for more variance in spelling accuracy than phoneme-based measures. Treiman et al. (2019) reported a replication study

with British English spellers where correctness (a binary correct/incorrect measure) was more predictive of single word spelling at Time 4 <u>assessment</u> (when the~~y~~ <u>children</u> attended the spring term of Year 2) ~~assessment~~ than non-binary measures. Binary measures might be useful for predicting who is likely to struggle with spelling but would not tell us why. In this case non-binary measures would give us more insight.

<u>Carvalhais et al. (2020) in their study with Portuguese students (also reported</u> ~~also earlier)~~ <u>found that orthographically related variables (such as stress mark errors and orthographic misspellings) were the most common errors in older learners (Grades 4, 6 and 7). In the English-</u>~~-~~<u>based studies reported above</u>~~earlier~~ <u>children were tested in the earlier school grades, so there is not much evidence for English on how the</u> ~~variables (phonological-and orthographic-related measures~~ <u>based) will</u> perform in~~for~~ <u>older</u> ~~in age learners~~<u>spellers.</u>

In summary, it has not been established if developmental changes in spelling errors apply equally to regular and irregular words and pseudowords. Moreover, several scoring methods have been found to relate to one another, for example, PhP and AMPR (Treiman et al., 2019)~~.,~~ ~~Y~~<u>y</u>et~~,~~ they have not been investigated concerning spelling errors with different letter string<u>/item</u> types. The current study addressed these gaps in the literature by exploring patterns of spelling errors in primary school children split into three age groups, Foundation Year/Key Stage 1 (F/KS1) (Kindergarten to grade 2), Early Key Stage 2 (EKS2) (Grade 3 to 4) and Advanced Key Stage 2 (AKS2) (Grade 5 to 6). Thus, the goal of the present study was to investigate the strategies children use when they spell different types of words<u>,</u> ~~(~~through analysis of their errors~~)~~, and how this may change from reception year to year six (5- to 12-years).

The current study

<u>We</u>~~The present study~~ aimed to investigate whether children rely on different processes for different types of words in their spelling, and explored whether this reliance

might change from Reception year to Year 6 (equivalent to Kindergarten to Grade 6 in the USA). Research has indicated that both lexical-semantic and sub-lexical processes are employed for spelling by beginner spellers, while the former processes seem to become more prevalent when children gain more experience with reading and spelling (Bruck & Waters, 1988; Caravolas et al., 2001; Kessler et al., 2012; Share, 1995). Treiman et al. (2016; 2019) found that lexically related variables were strong predictors of more advanced spelling (however, the older children only went up to Grade/Year 2). When investigating the spelling of different types of letter strings, it is likely that children rely more on sub-lexical processes for pseudowords, and lexical processes for irregular words, while the spelling of regular words will tap both processes (Niolaki et al., 20~~19~~22). Based on the research reviewed above, we aimed to test the following hypotheses. We expected that the sub-lexically related (phoneme--based) measure~~s~~ ~~(AMPR)~~ would be more important for spelling pseudowords than irregular word ~~spelling~~. We also anticipated that the sublexically- and lexically-related ~~measures~~ (letter based) measure ~~(LD)~~ would be important for ~~irregular and regular~~real words; but ~~lexically-related measures will be~~ less strongly associated with ps~~u~~eudoword spelling.

Specifically, we made the ~~aimed to answer the~~ following p~~research ques~~redictions:

- There will be a significant interaction between key stages and word types for spelling accuracy, letter-based and phoneme-based measures. ~~It is predicted that l~~Lower key stages should evidence~~will have~~ lower scores on accuracy and phoneme-based scores and higher scores on letter-based measures for each word type than higher key stages.

- There will be significant associations between phoneme-based and letter-based measures and accuracy for all word types in all key stages. The strength of associations will vary depending on key stage and letter string type.

Methods

Participants

Participants were 641 UK primary school children attending Reception Year to Year 6, from a mix of urban and rural schools (seven different state schools). School years were divided into three levels as follows: Group 1 comprised children in Foundation /Key Stage 1 (F/KS1), i.e., Reception Year to Year 2, Group 2 were children in the first half of Early Key Stage 2 (EKS2), i.e., Year 3 and 4, and Group 3 were children in the second half of Advanced Key Stage 2 (AKS2), i.e., Year 5 and 6. Pupils were grouped this way to allow for the identification of strategies used for spelling in developmental stages (beginning, early and late stages/advanced spellers). The number of children in each age group, together with their mean chronological age, is shown in Table 1. Data were collected after parents/carers of participating students returned consent forms and after children assented to participate in the study. The University Ethics' Committee granted ethical approval for the study.

<Table 1 here>

Materials

Data from a new interpretive spelling test for primary school children were collected. The test comprises consisteds of three sections; comprising each has comprising 36 irregular words (e.g., <yacht>), 36 regular words (e.g., <cat>), and 34 pseudowords (e.g., <trelfishly>). Spelling regularity was calculated based on the frequency of occurrence of sound-letter correspondences in the word (Spencer, 2009; Vousden et al., 2010)1) but also spelling instruction in UK schools. Pseudowords were formed by combining the first half of a regular word and the last half of another regular word. When spelled, these items follow regular words' structure, but they were unfamiliar to the children. For pseudoword accuracy

(using binary scoring) we categorised any plausible spelling as accurate. The first and third authors agreed on the phonologically plausible acceptable responses for the 36 pseudowords (for example the item <clep> was phonologically plausible if spelled as <clepp> and <klep>).

Items were matched across the three subtests on word length, and across regular and irregular words on zipfrequency and zipf contextual diversity (see ~~Authors,~~Niolaki et al., 2022[1], for details). The reliability of each subtest is high based on the accuracy scores of the binary assessment: irregular words *α=.97,* regular words *α=.96,* and pseudowords *α=.94.* Each spelling response was scored using binary, non-binary and categorical measures. Table 2 provides examples of scoring with the different measures.

Accuracy

Children's spellings were given a score of 0 for incorrect ~~spellings~~ and 1 for correct spellings. The maximum possible accuracy score was 106.

<Table 2 here>

Orthographic Measure

Letter distance (LD) LD was calculated using Ponto (available online at http://spell.psychology.wustl.edu/ponto/ (Kessler, 2009)). This online tool allocates points for each deletion, addition, transposition or substitution needed for the child's written response to be transformed into the conventional spelling. Mean distance scores were generated for each child ~~and~~ for regular words, irregular words and pseudowords separately.

Phonological Measures

Phonological Plausibility (PhP) An error was given a score of 1 if it was phonologically plausible and 0 if there were incorrect phoneme-grapheme correspondences

(phonemic error), or if additional elements were present or absent. Errors were calculated as PhP or not by the second author and agreed by the first and third authors.

Automated Measures for Phoneme Representation (AMPR). The AMPR scoring metric comprises the number of phonologically plausible phonemes in a word divided by the total number of phonemes. The AMPR calculates a lower score when phonologically implausible errors are made, for example, *life* spelled as *live* would receive a score of .66 (2/3 phonemes correctly represented) while *life* spelled <*life*> would receive a score of 1 (3/3 phonemes correctly represented, the highest score). The measure was generated using the online software (available at http://spell.psychology.wustl.edu/AMPR), yielding scores between 0 and 1, where scores nearer 1 represent a phonologically plausible response and scores closer to 0 represent a non-phonologically plausible response (Treiman & Kessler, 2004) (see Table 3). Mean scores for each word type per child were calculated.

Procedure

The data were collected in Spring/Summer term by researchers and trained research assistants testing children in small groups in urban and rural primary schools (at least three research assistants supervised the children, and the class teacher also helped during the procedure). For details onf the procedure, please see Authors Niolaki et al. (2022). Error analysis was computer-scored or hand-scored, as described above, by the authors. In the case of scoring by hand, categorisations were discussed and agreed upon between the first, second and third authors. Agreement between authors had to reach 100% for hand-scored variables.

Results

Total correct scores were computed for all participants for regular words, irregular words and pseudowords. Average scores for AMPR and LD were computed. Means and standard deviations per measure and word type are presented in Table 3.

<Table 3 here>

A two-way ANOVA was conducted to explore Key stage group differences in accuracy. The between groups variable was Key stage group (F/KS1 (Reception Year to Year 2), EKS2 (Year 3 and Year 4), AKS2 (Year 5 and Year 6)), and the within groups variable was word type (regular word, irregular word, pseudoword). There was a significant interaction between word type and Key stage groups (F (4, 1240) = 76.44, $p<.001$, $\eta_p^2=.198$). While pseudowords were spelled more accurately than irregular words in F/KS1 and EKS2, this reversed in AKS2 (see Figure 1a). Regular words were the most accurate in all key stages. There was a significant main effect of word type (F (2, 1240) = 402.7, p < .001, $\eta_p^2=.394$) – the children were more accurate in spelling regular words (M=23.8) than irregular words (M = 19.9) and pseudowords (M=17.9) ($p<.001$). There was also a significant main effect of group (F (2, 620) = 423.6, p < .001, $\eta_p^2=.577$), with AKS2 children having spelling significantly more items accurately overall ~~correct spellings in all word types~~ (M= 28.0) than EKS2 (M=23.0) and F/KS1 (M=10.6) (all p_s < .001). The difference between EKS2 and F/KS1 was also significant.

<Insert Figure 1 here>

The same analysis was carried out for the PhP, AMPR and LD scores (see Figures 1b, c, 1d for plots of the mean scores, ~~respectively~~). For PhP ~~scores~~ the scores were for regular and irregular words only as all phonologically plausible responses for pseudowords were counted as correct. ~~,t~~ The results of the two-way ANOVA revealed a significant interaction of word type ~~(regular, irregular only as pseudowords do not have a PhP score)~~ and Key stage group ~~was significant~~ (F (2, 601) = 23.3, $p<.001$, $\eta_p^2=.07$). ~~PhP scores were higher for irregular than regular words in each key stage.~~ The main effect of word type was significant

$(F (1, 601) = 341.9, p < .001, \eta_p^2=.36)$ as children made more PhP errors for irregular (M=42.1) than regular words (M=24.1). The main effect of group was significant $(F (2, 601) = 125.8, p < .001, \eta_p^2=.29)$. The difference between F/KS1 (M=46) and EKS2 (M=23) and AKS2 (M=11) was significant, as was that between EKS2 and AKS2. between all three groups (M = 46, M= 23 and M= 11 for F/KS1, EKS2, AKS2, respectively).

For the AMPR scores, the interaction of word type and key stage group was significant $(F (4, 1162) = 5.37, p<.005, \eta_p^2=.018)$. AMPR scores increased from F/KS1 to AKS2 for each word type but except pseudowords. In the case of pseudowords, whose scores increased from F/KS1 to EKS2 only. There was a significant main effect of word type $(F (2, 1162) = 64.8, p < .001, \eta_p^2=.100)$ with significantly lower scores for pseudowords (M=.81) than regular words (M=.85) and irregular words (M=.88) (all ps<.01 for all differences). The difference between regular and irregular words was also significant (p <.001). There was also a main effect of group $(F (2, 581) = 65.06, p < .001, \eta_p^2=.183)$. F/KS1 had significantly lower scores (M = .766) thant ESK (M=.875) and ASK2 (M=.905) but the difference between ESK2 and ASK2 did not reach significance.

Finally, for LD, there was a significant interaction between word type and key stage group $(F (4, 1220) = 125.9, p<.001, \eta_p^2=.29)$. LD scores decreased from F/KS1 to EKS2 for each word type. Irregular words had the highest score but in AEKS2 where pseudowords scored higher. There was a significant main effect of word type $(F (2, 1220) = 525.7, p < .001, \eta_p^2=.46)$. There were significantly higher scores for irregular words (M=1.71) than regular words (M=.91) and pseudowords (M=1.21) (all ps < .001). Thise difference was also significant between regular words and pseudowords (p < .001). There was a significant main effect of group $(F (2, 610) = 197.4, p < .001, \eta_p^2=.39)$. F/KS1 had significantly higher scores (M =2.45) thant EKS2 (M=.91) and AKS2 (M=.47) (all p < .001). The difference between EKS2 and AKS2 was also significant (p<.001).

Correlation analyses

Partial correlations were conducted for each KS group, controlling for grade, in order to explore the associations of the phoneme and letter-based measures with spelling accuracy for regular words, irregular words and pseudowords ~~the three types of word~~ for AMPR and LD, and for regular and irregular words for PhP. The results are presented in Table 4.

<Table 4 here>

For the F/KS1 group, significant associations were observed between accuracy for all word types and all measures. For EKS2, all associations were significant, with the exception that AMPR pseudoword scores w~~as~~ere not associated with pseudoword accuracy. Notably, for AKS2, associations with PhP and AMPR for regular words are lower than for the youngest age group. High levels of accuracy could explain these weaker associations in the older age groups. However, LD was consistently strongly correlated with accuracy across the age groups and letter string types.

In order to determine differences in the associations Eid et al.'s (2011) comparison of correlations (online calculator[32]) was used. LD and AMPR were compared, as PhP wa~~is~~s less strongly associated with~~to~~ accuracy according to the results of the correlations ~~conducted~~ (see Table 4).

Comparison of correlations for F/KS1

The relationship between irregular word accuracy and AMPR irregular word score (r=.290) was significantly ~~(z=5.1, p<.001)~~ less strong than that for regular word accuracy and AMPR regular word score (r=.529) (z=5.1, *p*<.001). The association between regular word accuracy and AMPR regular word score (r=.529) wa~~i~~s significantly ~~(z=1.9, p<.05)~~ lower than that for pseudoword accuracy and AMPR pseudoword score (r=.606) (z=1.9, p<.05)**.**

There were no significant differences in the associations between LD scores and accuracy for regular vs. irregular and pseudoword vs. irregular. Unexpectedly, the coefficient for LD regular words and regular word accuracy (r=-.682) was significantly (z=2.4, p<.01) less strong than the coefficient for LD pseudowords and pseudoword accuracy (r=-.751) (z=2.4, *p*<.01).

Comparison of correlations for EKS2

The coefficient for irregular word accuracy and AMPR irregular words (r=.456) was significantly higher than that for regular word accuracy and AMPR regular words (r=.237) (z=3.2, *p*<.001). This is different to what we found for F/KS1 children. The difference between the coefficient for LD irregular words/ and irregular word accuracy and the coefficient for LD regular word and/ regular word accuracy was not significant, indicating reliance on lexical processes for both word types. The coefficient of the association between LD and accuracy for pseudowords was significantly lower than for irregular words (z= 5.86, p <. 001) and regular words (z= 4.87, p < .001).

Comparison of correlations for AKS2

The difference between the coefficient for irregular word accuracy and AMPR irregular words (r=.411) and regular word accuracy and AMPR regular words (r =.187) was significant (z=10.1, *p*<.001). This is consistent with the findings for EKS2 but not for F/KS and this is to be expected due to the consistent sound – letter associations that the regular words have.

The coefficient for LD irregular words and irregular word accuracy (r=-.956) was significantly higher from than that for LD regular words and regular word accuracy (r=-.930), (z= 3.3, *p*<.001). This suggests that for irregular words there is more reliance on lexical processes than for regular words. The coefficient for LD irregular words and irregular word

accuracy was also significantly higher than the coefficient for LD pseudowords and pseudoword accuracy (-.927) (z= 3.3, *p*<.001). This indicates that pseudowords in comparison to irregular words rely less on lexical processes.

Discussion

This study aimed to examine the interactions between key stages, spelling accuracy and measures of sublexical (phonological plausibility: PhP, Automated Measures for Phoneme Representation: AMPR) and lexical spelling processes (Letter Distance: LD) and the associations between them. Spelling was examined in relation to the type of word, regular and irregular words and pseudowords, as they are differentially affected by lexicality. Beginning spellers seem towould rely more on sublexical processes for spelling, and as children gain experience, lexical processes will become more important. It was also expected that spelling of irregular words would be primarily associated with lexically-related variables, while spelling of pseudowords would be associated with sublexically-related variables. For regular words, there should reliance on both lexical and sublexical processing.

In terms of accuracy, similarly to past findings (de Bree & van der Boer, 2019), it was observed that children were the most accurate in spelling regular words, for all key stages. Whilste pseudowords were spelled more accurately than irregular words up to the second half of KS2, the AKS2 children were more accurate with irregular words than pseudowords. This suggests that as they gain more experience with spelling and reading, children rely more on orthographic (lexical) processes. This is supported by the significant interaction between KS and LD, with improving scores for LD for all letter string types from early KS to AKS. The improvement for irregular words was such that they outperformed pseudowords at AKS2. The results of the present study are in line with the findings of former studies which indicated that spelling development is continuous rather than stage driven, reflecting gradual

improvements in children's phonological and orthographic knowledge (Carvalhais et al., 2021; de Bree & van der Boer, 2019; McMurray, 2020).

AMPR also significantly improved from early to advanced KS for regular and irregular words. Pseudowords improved between F/KS1 and EKS2 but remained similar between EKS2 and AKS2, h. Highlighting that as children put more effort ionto applying lexical processes at AKS2, non phonologically implausible errors decrease more for irregular and regular words than pseudowords. This implies that for real words, phonological plausibility becomes strongly linked to lexical processing and automatic retrieval, which is also indicated by stronger associations between AMPR and accuracy as KS advances, when for pseudowords these associations become weaker. Another marker of sublexical processing, PhP, confirmed the regular words superiority over irregular oneswords, as irregular words haved consistently more phonologically appropriate errors than regular oneswords. However, this result does not capture the subtle changes and improvements as shown by the AMPR and KS interaction effects as AMPR can be used for pPseudowords. The results endorsed past research suggesting that non-binary measures are good metrics with which to monitor spelling development (Masterson & Apel, 2010; Werfel & Krimm, 2015).

To further confirm this, the strength of the relationships between spelling accuracy and phoneme- and letter-based measures differed as a function of KS and the type of letter string. Across KS groups, the strongest associations were found with LD, and less so for PhP (irregular and regular words) and AMPR. The discrepancy in the strength of associations between LD and phoneme-based measures might be also explained by the use inclusion of multisyllabic words, thatwords in the spelling test, that need more effort at the whole-word level (Heggie &Wade-Woolley, 2017) and the different types of letter stings that have differential reliance on lexical and sublexical processing. This may explain the difference to

Treiman et al. (2016; 2019), who found that the strongest predictors of spelling were accuracy followed by PhP and then LD.

For the early F/KS1 spellers the association between sublexical processes (i.e., AMPR) and accuracy for regular words and pseudowords iwas stronger than that for irregular words. This is expected as irregular words have less predictable letter-sound associations. At EKS2 the association between AMPR pseudoword and pseudoword accuracy was non-significant and smaller in comparison to the coefficients for regular and irregular words. This may suggest that at this stage knowledge of how the word is spelled is more important than sublexical phoneme-graphemePGC knowledge, which, by this KS level reaches a plateau.

Similarly, at AKS2, the association of irregular words accuracy association with AMPR wais stronger than the association forof regular word accuracy and AMPRs. Whilest the influence of phonics in Key Stage 2 becomes less pronounced, for irregular words sublexical processes can be still important due to the inconsistent PGCphoneme-grapheme correspondences. At KS2 there is some more reliance on lexical processes (reflected in LD scores) for irregular words than for pseudowords, which is consistent with the findings for F/KS1. It is also noteworthy that AMPR and PhP show consistently significant and fairly similar in strength associations for regular and irregular words in all KS. This finding confirms that the two variables tap the same construct.

At F/KS1 LD iwas more stronglyer associated with accuracy for pseudowords than regular words. This is likely because regular words benefit from being able to draw on both lexical and sublexical processes so will be the closest to the target spellings. In the case of pseudowords, they need to be spelled using phoneme-grapheme correspondences (PGCs) and we counted any legal alternative PGC as correct (unlike in the case of the words where it must be only the target one for that word in order to be correct). This lenient criterion for the pseudowords means they can be accurate phonologically while having many letters different

from the target. In the case of the irregular words, the children will be trying to use PGPCs but these will be disadvantageous and will lead to a big difference in letters compared to the target.

At EKS2 the difference in the associations between LD and accuracy for regular and irregular words did not reach significance indicating reliance on lexical processes for both word types, demonstrated also by the significantly stronger associations in comparison to the LD- and accuracy association for pseudowords. However, at AKS2 LD-accuracy irregular words associations were stronger than those for both regular words and pseudowords, showing a shift to relying more on lexical processes when spelling irregular words.

AMPR and LD are two non-binary measures that relate to spelling accuracy with all letter string types (except AMPR and pseudoword accuracy at EKS2). The strength of associations varied depending on the group and letter string category. AMPR was consistently less strongly associated with accuracy in comparison to LD. For pseudoword and regular word accuracy this is not an odd finding as for children with more competence in spelling, sublexical processing will be less critical for their spelling due to the straightforward PGCs. For irregular word spelling, the reliance to on lexical processes is strong at later stages, as reflected in the LD scores outperforming LD for pseudowords at AKS2. For FKS1 children there is also less influence of phoneme-based measures (AMPR) than LD for irregular word spelling even if the influence of phonics teaching is strong in this age group. The influence of phonics, although strong, is not the optimal strategy to spell irregular words. Regular words, on the other hand, at all KS, are the most accurate word type as they benefit from input from both lexical and sublexical processes, indicated also by the strong associations between LD scores and accuracy. The findings suggest that there is a gradual amalgamation unification of spelling processes (orthography, phonology and semantics) from KS1 to AKS2 as suggested by the lexical quality hypothesis (Perfetti & Hart, 2002) and the

'linguistic trilogy' that suggests spelling effectively ~~you need~~requires all three processes (Walter & Dilworth, 2013).

Finally, for pseudoword spelling and for F/KS1, there was influence of LD and for EKS2, LD was more strongly associated with pseudoword accuracy than with regular and irregular word ~~spelling~~ accuracy. This may be partly explained to the way pseudowords were devised for the spelling test, which implicates lexical processing. In an attempt to spell pseudowords, that somewhat~~partly~~ resemble real words, children employ ~~more strongly the~~ lexical processes to a greater extent than for regular and irregular words, as the latter two are more easi~~ly~~er recognised. Lexical processing becomes a strategy that can be transferred to other types of letter strings. Another potential interpretation could be that pseudowords need to be spelled using PGCs and we counted any alternative PGC as correct ~~(unlike in the case of the words where it must be only the target one for that word in order to be correct)~~. This lenient criterion for the pseudowords means they can be accurate phonologically while having many letters different from the target.

This important finding agrees with results from the longitudinal study of Carvalhais et al.'s (2021) ~~longitudinal study~~. Those researchers ~~which~~o found for the older cohort of Portuguese learners that orthographic and stress errors were mos~~tr~~e prevalent than phonological errors, indicating that the influence of lexical semantic processes are~~are~~ stronger as phonological skills are by now very well mastered. That means from a school perspective that spelling instruction should foster more orthographic strategies in these~~se~~ later stages of education ~~more the orthographic strategies~~.

Use of non-binary ways to score spelling errors can provide the teachers with a powerful tool to unpick the type of errors children make and if they need more support in phonological and lexical semantic (orthographic elements) processes. The spelling test with irregular and regular words and pseudowords can further uphold the teacher's work in finding

the linguistic gaps that children might have and tailor an appropriate bespoke intervention. It is also evident that in addition to phonological processes, orthographic knowledge plays a significant role in learning to spell. Therefore, an educational system that emphasises solely alphabetic strategies may put children at a disadvantage and it might also be of disservice.

As with any study, this one is not free of limitations. Several variables were generated by computer programs; however, some errors can still be difficult to categorise. Also, the data are cross-sectional while a longitudinal design could be more informative. However, these limitations are minimised by strengths, such as the reliable measures that were used to score children's spellings. PhP has been employed in many earlier studies examining the type of errors children make. The large sample size (N=641) and items in the spelling test (N=106) produced a large number of spellings (- the total number of words including correct answers and no responses exceeded 65,000). To further our understanding of spelling development, researchers could apply a similar method in a cross-cultural study where spellings of native speakers of other orthographies are examined. This will allow us to ascertain whether these findings are universal or not.

Many years of schooling need to take place in order for spellings to become established entries in the mental lexicon. This study is one of the first to demonstrate when the lexical and sublexical processes based on the DR model (Martin & Barry, 2012) start to function in a more integrated and interdependent way. It will be informative for further studies to be conducted with more transparent and opaque orthographies as these can indicate if there is a universal age at which this occurs.

Conclusion

The scoring measures included previously showed primary school children's reliance on lexical and sublexical spelling processes (Caravolas et al., 2001; Treiman et al., 2016; Treiman, et al., 2019). The measures assign credit for partially correct spellings and can allow researchers and educators to capture changes in the development of spelling ability, and to analyse spelling performance (Kessler, 2009; Masterson & Apel, 2010; Treiman & Kessler, 2006).

~~The current study investigated strategies primary school children use when they spell different letter string types; the novel aspect of the study is that error analyses were conducted rather than adopting a dichotomous correct versus incorrect scoring system. This type of analysis is a stronger indicator of the types of errors the children make.~~

The fine-tuned measures employed in calculating the spelling errors, in addition to use of an assessment that differentiates between different word types, means we were able to capture changes as the children become more advanced in spelling. It was observed that as spelling skill increased, and the influence of systematic synthetic phonics ~~is~~ presumably became less strong, children appeared to become more reliant on~~tend to use~~ lexical processes. We believe that our study will inspire more research in spelling which will help unlock the mystery of learning to spell conventionally, and ~~as we say in our title~~highlight that, every letter counts for understanding spelling processes.

Note: [1] A list of the items is in the appendix of the published paper Niolaki et al. (2022)

[2] LD: calculated using Ponto http://spell.psychology.wustl.edu/ponto/ (Kessler, 2009)

[3] Eid et al. (2011): https://www.psychometrica.de/correlation.html#dependent

References

Authors et al. (2019, 2022)

Barry, C. (1994). Spelling routes (or roots or rutes). *Handbook of spelling: Theory, process and intervention,* 27-49.

Barry, C. and Seymour, P. (1988). Lexical Priming and Sound-to-Spelling Contingency Effects in Nonword Spelling. *The Quarterly Journal of Experimental Psychology Section A, 40* (1), 5-40. DOI:10.1080/14640748808402280

Bruck, M. and Waters, G. (1988). An analysis of the spelling errors of children who differ in their reading and spelling skills. *Applied Psycholinguistics 9*(1), 77-92. DOI: https://doi.org/10.1017/S0142716400000461

de Bree, E. and van den Boer, M. (2019). Knowing What We Don't Know: Cognitive Correlates of Early Spelling of Different Target Types. *Reading and Writing 32*(8), 2125-2148. https://doi.org/10.1007/s11145-019-09936-9

Caravolas, M., Hulme, C. and Snowling, M. (2001). The Foundations of Spelling Ability: Evidence from a 3-Year Longitudinal Study. *Journal of Memory and Language 45* (4), 751-774. https://doi.org/10.1006/jmla.2000.2785

Carvalhais, L., Limpo, T., Richardson, U., & Castro, S. L. (2020). Effects of the Portuguese GraphoGame on Reading, Spelling, and Phonological Awareness in Second Graders Struggling to Read. *The Journal of Writing Research, 12*(1), 9-34. https://doi.org/10.17239/jowr-2020.12.01.02

Carvalhais, L., Limpo, T., & Pereira, L. Á. (2021). The contribution of word-, sentence-, and discourse-level abilities on writing performance: a 3-year longitudinal study. *Frontiers in Psychology*, 2924. https://doi.org/10.3389/fpsyg.2021.668139

Clemens, N., Oslund, E., Simmons, L. and Simmons, D. (2014). Aassessing spelling in kindergarten: Further comparison of scoring metrics aand their relation to reading skills. *Journal of School Psychology 52* (1), 49-61. doi: 10.1016/j.jsp.2013.12.005.

Frisby, C. (2016). An empirical comparison of the words spelled correctly and correct letter sequence spelling scoring methods in third- and fourth-grade classrooms. *Journal Of of Applied School Psychology 32*(2), 101-121. DOI: 10.1080/15377903.2016.1151847

Heggie, L., & Wade-Woolley, L. (2017). Reading longer words: Insights into multisyllabic word reading. *Perspectives of the ASHA Special Interest Groups, 2*(1), 86-94. DOI: 10.1044/persp2.SIG1.86

Georgiou, G. K., Torppa, M., Manolitsis, G., Lyytinen, H., & Parrila, R. (2012). Longitudinal predictors of reading and spelling across languages varying in orthographic consistency. *Reading and Writing, 25*, 321-346. DOI: 10.1007/s11145-010-9271-x

Georgiou, G. K., Torppa, M., Landerl, K., Desrochers, A., Manolitsis, G., de Jong, P. F., & Parrila, R. (2020). Reading and spelling development across languages varying in orthographic consistency: Do their paths cross?. *Child development, 91*(2), e266-e279. DOI: 10.1111/cdev.13218

Joye, N., Dockrell, J. E., & Marshall, C. R. (2020). The spelling errors of French and English children with developmental language disorder at the end of primary school. *Frontiers in Psychology*, 1789. doi: 10.3389/fpsyg.2020.01789

Kessler, B. (2009). Statistical Learning of Conditional Orthographic Correspondences. *Writing Systems Research 1* (1), 19-34. https://doi.org/10.1093/wsr/wsp004

Kessler, B., Pollo, T., Treiman, R. and Cardoso-Martins, C. (2012). Ffrequency analyses of prephonological spellings as predictors of success in conventional spelling. *Journal oOf Learning Disabilities, 46*(3), 252-259. https://doi.org/10.1177/0022219412449440

Masterson, J. and Apel, K. (2010). Tthe spelling sensitivity score: noting developmental changes iin spelling knowledge. *Assessment for effective intervention 36*(1), 35-45. DOI: 10.1177/1534508410380039

Martin, D. and Barry, C. (2012). Writing nNonsense: the interaction between lexical and sublexical knowledge in the priming of nonword spelling. *Psychonomic Bulletin & Review 19*(4), 691-698. https://doi.org/10.3758/s13423-012-0261-7

McMurray, S. (2020). Learning to spell for children 5–8 years of age: The importance of an integrated approach to ensure the development of phonic, orthographic and morphemic knowledge at compatible levels. *Dyslexia*, *26*(4), 442-458. DOI: 10.1002/dys.1663

Nag, S., Vagh, S., Dulay, K. and Snowling, M. (2018). Home language, school language and children's literacy attainments: A systematic review oof evidence from low- and middle-income countries. *Review oOf Education*. https://doi.org/10.1002/rev3.3130

Niolaki, G. Z., & Masterson, J. (2012). Transfer effects in spelling from transparent Greek to opaque English in seven-to-ten-year-old children. *Bilingualism: Language and Cognition*, *15*(4), 757-770. doi:10.1017/S1366728911000721

Niolaki, G. Z., Terzopoulos, A. R., & Masterson, J. (2014). Varieties of developmental dyslexia in Greek children. *Writing Systems Research*, *6*(2), 230-256. https://doi.org/10.1080/17586801.2014.893862

Niolaki, G. Z., Vousden, J., Terzopoulos, A. R., Taylor, L. M., Sephton, S., & Masterson, J. (2020). Predictors of single word spelling in English speaking children: a cross sectional study. *Journal of Research in Reading*, *43*(4), 577-596. DOI: 10.1111/1467-9817.12330

Niolaki, G., Vousden, J., Terzopoulos, A., Shepherd, D. L., Debney, L., & Masterson, J. (2022). Spelling predictors; investigating the role of phonological ability and rapid naming in a large cross-sectional British study. *Learning and Instruction*, *80*, 101635. https://doi.org/10.1016/j.learninstruc.2022.101635

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. *Precursors of functional literacy*, *11*, 67-86.

Ritchey, K., Coker, D. and McCraw, S. (2009). A cComparison of metrics for scoring beginning spelling. *Assessment fFor Effective Intervention, 35*(2), 78-88. https://doi.org/10.1177/1534508409336087

Rønneberg, V., & Torrance, M. (2019). Cognitive predictors of shallow-orthography spelling speed and accuracy in 6th grade children. *Reading and Writing*, *32*(1), 197-216. https://doi.org/10.1007/s11145-017-9751-3

Spencer, K. A. (2009). Feedforward,-backward, and neutral transparency measures for British English. *Behavior Research Methods*, *41*(1), 220-227. https://doi.org/10.3758/BRM.41.1.220

Treiman, R. (2018). Statistical learning and spelling. *Language, Speech, and Hearing Services in Schools*, *49*(3S), 644-652. https://doi.org/10.1044/2018_LSHSS-STLT1-17-0122

Treiman, R., & Bourassa, D. C. (2000). The development of spelling skill. *Topics in language disorders, 20*(3), 1-18.

Treiman, R. and Kessler, B. (2004). The case oof case: children's knowledge aand use oof upper- aand lowercase letters. *Applied Psycholinguistics,* 25(3), 413-428. DOI: https://doi.org/10.1017/S0142716404001195

Treiman, R., Kessler, B. and Caravolas, M. (2019). What methods of scoring young children's spelling best predict later spelling performance? *Journal of Research in Reading, 42*(1), 80-96. doi:10.1111/1467-9817.12241.

Treiman, R., Kessler, B., Pollo, T., Byrne, B. and Olson, R. (2016). Mmeasures of kindergarten spelling and their relations to later spelling performance. *Scientific Studies of Reading 20*(5), 349-362. doi: 10.1080/10888438.2016.1186168

Vousden, J.I., Ellefson, M.R., Solity, J.E., & Chater, N. (2011). Simplifying Reading: Applying the Simplicity Principle to Reading. *Cognitive Science*, 35, 34-78. https://doi.org/10.1111/j.1551-6709.2010.01134.x

Werfel, K. & Krimm, H. (2015). Utility of the Spelling Sensitivity Score to Analyse Spellings of Children with Specific Language Impairment. *Australian Journal of Learning Difficulties, 20*(1), 39-53. DOI: 10.1080/19404158.2015.1047871

Wolter, J. A., & Dilworth, V. (2014). The effects of a multilinguistic morphological awareness approach for improving language and literacy. *Journal of Learning Disabilities, 47*(1), 76-85. https://doi.org/10.1177/0022219413509972

Zoccolotti, P., De Luca, M., Marinelli, C. V., & Spinelli, D. (2020). Predicting individual differences in reading, spelling and maths in a sample of typically developing children: A study in the perspective of comorbidity. *PLoS One, 15*(4), e0231937. https://doi.org/10.1371/journal.pone.0231937

Table 1

Number and mean chronological age of participants per age group (standard deviations are in parentheses)

	F/KS1	*EKS2*	*AKS2*
Number of Participants	309	165	167
Age (years)	6.34 (.95)	8.77 (.94)	10.78 (.56)

Table 2

Examples of scoring using all scoring measures

Target word	Child's spelling	Accuracy	PhP	AMPR	LD
Life	*live*	0	0	0.66	1.4
Nature	*nocher*	0	0	0.66	6.2
Flavour	*flaver*	0	1	1	2.4
Aspire	*aspier*	0	1	1	2

Note. PhP= Phonological Plausibility; AMPR = Automated Measure for Phoneme Representation; LD= Letter Distance

Table 3

Means for all scoring measures (standard deviations are in parentheses)

Year_Groups	Type of word	Accuracy%	PhP%	AMPR	LD
	Irregular	17.1 (19.9)	26 (18)	.78 (.19)	3.28 (1.55)
F/KS1	Regular	38.5 (25.9)	16 (13.5)	.77 (.18)	1.86 (1.57)
	Pseudoword	34.4 (22.7)	-	.72 (.17)	2.12 (1.66)
	Irregular	54.8 (23.1)	42.3 (20.8)	.89 (.07)	1.31(.88)
EKS2	Regular	74.2 (18.8)	23.7 (21)	.86 (.09)	.57 (.58)
	Pseudoword	62.7 (17.1)	-	.86 (.26)	.83 (.65)
	Irregular	76.9 (17.2)	61.8 (24.9)	.95 (.04)	.54 (.47)

AKS2	Regular	86.4 (11.4)	34.1 (29.7)	.91 (.06)	.24 (.22)
	Pseudoword	70.1 (12.8)	-	.85 (.07)	.61 (.37)

Note. PhP= Phonological Plausibility; LD=Letter Distance; AMPR = Automated Measure for Phoneme Representation

Table 4

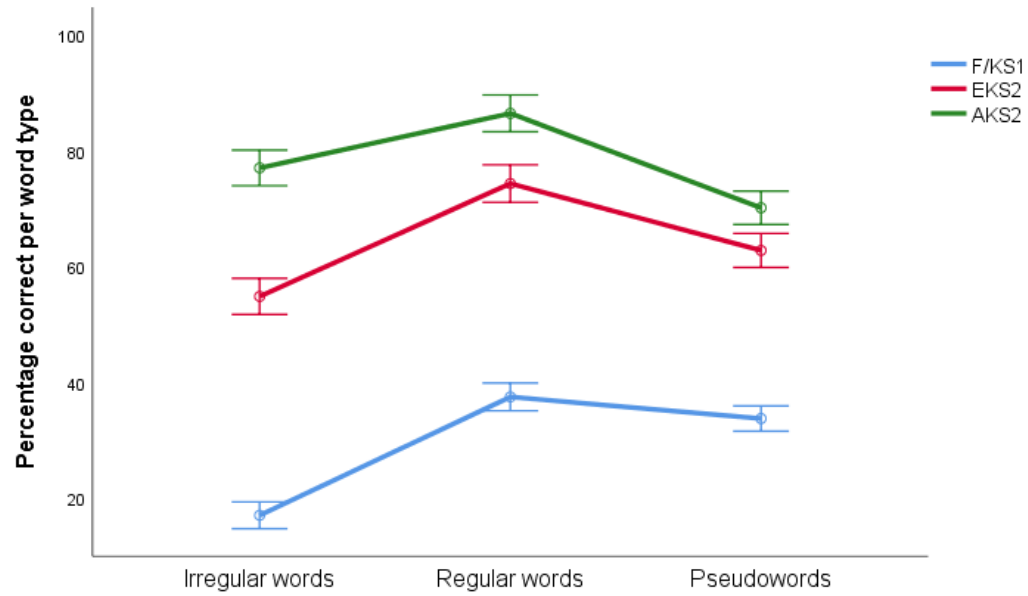*Correlations between spelling accuracy and scoring metrics controlling for year_group*

	F/KS1 Accuracy			EKS2 Accuracy			AKS2 Accuracy		
Scoring Metric	*Ir*	*Rg*	*Pw*	*Ir*	*Rg*	*Pw*	*Ir*	*Rg*	*Pw*
PhP Ir	.391***	-		.435***	-		.446***	-	
PhP Rg	-	.420***		-	.334***		-	.202*	
AMPRIr	.290***	-	-	.456***	-	-	.411***	-	-
AMPR Rg	-	.529***	-	-	.237**	-	-	.187*	-
AMPRPw	-	-	.606***	-	-	-.073	-	-	.170*
LD Ir	-.719***	-	-	-.951***	-	-	-.956***	-	-
LDRg	-	-.682***	-	-	-.943***	-	-	-.930***	-
LD Pw	-	-	-.751***	-	-	-.882***	-	-	-.927***

Note: * *p*<.05, ** *p*<.01, ***p<.001 Ir. =Irregular; Rg.=regular; Pw=Pseudowords; PhP= Phonological Plausibility; AMPR = Automated Measure for Phoneme Representation; LD= Letter Distance
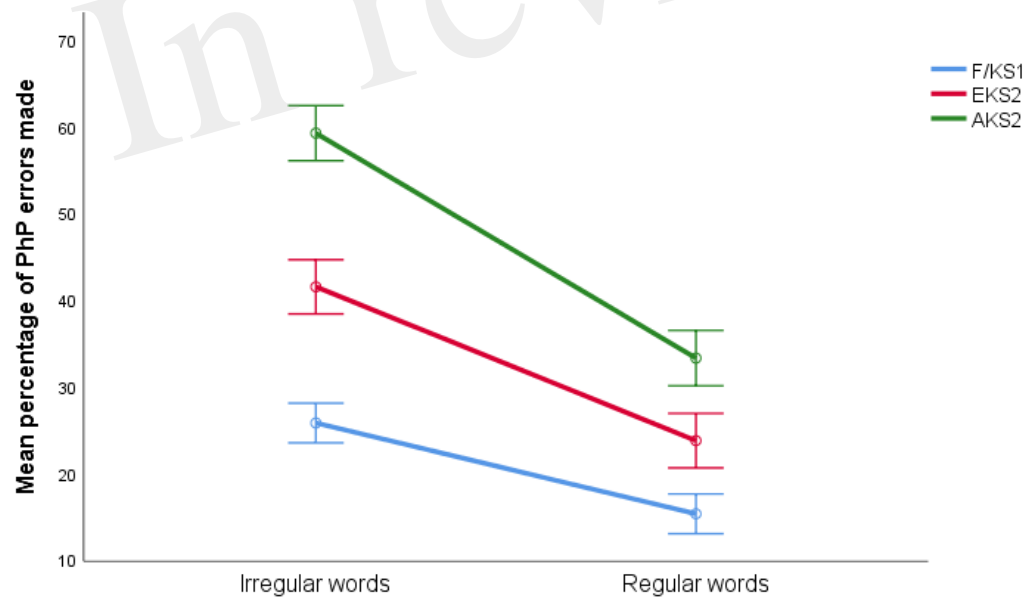
Figure 1

Mean Accuracy, PhP, AMPR and LD score for each word category and group.
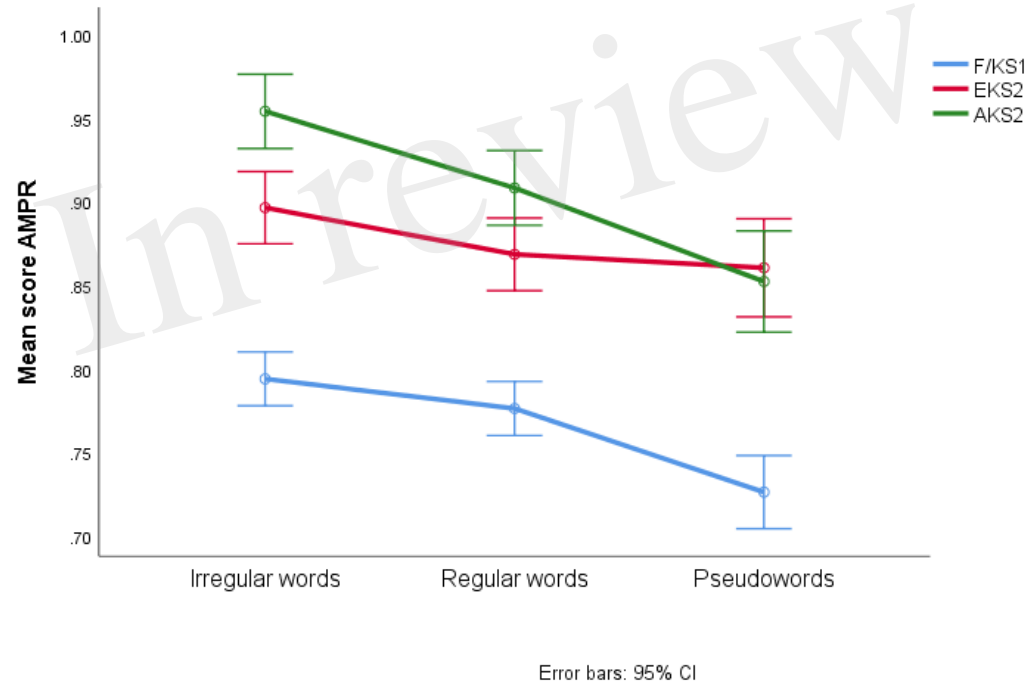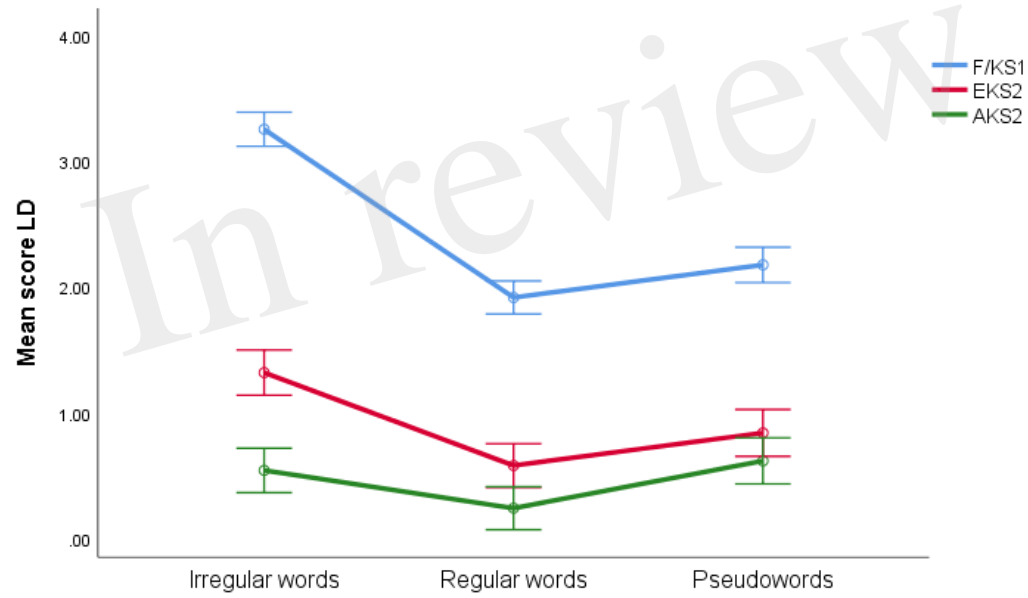
    a. Accuracy

Error bars: 95% CI

b. PhP

Error bars: 95% CI

c. AMPR

Error bars: 95% CI

d. LD

Error bars: 95% CI