AOSIS

Expanding the neighbourhood watch: Orthographic neighbours in isiXhosa reading and spelling

CrossMark
click for updates

Authors:
Paige S. Cox[1]
Tracy N. Bowles[1]

Affiliations:
[1]Department of Linguistics and Applied Language Studies, Faculty of Humanities, Rhodes University, Makhanda, South Africa

Corresponding author:
Paige Cox,
paige9367@gmail.com

Background: Lexical properties such as orthographic neighbours have been shown to have an influence on reading and writing; however, this phenomenon is yet to be explored in the Southern Bantu languages.

Objectives: We investigate the role of orthographic neighbourhood density and neighbourhood frequency in reading and spelling in Grade 3 isiXhosa home-language learners. The aim is to establish whether orthographic neighbours facilitate or hinder reading and spelling. The dual-route model of orthographic processing is used to interpret our findings.

Method: Data were collected from 97 Grade 3 isiXhosa home-language learners. Lexical decision, word reading, and spelling tasks were administered. Tasks included both real and pseudowords varying in orthographic neighbourhood density and neighbourhood frequency, while controlling for word length and word frequency.

Results: Findings show a significant inhibitory effect of orthographic neighbourhood frequency and word length for spelling accuracy. Longer words with higher frequency neighbours were more likely to be spelt incorrectly. No effect was recorded for the lexical decision and word reading tasks.

Conclusion: Our findings suggest that spelling in isiXhosa relies partially on lexical processing strategies, which allows for the inhibitory effects of orthographic neighbours to be observed. While in reading, sublexical processing is more prominent, therefore mitigating the effect of orthographic neighbours.

Contribution: Empirical evidence for the effects of orthographic neighbours, specifically in a Southern Bantu language, provides a clearer picture of the underlying cognitive-linguistic processes involved in reading and writing. This evidence may contribute to the development of targeted pedagogical practices to address spelling errors.

Keywords: Orthographic neighbourhood density; orthographic neighbourhood frequency; neighbourhood effects; word length; word frequency; early literacy; isiXhosa; spelling; word reading; lexical decision.

Introduction

The relevance of literacy research within the South African context is by now well known, as the South African literacy crisis has been acknowledged substantially in academic sources, and widespread in popular media sources (Biesman-Simons et al. 2020; Department of Basic Education 2023). In the last two decades, research efforts have concentrated on the underlying factors of this crisis. This includes studies on macro-social factors such as teaching, and home and classroom literacy practices, as well as on cognitive-linguistic factors such as phonological awareness, morphological awareness and reading fluency (Pretorius 2018; Schaefer, Probert & Rees 2020; Wilsenach 2013; 2019). Contrastingly, in the South African context, little attention has been given to the role of lexical properties such as word length, word frequency, and orthographic neighbours within a cognitive-linguistic approach. These properties are, however, widely acknowledged as playing an important role in cognitive processing involved in reading and writing (Beinborn, Zesch & Gurevych 2016; Ferrand et al. 2018; Kuperman, Schroeder & Gnetov 2024). Specifically, in languages such as English, Dutch and French, orthographic neighbours have been shown to facilitate reading and writing, resulting in shorter reading and writing times and more accurate encoding and decoding (Barnhart & Goldinger 2015; Brysbaert et al. 2015; Ferrand et al. 2018). Alternatively, in other languages such as Spanish, Turkish and Malay, neighbours appear to hinder reading and writing, resulting in longer and less accurate encoding or decoding (Aguasvivas et al. 2020; Erten, Bozsahin & Zeyrek 2014; Yap et al. 2010). To the best of the authors'

Note: Special Collection: Literacy in practice.

knowledge, only two studies have started to look at lexical properties, such as orthographic neighbours, in Southern Bantu languages, and in isiXhosa specifically (Berghoff 2023; 2024). In her research, Berghoff (2023; 2024) describes the derivation of lexical properties that impact word recognition, including lexical and sublexical frequencies and orthographic neighbours in isiXhosa, but she does not present any experimental evidence for the effect of these properties on literacy. We conducted a quantitative cross-sectional analysis to determine the unique effects of the lexical properties of words, specifically orthographic neighbours, for both reading and spelling in Grade 3 isiXhosa home-language learners.

Orthographic neighbours and neighbourhood effects

Orthographic neighbours are words that share similar spelling patterns, for example the words *chest, cheater,* and *heat* are considered orthographic neighbours of the word *cheat*. Coltheart et al. (1977) first defined orthographic neighbours, somewhat stringently, as words which differ by one letter while maintaining letter positions as well as word length, for example *heat* and *meat* would be considered neighbours, but not *heat* and *cheat*. This rigid definition has evolved to a more flexible definition in recent literature which considers orthographic neighbours on a graded scale of similarity (Yarkoni, Balota & Yap 2008). In such definitions, words with different word lengths (formed through letter additions or deletions) such as *plan* and *planet*, as well as letter transpositions such as *trial* and *trail* (Yarkoni et al. 2008), are considered to be orthographic neighbours.

Research on orthographic neighbours, from a psycholinguistic perspective, is primarily interested in measuring the effect that word similarity has on reading and writing. To this end, researchers focus on the neighbourhood characteristics of words, called orthographic neighbourhood effects. This study examines two orthographic neighbourhood effects, namely neighbourhood density and neighbourhood frequency, as well as their interaction with reading and spelling in isiXhosa.

Orthographic neighbourhood density

Orthographic neighbourhood density is a measure of the average similarity of a word's neighbours to the word itself. The Orthographic Levenshtein Distance 20 (hereafter OLD20) metric (Yarkoni et al. 2008) can be used to calculate orthographic neighbourhood density, and was the measure adopted in this study. The OLD20 is calculated by taking the average Levenshtein distance of the 20 closest neighbours to a target word. Levenshtein distance is a computer science metric which measures the difference between two strings (words) by characters (letters). For example, in isiXhosa *ubanzi* (rock-hyrax) and *amanzi* (water) have a Levenshtein distance of 2, or in other words, they differ by two letter substitutions. Importantly, a lower OLD20 number reflects a denser neighbourhood, with a higher OLD20 number reflecting a sparser neighbourhood. For example, the word

ingonyama (lion) has an OLD20 value of 1.75. This means that, on average, the neighbours of the word *ingonyama* differ from this target word by 1.75 letters. In contrast a word such as *intliziyo* (heart) has a very sparse neighbourhood, with an OLD20 of 4. This means that there are few words similar to *intliziyo* in isiXhosa. For example, one of the closest neighbours for *intliziyo* would be *intlanzi* (fish) which has a Levenshtein distance of 4. Levenshtein distance has also been used in other applications in linguistics, for example as a measure of general orthographic similarity across dialects and languages (Zulu et al. 2008).

Most studies on neighbourhood density report facilitatory effects for reading (words which share many letters are read faster) and this is somewhat consistent across languages (cf. Barnhart & Goldinger 2015; Lim 2016; and Parker et al. 2021 for English; Boot & Pecher 2008 and Brysbaert et al. 2015 for Dutch; Ziegler, Perry & Coltheart 2003 and Ferrand et al. 2018 for French). This facilitatory effect is observed in words with dense neighbourhoods, which are read faster and more fluently (i.e. with a reduced number of errors). Similar findings have been reported for spelling, although fewer studies have focused on spelling. For example, Khanna et al. (2023) report a significant facilitatory effect of orthographic neighbourhood density for spelling accuracy in English adults, and Roux and Bonin (2009) report facilitatory effects for spelling speed in French adults. Words with dense neighbourhoods are spelt faster and more accurately than those with sparser neighbourhoods. In writing more generally, research findings suggest a similar interaction, for example Scaltritti et al. (2016) in their study on Italian found that interkeystroke intervals, when typing, were faster for words with denser orthographic neighbourhoods.

Research in less commonly studied languages on this topic, such as Turkish, Malay and Greek, report contradicting results. For example, Erten et al. (2014) found an inhibitory effect of neighbourhood density (words that share many of the same letters are slower to read) for Turkish in a lexical decision task. The authors suggest that this finding may be a result of the agglutinative morphology of Turkish, 'where unique non-stem combinations (unique suffix groups) can be quite large' (Erten et al. 2014:4). In other words, many words in Turkish share a root form, and differ by unique suffix combinations. For example, *arabalar* (cars), *arabam* (my car), *arabamiz* (our car), and *arabada* (in the car) (Oflazer, Göçmen & Bozsahin 1994). These 'morphological neighbours' appear to compete for lexical access when reading in Turkish, resulting in inhibitory effects (Erten et al. 2014). This is supported by other studies on agglutinative languages such as Malay (Yap et al. 2010). We hypothesise that a similar effect may be found for isiXhosa due to the agglutinative nature of the language. In isiXhosa, like in Turkish, many words share a root and differ by both unique suffix combinations, for example *hamba* (walk or go), *hambisa* (to deliver), and *hambela* (to visit), as well as unique prefix combinations as in *ukuhamba* (to walk or go), *ihambo* (journey), *umhambi* (traveller), and *abahambi* (travellers).

In comparison, in inflectional languages such as Greek, null orthographic neighbourhood density effects have been reported. Kapnoula et al. (2017) explain this finding as owing to the prevalence of long words in Greek as a result of its inflection. The authors explain that the longer words result 'in much sparser neighbourhoods than English, especially if inflectional variants are excluded, and therefore, fewer opportunities for neighbourhood effects' (Kapnoula et al. 2017:9). Such an argument could also be applied to isiXhosa, which has many long words owing to its conjunctive orthography, where one orthographic word can stand for an entire sentence. For example, the word *ndiyabafundisa* (I teach them).

Orthographic neighbourhood frequency

Orthographic neighbourhood frequency is understood as the relative frequency of a word's neighbours. Specifically, one takes the average word frequency score of the 20 closest neighbours of a target word. The resulting metric is coined OLDF (Orthographic Levenshtein Distance Frequency) (Balota et al. 2007).[1] We adopted OLDF as a measure of orthographic neighbourhood frequency in this study.

Findings for the effects of orthographic neighbourhood frequency in reading and writing are less clear cut than that of orthographic neighbourhood density, particularly for English, with different studies reporting facilitatory (Siakaluk, Sears & Lupker 2002), inhibitory (Newman, German & Jagielko 2017), and null (Huntsman & Lima 2002) effects. Interestingly, for many other languages, including Dutch (De Moor & Brysbaert 2000), French (Dujardin & Mathey 2022), and Spanish (Acha & Perea 2008), robust inhibitory orthographic neighbourhood frequency effects are reported for reading.

Much like for orthographic neighbourhood density, the focus on neighbourhood frequency in spelling is significantly less than that on reading. In their study on French spelling, Roux and Bonin (2009) report null orthographic neighbourhood frequency effects. Other studies on neighbourhood effects in spelling have looked at phonographic neighbours (words which differ by graphemes and phonemes). For example, Lété, Peereman and Fayol (2008) report inhibitory effects for words with a high phonographic neighbourhood frequency in French children. Words with higher frequency neighbours were spelt incorrectly.

Very little explicit explanation is given for the variation in findings across languages for both neighbourhood effects, which speaks to a need for cross-linguistic reviews of these effects. Some mention is made of this issue in Andrews's (1997) review of neighbourhood effects. Andrews posits that the inhibitory orthographic neighbourhood frequency effect in languages like Spanish and French, but not English, is owed to the orthographic transparency of the former languages in comparison to English. However, this argument is more relevant in explaining a facilitatory orthographic

---

1.The ELP was updated in 2019 to include the OLDF and OLD20 metrics. Available at: https://elexicon.wustl.edu/about.html

---

neighbourhood density effect in English. The lack of orthographic transparency in English, combined with the prevalence of body units in the language in words such as '-at' in *cat, mat, rat,* and *sat,* means that readers of English can benefit from body neighbours, since they do not need to rely on grapheme–phoneme decoding (Andrews 1997). In French and Spanish, these body neighbours are not as frequent. This, however, does not provide evidence for why facilitatory orthographic neighbourhood density effects are still reported in these languages; nor does it explain why readers of English do not encounter competition from higher frequency neighbours, whereas readers of French and Spanish consistently experience these inhibitive effects. If we were to follow Andrews's (1997) reasoning, we should expect to see inhibitory neighbourhood frequency effects for isiXhosa, but facilitatory neighbourhood density effects.

Perhaps another explanation for the contradictions cross-linguistically is methodological. Specifically, the majority of studies pre-2008 and even more recently (cf. Berghoff 2023, 2024) use Coltheart's N (which considers neighbours on the basis of one-letter substitutions only, and sometimes as one-letter additions and deletions) to measure orthographic neighbourhood density, and consequently orthographic neighbourhood frequency. However, this metric limits the psychological effects of neighbours to a one-letter radius, rather than acknowledging that the effect of word similarity is more gradient. This argument is the foundation for the development of the alternative OLD20 density metric by Yarkoni et al. (2008). This is especially true in isiXhosa, where neighbours are likely to differ by two or more letters, owing to the way in which words are formed by attaching affixes to root forms. For example, neighbours for the word *ufunda* (he/she reads) would include *bafunda* (they read), *umfundi* (student) *umfundisi* (priest) which all share the root *-fund*. Thus, the Coltheart's N metric is not well suited to measure neighbourliness in agglutinating languages (such as isiXhosa), nor arguably in highly syllabic languages like French and Spanish, where words differ by one or more syllables (Andrews 1997).

Our study on isiXhosa constitutes a unique contribution to the field on orthographic neighbourhood effects, because: (1) the rich morphology of the language means that it shares features with some less-studied languages such as Turkish, Malay and Greek, meaning that it adds to an understanding of these effects in more synthetic languages; and (2) isiXhosa has a very transparent orthography and is mostly feedback (sound to spelling) and feedforward (spelling to sound) consistent, even more so than French and Spanish, and so this research will add to our understanding of how orthographic consistency interacts with orthographic neighbourhood effects.

Theoretical framework: The Dual-Route Model

Neighbourhood effects in reading and spelling are explained as the activation of word-level linguistic knowledge when

decoding or encoding written text (Parker et al. 2021). Different conceptual models have been put forward to predict these effects, which attempt to capture how orthographic information is organised and processed in the mind.

One of the most prominent theoretical frameworks in psycholinguistics is the dual-route model (Coltheart 1978; Morton & Patterson 1980). This model is of particular interest to our study, as it is used to explain both visual word processing (reading) and auditory word processing (spelling) (Coltheart et al. 2001; Houghton & Zorzi 2003). The dual-route model's architecture posits two processing routes for both visual and spoken words: a sublexical route and a lexical route. However, more recent renditions of the model suggest an additional interactive mechanism between the sublexical and lexical levels, which has led to a connectionist dual-route framework for reading and spelling (Folk, Rapp & Goldrick 2002; Rapcsak et al. 2007; Rapp, Epstein & Tainturier 2002). In the case of reading, the sublexical route involves the decoding of letters, using grapheme-to-phoneme mapping rules which may or may not activate relevant lexemes in either the orthographic or phonological lexicon. For example, when reading the word *intle* (beautiful), this may be broken down into graphemes {i, n, tl, e} before being mapped onto corresponding phonemes /i, n, tɫ, e/ and consequently read aloud. Importantly, this route does not necessitate accessing the whole word form – either orthographic or phonological – before reading aloud can occur. This is particularly true for languages with transparent orthographies, such as isiXhosa, where graphemes are mostly consistently mapped onto phonemes (Probert & De Vos 2016). An interaction between lexical and sublexical information is, however, still possible in the sublexical route, as indicated in Figure 1 by the double-sided arrow.

On the other hand, the lexical processing route involves the mapping of visual words onto a lexical entry in the orthographic lexicon on the basis of semantic specifications, which are activated from the orthographic lexeme (Rapcsak et al. 2007). Here there is a back-and-forth interaction of lexical and sublexical information between lexemes and letters which leads to the activation of a specific lexical item in the orthographic lexicon. This is then mapped directly onto the corresponding phonological lexeme. Additional back-and-forth interaction ensues between phonological lexemes and phonemes until an output is produced in the form of a spoken word. For example, the word *intle* would be mapped onto an orthographic lexeme, which encompasses the abstract orthographic form of the word {intle}; this is achieved with the aid of sublexical information from the letter level. This would then be mapped onto the abstract phonological form of the word /intɫe/ with the aid of sublexical information from the phoneme level. Finally, the word *intle* is read aloud.

In the case of spelling, the dual-route model proposes the same two processing routes as for reading, albeit in reverse. A spoken word is processed either along the sublexical route through the use of phoneme to grapheme mappings, or through the lexical route, where a phonological lexeme is activated and mapped onto its corresponding orthographic form until an output is produced, that is a written word.

As much as the dual-route model may suggest that reading and spelling are two sides of the same coin, there are some key differences which underlie these two literacy skills. Most noticeably, researchers have noted that spelling requires a more advanced knowledge of phoneme-grapheme mappings than reading (Conrad 2008; Schaars, Segers & Verhoeven 2017). Reasons given for this are due to the production component of spelling, which requires the speller to not only recognise phonemes and graphemes, but also apply and reproduce this knowledge. For this reason, spelling is seen as a more advanced literacy skill (Schaars et al. 2017).

A challenge in the literature on orthographic neighbours specifically is attempting to capture orthographic neighbourhood effects within a theoretical model of reading and spelling. There have been multiple theoretical explanations put forward to explain the contradictory findings in the field, but no consensus has been reached so far. It is generally agreed, however, that orthographic neighbours are active at the lexical level of processing, that is, when looking up a word in the lexicon, candidate sets of words are activated, based on their similar spelling (Meade, Grainger & Declerck 2021). In the dual-route model in Figure 1, these neighbour activations would take place within the orthographic lexicon. Facilitatory effects are explained as the strengthening of interactions between lexical and sublexical information as a result of neighbours. In other words, neighbours in the orthographic lexicon send feedback to corresponding graphemes at the sublexical level, resulting in facilitatory effects such as increased reading or spelling speed and accuracy (Roux & Bonin 2009). For example, reading the word *intle* (beautiful) would activate other orthographic lexemes
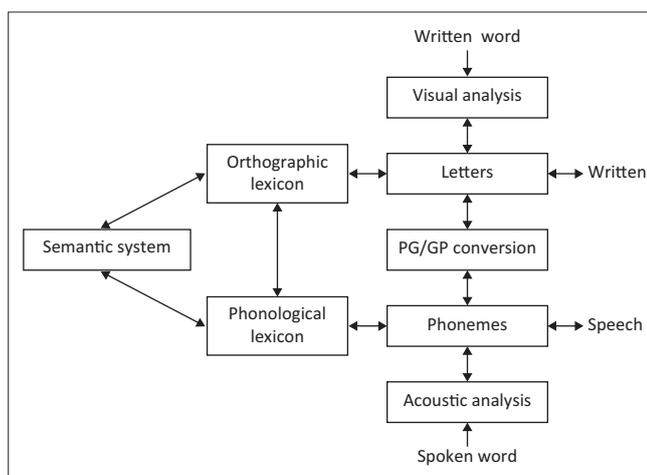
Source: Rapcsak, S.Z., Henry, M.L., Teague, S.L., Carnahan, S.D. & Beeson, P.M., 2007, 'Do dual-route models accurately predict reading and spelling performance in individuals with acquired alexia and agraphia?', *Neuropsychologia* 45(11), 2519–2524. https://doi.org/10.1016/j.neuropsychologia.2007.03.019

PG, phoneme – grapheme; GP, grapheme–phoneme conversion.

FIGURE 1: Dual-route cognitive model of reading and spelling.

such as *zintle* (they are beautiful), *intlanzi* (fish)*,* and *indlebe* (ear). These word neighbours would provide feedback activation to those letters that are shared with the target word, that is {*i, n, tl, e*}, increasing the speed and accuracy with which a word is read or spelt. In comparison, a word with fewer neighbours would not receive the same top-down strengthening.

Contrastingly, inhibitory neighbourhood effects are explained as a lateral inhibition effect at the lexical level, which occurs when similar orthographic forms compete for lexical access (Parker et al. 2021). Specifically in the case of inhibitory neighbourhood frequency effects which are often reported in the literature (Acha & Perea 2008; De Moor & Brysbaert 2000; Dujardin & Mathey 2022), this finding is explained as an outcome of 'intralevel inhibition between the lexical units of the model (which) delay the activation of a word with higher frequency neighbours' (Sears, Hino & Lupker 1999:222). In other words, as a target word is activated in the orthographic lexicon, its higher frequency neighbours are activated simultaneously. These neighbour activations will then need to be inhibited before lexical access can occur, which results in longer response time latencies (De Moor & Brysbaert 2000). In the case of the word *intle*, if any of the neighbours have a higher frequency than the target, this could potentially result in competition in the lexicon, resulting in slower and less accurate reading or writing.

Present study

In this study, we investigate the role of orthographic neighbourhood density and neighbourhood frequency on reading and spelling in Grade 3 isiXhosa home-language learners. The aim is to establish whether orthographic neighbours facilitate or hinder reading and spelling, and to situate these findings within the dual-route model of orthographic processing. By addressing these questions, the study aims to make practical contributions to understanding orthographic neighbourhood effects in isiXhosa, which can help to inform linguistic theory as well as pedagogical practices.

Research methods and design
Participants

Data were collected from 97 Grade 3 isiXhosa home-language learners from five schools in Kwanobuhle township in the Eastern Cape province. All five schools were Quintile 3 schools, that is they are no-fee schools and rely on government funding. The schools also have a feeding scheme, which provides children with a meal at school. The selection criteria for the schools required that all schools used isiXhosa as the Language of Learning and Teaching (LoLT) in the Foundation Phase (Grades 1 to 3), and that they were located within the same geographical area. The selection criteria for participants were that learners were present at the schools for all administered assessments at the time of data collection in February 2022. Grade 3 learners were selected because the South African language in education policy encourages the use of home language instruction up to Grade 3, whereafter a switch is often made to the use of English. Given the focus on isiXhosa reading and spelling, we sought to assess learners prior to them making the switch to English as the language of instruction.

Instruments and procedures

To test whether orthographic neighbourhood density and orthographic neighbourhood frequency had a relationship with reading and spelling, we developed a database of orthographic neighbours in isiXhosa. This then informed the stimuli for the reading and spelling instruments used in our research. The database of orthographic neighbours in isiXhosa was drawn from an isiXhosa corpus by Rees and Randera (2017). The corpus consisted of over 100 000 tokens from multiple Foundation Phase reading sources, including African Story Book, Nal'ibali, Department of Basic Education readers, Department of Education readers, Bible stories, and Story Weaver. These sources are aimed at children who are in the learning to read phase, and as such the resulting corpus is well suited for our study. The orthographic neighbourhood database is available on request.

From this corpus, we compiled a database of 30 isiXhosa words, varying in orthographic neighbourhood density and orthographic neighbourhood frequency. Word length in letters and syllables, as well as word frequency were controlled for. Word frequency was measured as the raw frequency of hits in the corpus. We also created 30 isiXhosa pseudowords (e.g. *nokhube, ucukela, hlenama*). These were chosen as stimuli for the reading and spelling instruments, and all pseudowords conformed with the orthographic and phonological conventions of the language. The advantage of using pseudowords is that they mitigate the influence of word frequency, since the words do not exist in the language, and as such have no semantic influence on results. Orthographic neighbours were computed for each pseudoword from the Rees and Randera (2017) corpus of real words. We used the Visual Basic for Application (VBA) module developer in MS Excel to calculate Levenshtein distance for the neighbours. The resulting orthographic neighbourhood database was 30 real and 30 pseudowords, each with 20 neighbours and corresponding neighbourhood statistics for orthographic neighbourhood density and orthographic neighbourhood frequency. The descriptive statistics for the orthographic neighbourhood database are presented in Table 1.

TABLE 1: Descriptive statistics for the lexical features of words and pseudowords (*n* = 60) in the orthographic neighbourhood database.

Lexical features	Mean	SD	Median	Min	Max
Word frequency	63.60	125.07	24.00	1.00	66.00
Word length (letters)	7.22	1.22	7.00	5.00	10.00
Word length (syllables)	3.27	0.45	3.00	3.00	4.00
Orthographic neighbourhood density	2.60	0.64	2.55	1.60	4.00
Orthographic neighbourhood frequency	22.10	25.03	13.75	4.40	115.90

Note: Only real words were considered in the statistics for word frequency.

SD, standard deviation.

This database informed the test instruments for reading and spelling. Word reading, lexical decision, and spelling tasks were designed and administered to participants over 3 days. In addition, an oral reading fluency (ORF) task was administered with the learners. We report on the ORF data in order to gauge the reading fluency of the sample. All instruments were administered in the learners' home language, isiXhosa, by trained isiXhosa home-language research assistants. Task-specific design, procedures and data coding are discussed under each task.

Word reading task

We developed an original untimed word reading task to measure word reading accuracy in isiXhosa, without interference from reading speed. Research assistants assessed learners individually. Each learner was presented with laminated flashcards, one at a time, each with an individual word printed on it. The assistant then asked the learner to read the words out loud. In total there were 20 flashcards: 10 of the cards had real isiXhosa words and 10 had pseudowords. All 20 words were selected from the orthographic neighbourhood database, while ensuring a range of orthographic neighbourhood density and orthographic neighbourhood frequency scores across words for the task. The research assistants captured any occurrences of errors on Tangerine, an open-source data collection software. Incorrect responses were coded as 0 and correct responses as 1. The task was found to be reliable as measured by Cronbach's alpha: 0.97.

Lexical decision task

The lexical decision task was an originally developed instrument. We designed it using Psychopy builder (Pierce et al. 2019), an open-source experimental design software. The task was run online using Pavlovia.org and administered to learners individually on touchscreen tablets. For the task, a word would appear on the screen and the learner would have to decide whether it constituted a real isiXhosa word or not, by tapping the appropriate button (a tick for if they thought it was a real word, and a cross if they thought it was not a real word). An example of a pseudoword stimulus is given in Figure 2. Learners were given the opportunity to practise on four words before beginning the task, which consisted of 20 possible words (10 real words, and 10 pseudowords). The response times for each word were measured automatically in seconds from when a word appeared on the screen, until the learner
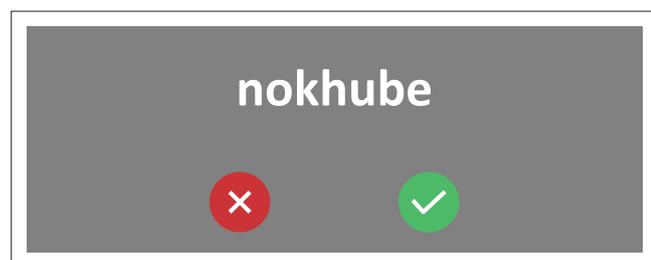

FIGURE 2: An example pseudoword stimulus from the lexical decision task.

tapped a button. The reliability of the task, as measured by Cronbach's alpha, was 0.95.

Lexical decision tasks measure the latency of lexical access, in other words, the time taken from when a child sees a written word, to when they access (or fail to access) the word in their lexicon (Balota et al. 2004). The lexical decision task has received some criticism concerning the additional decision-making component, which is both cognitively demanding and unrelated to the lexical access process (Balota & Chumbley 1984; Carreiras, Perea & Grainger 1997). In spite of these criticisms, the lexical decision task remains a well-used tool in the literature for measuring neighbourhood effects on word reading (Aguasvivas et al. 2020; Brysbaert et al. 2015; Ferrand et al. 2018; Lim 2016; Marian 2017; Parker et al. 2021). Our study includes a lexical decision task to allow for a comparison of our findings to previous studies on neighbourhood effects. Further, the use of a word reading task, to measure accuracy of lexical access, compliments the possible shortfalls of the lexical decision task.

Spelling task

Spelling accuracy was assessed using an originally developed handwritten spelling test. Each learner was given a worksheet with 20 spaces to write 20 words. The research assistant read each word out loud, and the learners were asked to write these down. As in the other two tasks, there were 10 real isiXhosa words and 10 pseudowords in a random order. The task was group administered; learners were seated together in their classroom. The research assistant read each word aloud twice, using a normal speech rate. The learners' worksheets were checked for spelling errors and an accuracy score was recorded for each word per participant, with incorrectly spelt words coded as 0 and correct words coded as 1. Cronbach's alpha indicated that the task was reliable: 0.94.

Oral reading fluency (ORF) task

Oral reading fluency was measured using a 1-minute timed task. Learners were required to read a Grade 3-level text (*Kutheni Imvubu zingenazo inwele,* 'How the Hippo lost his fur'), which was 132 orthographic words long, with a mean of 4.7 words per sentence. Any occurrences of errors were captured on Tangerine, from which a words-correct-per-minute-score (wcpm) was calculated automatically by the software, by subtracting the errors learners made from the total number of words read aloud in 1 minute. The data from this task are included in this study, to provide an indication of learners' reading fluency levels.

Data analysis

Multilevel regression analyses were conducted to test for a relationship between orthographic neighbourhood density, orthographic neighbourhood frequency, and the three literacy outcomes: word reading accuracy, lexical decision response time, and spelling accuracy. Specifically, a multilevel logistic regression was conducted for both word

reading accuracy and spelling accuracy, as these are both coded as binary outcomes (correct or incorrect). A general linear multilevel regression (GLMM) was conducted for lexical decision response time, as the outcome is continuous. General linear multilevel regressions are also robust to non-normally distributed data (Lo & Andrews 2015). Multilevel models account for across item and across participant variation, and as such are considered a more valid statistical approach than traditional mean-based ANOVAs (analyses of variance) (Lo & Andrews 2015). Word length, in letters and syllables, and word frequency were included in the models as fixed effects control variables. Likelihood ratio tests were conducted to determine which of the variables in the model significantly predicted each outcome variable. For the multilevel logistic regressions, the odds ratio or exponentiated Beta coefficients (e^{β}) are presented to determine effect size. This is because the Beta estimate cannot be interpreted in the same way as for a multilevel linear model. The odds ratio of a variable indicates the probability of some outcome x relative to some outcome y when all other variables are held constant. Specifically, an odds ratio of 1 is interpreted as a null effect; an odds ratio > 1 indicates a positive or facilitative relationship between the predictor and outcome, whereas an odds ratio < 1 indicates a negative or inhibitory relationship between the predictor and outcome. To make the effect more tangible, one can calculate the change in odds percentage, by using the following formula: $e^{\beta}-1*100$. The change in odds is interpreted similarly to a standardised Beta coefficient and is a calculation of the percentage change in the odds of an outcome; that is associated with a one standard deviation increase in the predictor variable. The results of the models are reported following the standards outlined by Sonderegger (2022).

Ethical considerations

We received ethical clearance through Rhodes University and the Eastern Cape Department of Basic Education. The ethics code for this study is 2020-1195-3307. Informed consent was obtained from the learners' guardians. In addition, informed consent was obtained from the principals and teachers at the schools involved. Lastly, verbal assent was obtained from each participant before administering the task.

Results

Sample performance on literacy measures

Descriptive statistics are presented in Table 2 to provide an overview of the literacy levels of learners in the sample.

The mean reading fluency for learners in our sample was 16.2 wcpm, which is below the national threshold (20 wcpm)

TABLE 2: Descriptive statistics for oral reading fluency, word reading and spelling.

Variable	N	mean	SD	p25	p50	p75
ORF	92	16.2	13.60	4.00	15.5	25.00
Word reading	97	14.2	7.15	13.00	17.0	20.00
Spelling	97	10.9	6.42	6.00	13.0	16.00

ORF, oral reading fluency, SD, standard deviation, p25, 25th percentile

for reading at the end of Grade 2 for Nguni languages (Ardington et al. 2020). Learners had a high mean accuracy for word reading of 14 words correct out of 20, or 70%, with a mean spelling accuracy of 55%.

Word reading accuracy

A multilevel logistic regression analysis was conducted to test for a statistically significant relationship between neighbourhood effects and word reading accuracy. Assumptions of linearity and no multicollinearity were met sufficiently. We fit a crossed-random effects model, with data grouped by participants and items. Firstly, an empty model was run with no predictor variables, thereafter a second model was run to test whether the inclusion of predictor variables improved model fit. Initially the predictor model did not converge. However, after scaling and centring word frequency, the model convergence issue was solved. The statistics for the random and fixed effects of the predictor model are presented in Table 3.

A likelihood ratio test was run to test whether the addition of the predictors to the empty model improved the model fit. This showed no significant improvement in model fit (χ^2 (5) = 3.91, p = 0.56).

Therefore, none of the predictors explained significant variance in the model. This indicates a null effect of orthographic neighbourhood density and neighbourhood frequency for word reading accuracy.

Lexical decision response time

A GLMM was run to test for a statistically significant relationship between neighbourhood effects and lexical decision response time. As was done for word reading, we fit a crossed-random effects model. Firstly, an empty model was run before the predictor model was fit. Table 4 presents the

TABLE 3a: Random effects of the multilevel model for word reading accuracy.

Random effects			
Groups	Variance	SD	ICC
Participants	15.37	3.92	0.80
Items	0.66	0.81	0.03

No. of observations = 1940, groups: participants, 97; items, 20.
SD, standard deviation; ICC, intra-class correlation.

TABLE 3b: Fixed effects of the multilevel model for word reading accuracy.

Fixed effects	b	SE	z	p	CI	
					5%	95%
(Intercept)	3.53	1.67	2.10	0.04	0.24	6.81
Word length (letters)	0.05	0.36	0.14	0.89	-0.65	0.75
Word length (syllables)	-0.26	0.62	-0.41	0.68	-1.48	0.97
Word frequency	0.14	0.30	0.48	0.63	-0.45	0.74
Neighbourhood frequency (OLDF)	-0.01	0.01	-0.64	0.52	-0.02	0.01
Neighbourhood density (OLD20)	-0.38	0.63	-0.61	0.54	-1.63	0.86

No. of observations = 1940, Pseudo R^2(fixed effects) = 0.01, Pseudo R^2(total) = 0.83, AIC = 1245.04, BIC = 1289.61, b = regression coefficient.
SE, standard estimate; CI, confidence interval; OLDF, Orthographic Levenshtein Distance Frequency; OLD20, Orthographic Levenshtein Distance 20.

statistics for the random and fixed effects of the predictor model.

A likelihood ratio test was run to test whether the addition of the predictors to the empty model improved the model fit. This returned a significant result (χ^2 (5) = 12.55, $p < 0.05$), indicating that the predictors contributed significant variance to lexical decision response time. To assess which of the variables in the model were significant predictors, additional likelihood ratio tests were conducted. This showed that of the five variables in the predictor model, only word length in letters (χ^2 (1) = 5.08, $p < 0.05$) and word frequency (χ^2 (1) = 4.63, $p < 0.05$) significantly predicted response time. The Beta coefficients were calculated to check the effect size of these significant predictors. It was found that word length predicted 19% (β = 0.19, 95% confidence interval [CI]: 0.01, 0.36) of lexical decision response time. This means that there is a significant inhibitory effect of word length, with longer words resulting in longer response times. Further, word frequency predicted an additional 8% (β = −0.08, 95% CI: −0.17, 0.00). The negative Beta value indicates that higher frequency words result in shorter response times; thus, word frequency has a facilitatory effect on lexical decision response time. There was no significant relationship between neighbourhood effects and response time.

Spelling accuracy

For spelling accuracy, the same analysis was followed as for word reading accuracy, using a multilevel logistic regression model. Table 5 presents the results of the random and fixed effects for spelling accuracy.

A likelihood-ratios test was run to test whether the addition of the predictors improved the model fit. This returned a significant result (χ^2 (5) = 18.78, $p = 0.002$) which showed the predictors contributed significant variance to spelling accuracy. To assess which of the variables in the model were significant predictors, additional likelihood ratio tests were conducted. These showed that four of the five variables in the

model significantly predicted spelling accuracy: word length in letters (χ^2 (1) = 11.17; $p < 0.001$), word length in syllables (χ^2 (1) = 7.35; $p < 0.05$), word frequency (χ^2 (1) = 6.5; $p < 0.05$), and orthographic neighbourhood frequency (χ^2 (1) = 8.03, $p < 0.005$). The odds ratios of these significant predictor variables were calculated to test for effect size. These values are presented in Table 6.

From Table 6, word length in letters had an odds ratio of $e^\beta = 0.30$ which indicates an inhibitory relationship. To make the effect more tangible, the change in odds percentages is reported here. For word length in letters, this results in a change in odds of 70%. Therefore, an increase in the length of a word (in letters) results in a 70% decrease in the odds of spelling that word correctly. Interestingly, word length in syllables has an odds ratio of $e^\beta = 2.19$, which indicates a facilitatory effect. The change in odds shows that an increase in the length of a word (in syllables) results in a 119% increase in the odds of spelling that word correctly. Thus, word length when measured in letters has an inhibitory effect on spelling in this data set; however, when measured in syllables, it appears to have a facilitatory effect. This seemingly contradictory finding is interrogated in the discussion.

Word frequency has an odds ratio of $e^\beta = 2.27$, which indicates a facilitatory effect. The change in odds for word frequency showed that an increase in the frequency of a word is

TABLE 5a: Random effects of the multilevel model for spelling accuracy.

Random effects			
Groups	Variance	SD	ICC
Participants	7.37	2.72	0.63
Items	1.04	1.02	0.09

No. of observations = 1940, groups: participants, 97; items, 20.
SD, standard deviation; ICC, intra-class correlation.

TABLE 5b: Fixed effects of the multilevel model for spelling accuracy.

Fixed effects	b	SE	z	p	CI	
					5%	95%
(Intercept)	1.04	2.64	0.39	0.69	-4.14	6.22
Word length (letters)	-1.20	0.31	-3.83	0.001	-1.82	-0.59
Word length (syllables)	1.96	0.66	2.98	0.003	0.672	3.25
Word frequency	0.02	0.01	2.77	0.01	0.01	0.03
Neighbourhood frequency (OLDF)	-0.08	0.02	-3.12	0.002	-0.12	-0.03
Neighbourhood density (OLD20)	0.91	0.55	1.65	0.10	-0.17	1.99

No. of observations = 1940; Pseudo R²(fixed effects) = 0.13; Pseudo R²(total)= 0.76; AIC= 1624.72; BIC= 1669.29; b= regression coefficient.
SE, standard estimate; CI, confidence interval; OLDF, Orthographic Levenshtein Distance Frequency; OLD20, Orthographic Levenshtein Distance 20.

TABLE 4a: Random effects of the multilevel model for lexical decision response time.

Random effects			
Groups	Variance	SD	ICC
Participants	24.56	4.96	0.46
Items	1.20	1.10	0.02

No. of observations = 1940, groups: participants, 97; items, 20.
SD, standard deviation; ICC, intra-class correlation.

TABLE 4b: Fixed effects of the multilevel model for lexical decision response time.

Fixed effects	b	SE	t	df	p	CI	
						5%	95%
(Intercept)	3.710	2.620	1.42	15.09	0.18	-1.42	8.85000
Word length (letters)	1.150	0.560	2.06	14.00	0.06	0.06	2.24000
Word length (syllables)	-0.460	0.950	-0.49	14.00	0.63	-2.32	1.40000
Word frequency	-0.004	0.002	-1.95	14.00	0.07	-0.01	0.00001
Neighbourhood frequency (OLDF)	-0.004	0.010	-0.33	14.00	0.75	-0.03	0.02000
Neighbourhood density (OLD20)	-1.100	0.750	-1.46	14.00	0.17	-2.58	0.37000

No. of observations = 1940, Pseudo R²(fixed effects) = 0.02, Pseudo R²(total) = 0.49, AIC = 12277.07, BIC = 12327.20, b = regression coefficient.
SE, standard estimate; CI, confidence interval; OLDF, Orthographic Levenshtein Distance Frequency; OLD20, Orthographic Levenshtein Distance 20.

TABLE 6: Odds ratios for the significant predictors of spelling accuracy.

Parameter	Odds ratio (e$^\beta$)	CI	
		5%	95%
(Intercept)	1.11	0.54	2.28
Word length (letters)	0.30	0.16	0.56
Word length (syllables)	2.19	1.31	3.68
Word frequency	2.27	1.27	4.05
Orthographic neighbourhood frequency	0.42	0.25	0.73

CI, confidence interval

associated with a 127% increase in the odds of spelling that word correctly. Lastly, orthographic neighbourhood frequency had an odds ratio of e$^\beta$ = 0.42, which indicates an inhibitory effect. Further, the change in odds shows that an increase in the neighbourhood frequency of a word results in a 58% decrease in the odds of spelling that word correctly. These results indicate an inhibitory effect of both neighbourhood frequency and word length in letters for spelling accuracy in isiXhosa, and a facilitatory effect of word length in syllables and word frequency for spelling.

Discussion

Our findings indicate a significant inhibitory orthographic neighbourhood frequency effect for spelling, with null neighbourhood effects observed for word reading and response time. That is, words with higher frequency neighbours are harder to spell than those with no such neighbours. Additionally, word length (when measured in letters) had a significant inhibitory effect for both lexical decision response time and spelling, but when measured in syllables, had a significant facilitatory effect for spelling. Finally, word frequency had a significant facilitatory effect for lexical decision response time and spelling.

Orthographic neighbourhood effects in isiXhosa

The finding of an inhibitory effect of neighbourhood frequency for spelling is consistent with studies on orthographic neighbours in reading, which report a competitive effect of higher-frequency neighbours (cf. Slattery 2009; Newman et al. 2017 for English; De Moor & Brysbaert 2000 for Dutch; Grainger & Jacobs 1996: Experiments 1 and 3 for French; and Carreiras et al. 1997 and Acha & Perea 2008 for Spanish). That is, higher frequency neighbours compete for lexical access when reading. This finding also suggests an extension of this competitive effect to spelling. Further, this finding is consistent with research on phonographic neighbours in spelling (Lété et al. 2008; Maggio et al. 2012). Other research on orthographic neighbours in spelling using linguistic error analysis reports a prevalence of spelling errors which can be attributed to competition from neighbours (Andrews & Hersch 2010; Burt & Blackwell 2008; Folk et al. 2002) much like the results of the present study. Drawing on the dual-route framework, it could be argued that learners in this study draw on a partially lexical route when spelling words in isiXhosa – in which neighbours are activated. When spelling a word, its frequent neighbours are simultaneously activated, which results in competition within the orthographic lexicon and

subsequently the letter output (written word). That is, spelling errors occur due to the presence of more frequent neighbours which outcompete the spelling of the target word.

The lack of a neighbourhood effect for reading in isiXhosa is more puzzling. This finding is inconsistent with the literature on neighbourhood effects, which has consistently reported some effect of orthographic neighbours on reading, be it facilitatory or inhibitory (Aguasvivas et al. 2020; Brysbaert et al. 2015; Ferrand et al. 2018; Lim 2016 and Parker et al. 2021).

The disparity between neighbourhood effects in spelling and word reading for isiXhosa suggests that different linguistic knowledge may be drawn on when spelling, in comparison to that used when reading. While we did not look into this, we suggest that future researchers investigate in some detail the cognitive components that underlie these two key literacy skills. However, one possibility that could explain this disparity is that the learners in this study draw on a strictly sublexical route when reading, while adopting a partially lexical route when spelling. That is, when presented with a visual word (reading), isiXhosa Grade 3 learners rely on a sublexical route, due to the transparency of the language, as was found in Probert and De Vos (2016). Learners need not even access the corresponding word from their lexicon – which has been found to be the case for many South African learners, who are said to 'bark at print' (Pretorius & Spaull 2016). The fluency levels of our sample (16.2 wcpm) were below the reading fluency benchmark for the end of Grade 2 (20 wcpm) (Ardington et al. 2020). However, the learners demonstrate high accuracy levels when reading words in isolation (71%). This suggests that learners are able to decode accurately, particularly when reading words in isolation in an untimed task, but stumble when reading connected text in a timed task. This again suggests that the learners in our sample may be relying on the sublexical processing route when reading, resulting in the trade-off between accuracy and speed. Furthermore, this could explain why neighbourhood effects were not found for reading in our study, since neighbours are activated at the lexical level, and thus require some reliance on the lexical processing route. Further studies should investigate whether there is indeed a link between reading fluency levels and orthographic neighbourhood effects to confirm this. In comparison to reading, a complete reliance on the sublexical route may not be feasible for novice spellers. This is because there is no visual component which learners can rely on (i.e. letters on the page) when spelling. Therefore, learners may need to draw more on lexical information in order to access the spelling of word forms, before being able to produce a written word. This partial reliance on the lexical processing route could explain the presence of neighbourhood effects for spelling in our study. What is evident though is that orthographic neighbours, specifically higher-frequency neighbours, are activated in the lexicon when spelling in isiXhosa, but not in reading.

Another possible interpretation of our finding of inhibitory neighbourhood effects for spelling, but null effects for reading, could be attributed to the presence of a masked phonological neighbourhood effect. IsiXhosa has a transparent orthography, which means that there is likely a substantial overlap between orthographic neighbours and phonological neighbours, with neighbours differing by a similar number of phonemes, as letters from a target word. Thus, an orthographic neighbourhood effect may in fact be masking a phonological neighbourhood effect. This potentially confounding issue has been brought up by other researchers, who note that 'in reality the effect of orthography is not limited to visual word processing but also extends to auditory processing' (Marian 2017:9), and vice versa. Thus, the study of orthographic neighbours will always overlap to some degree with phonological neighbours, especially when the language in question has a transparent orthography. Further, because the spelling task includes an auditory component, in which the participant must first listen to the word spoken aloud, before writing the word down, phonological neighbours may be more active during an auditory spelling task than they are for a visual word reading task. Thus again, the observed orthographic neighbourhood frequency effect for spelling may be indicative of a masked phonological neighbourhood effect, which could potentially explain why no such effect was observed for word reading. Future research should attempt to control for phonological neighbours.

The finding of a null orthographic neighbourhood density effect for both literacy skills aligns with the findings of Kapnoula et al. (2017) in their study on Greek. Kapnoula et al. (2017) attribute this to the long words in Greek, which mitigate neighbourhood density effects, since longer words usually have more distant neighbours, and thus sparser neighbourhoods. We suggest that our finding for isiXhosa can be explained in much the same way. To corroborate this, we calculated the average density of a word in the Rees and Randera (2017) corpus. We found that average words in isiXhosa have a neighbourhood density of 3.04 (standard deviation [SD] = 1.20). That is, words differ on average by three letters from other words in the language. This means that neighbourhoods in isiXhosa are mostly very sparse, which could explain why density does not play a role in reading and writing in isiXhosa. With reference to the dual-route model, this finding suggests that there is little top-down strengthening from neighbours at the lexical level to graphemes at the sublexical level, because the neighbours do not share as many letters with the target word.

Word length and word frequency effects in isiXhosa

We included two metrics for word length in our study: the number of syllables, and the number of letters. Our findings indicate that word length in letters had a significant inhibitory effect on spelling accuracy and lexical decision response time, whereas word length in syllables had a significant facilitatory effect for spelling accuracy. Therefore, words with more letters were spelt less accurately and resulted in longer response times, but words with more syllables were spelt more accurately. If one holds the number of letters constant, words with more syllables will naturally have fewer letters per syllable. For example the words *elininzi* (a lot) and *umngxuma* (hole) are both eight letters in length; however, *e.li.ni.nzi* has four syllables, whereas *u.mngxu.ma* has three syllables. The average length of the syllables in *elininzi* is two letters, whereas the average length of the syllables in *umngxuma* is 2.6 letters. Research has shown that complex consonant graphemes such as those in the onset of the second syllable of *umngxuma*, often result in spelling errors (Daries & Probert 2020). Thus, when keeping letter length constant, a word such as *elininzi* will be easier to spell than *umngxuma*, even though it has more syllables than the former. In this way, an increase in the number of syllables in a word can actually facilitate the spelling of that word, as reported in our findings.

The absence of a syllable word length effect in the lexical decision task, is likely owed to the type of linguistic processing that is required by the task. That is, the lexical decision task is a purely visual task and does not involve any auditory processing. Since syllables are a phonological unit, the number of syllables in a word does not appear to affect the processing time for words in the lexical decision task. The absence of any word length effect for word reading accuracy is less clear; however, it likely suggests that word length does not impact the accuracy of decoding, particularly if learners are relying on grapheme-to-phoneme mappings when reading. It is also arguable that the task stimuli (3 and 4 syllable long words) were too short to experience any negative influence on accuracy from word length. Whereas the speed of reading is negatively impacted by word length, as evident by the inhibitory word length effect for lexical decision response time.

We report a significant facilitatory effect of word frequency in the lexical decision and spelling tasks, but not in the word reading accuracy task. Words with higher frequency scores benefit from decreased response times and increased spelling accuracy. Since word frequency operates at the lexical level, the absence of a word frequency effect for reading accuracy confirms our hypothesis that the learners in our sample may in fact be 'barking at print', and relying solely on grapheme-to-phoneme mappings when reading. As such, learners cannot yet benefit from frequency effects for reading.

Conclusion

This article investigated the influence of orthographic neighbours for reading and spelling in Grade 3 isiXhosa home-language learners.

No effect of orthographic neighbourhood density, or neighbourhood frequency, was found for either word reading task, but a significant inhibitory effect of neighbourhood frequency on spelling accuracy was found. The lack of neighbourhood effects in word reading accuracy and lexical decision response time is likely owed to learners' reliance on

sublexical processing when reading – which can be attributed to the orthographic transparency of the language and visual aid when decoding. With spelling, however, there appears to be a greater reliance on lexical processing – resulting in competitive effects at the lexical level, and hence greater spelling errors. Since orthographic neighbours only interact at the lexical level of orthographic processing, it follows that these effects are not present for word reading. It is our recommendation that future research investigates the interaction between literacy levels and orthographic neighbourhood effects.

The relevance of these findings in the context of the South African literacy landscape, is that they contribute to a more thorough understanding of reading and writing in a setting where learners are continuously underperforming in reading literacy (DBE 2023). Further, these findings provide empirical evidence which can inform linguistic theory, and importantly, may also contribute to the development of targeted pedagogical practices, which can address learners' spelling errors. It is our recommendation that spelling reforms implement lists of orthographic neighbours when teaching novice words, such that the nuances between word spellings are made more explicit. This will of course require a more extensive database of orthographic neighbours in isiXhosa, to be made available to educational practitioners, such that linguistically sound teaching resources may be developed.

Acknowledgements

Competing interests

Authors' contributions

Both P.S.C. and T.N.B. were involved in the conceptualisation of the study, as well as in the design of the methodology.

Data were analysed by P.S.C. with supervision from T.N.B. The original manuscript was written by P.S.C. and both authors contributed to the revision and editing of the final manuscript.

Funding information

Data availability

The data that support the findings of this study are available on reasonable request from the corresponding author, P.S.C. The data are not publicly available due to their containing information that could compromise the privacy of research participants.

Disclaimer

References

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

Aguasvivas, J., Carreiras, M., Brysbaert, M., Mandera, P., Keuleers, E. & Duñabeitia, J.A., 2020, 'How do Spanish speakers read words? Insights from a crowdsourced lexical decision megastudy', *Behavior Research Methods* 52(5), 1867–1882. https://doi.org/10.3758/s13428-020-01357-9

Andrews, S., 1997, 'The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts', *Psychonomic Bulletin and Review* 4(4), 439–461. https://doi.org/10.3758/BF03214334

Andrews, S. & Hersch, J., 2010, 'Lexical precision in skilled readers: Individual differences in masked neighbor priming', *Journal of Experimental Psychology: General* 139(2), 299–318. https://doi.org/10.1037/A0018366

Ardington, C., Wills, G., Pretorius, E., Deghaye, N., Mohohlwane, N., Menendez, A., Mtsatse, N. & Van der Berg, S., 2020, Summary report: Benchmarking early grade reading skills in Nguni languages: ReSEP, Stellenbosch University. Cape Town: SALDRU, University of Cape Town. Chicago: NORC at the University of Chicago. Pretoria: Department of Basic Education.

Balota, D.A. & Chumbley, J.I., 1984, 'Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage', *Journal of Experimental Psychology: Human Perception and Performance* 10(3), 340–357. https://doi.org/10.1037/0096-1523.10.3.340

Balota, D.A., Cortese, M.J., Sergent-Marshall, S.D., Spieler, D.H. & Yap, M.J., 2004, 'Visual word recognition of single-syllable words', *Journal of Experimental Psychology* 133(2), 283–316. https://doi.org/10.1037/0096-3445.133.2.283

Balota, D.A., Yap, M.J., Cortese, M.J., Hutchison, K.A. & Kessler, B.L., 2007, 'The English Lexicon Project', *Behavior Research Methods* 39(3), 445–459. https://doi.org/10.3758/BF03193014

Barnhart, A.S. & Goldinger, S.D., 2015, 'Orthographic and phonological neighborhood effects in handwritten word perception', *Psychonomic Bulletin and Review* 22(6), 1739–1745. https://doi.org/10.3758/s13423-015-0846-z

Beinborn, L., Zesch, T. & Gurevych, I., 2016, 'Predicting the spelling difficulty of words for language learners', in *Proceedings of the 11th workshop on innovative use of NLP for building educational applications*, pp. 73–83, BEA@NAACL-HLT Association for Computational Linguistics. San Diego. https://doi.org/10.18653/v1/W16-0508

Berghoff, R., 2023, 'Deriving lexical statistics for psycholinguistic research on isiXhosa', *Journal of the Digital Humanities Association of Southern Africa* 4(1), 1–7. https://doi.org/10.55492/dhasa.v4i01.4442

Berghoff, R., 2024, 'Lexical and sub-lexical frequencies in isiXhosa-medium children's stories', *Southern African Linguistics and Applied Language Studies*, 1–18. https://doi.org/10.2989/16073614.2023.2252010

Biesman-Simons, C., Dixon, K., Pretorius, B. & Reed, Y., 2020, 'Pitfalls and possibilities in literacy research: A review of South African literacy studies, 2004–2018', *Reading and Writing (South Africa)* 11(1), 1–9. https://doi.org/10.4102/rw.v11i1.238

Boot, I. & Pecher, D., 2008, 'Word recognition is affected by the meaning of orthographic neighbours: Evidence from semantic decision tasks', *Language and Cognitive Processes* 23(3), 375–393. https://doi.org/10.1080/01690960701527317

Brysbaert, M., Stevens, M., Mandera, P. & Keuleers, E., 2015, 'The impact of word prevalence on lexical decision times: Evidence from the Dutch lexicon project 2', *Journal of Experimental Psychology: Human Perception and Performance* 42(3), 441–458.

Burt, J.S. & Blackwell, P., 2008, 'Sound-spelling consistency in adults' orthographic learning', *Journal of Research in Reading* 31(1), 77–96. https://doi.org/10.1111/J.1467-9817.2007.00362.X

Carreiras, M., Perea, M. & Grainger, J., 1997, 'Effects of orthographic neighborhood in visual word recognition: Cross-task comparisons', *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23(4), 857–871. https://doi.org/10.1037//0278-7393.23.4.857

Coltheart, M., 1978, 'Lexical access in simple reading tasks', in G. Underwood (ed.), *Strategies of information processing,* pp. 151–216, Academic Press, San Diego, CA.

Coltheart, M., Davelaar, E., Jonasson, J.T. & Besner, D., 1977, 'Access to the internal lexicon', in S. Dornic (ed.), *Attention and performance VI*, pp. 535–555, Lawrence Erlbaum Associates, Hillsdale, NJ.

Coltheart, M., Rastle, K., Perry, C., Langdon, R. & Ziegler, J., 2001, 'DRC: A Dual Route Cascaded model of visual word recognition and reading aloud', *Psychological Review* 108(1), 204–256. https://doi.org/10.1037/0033-295X.108.1.204

Conrad, N.J., 2008, 'From reading to spelling and spelling to reading: Transfer goes both ways', *Journal of Educational Psychology* 100(4), 869–878. https://doi.org/10.1037/a0012544

Daries, M.A. & Probert, T.N., 2020, 'A linguistic analysis of spelling errors in grade 3 isiXhosa home-language learners', *Reading and Writing (South Africa)* 11(1), 1–10. https://doi.org/10.4102/rw.v11i1.265

De Moor, W. & Brysbaert, M., 2000, 'Neighborhood-frequency effects when primes and targets are of different lengths', *Psychological Research* 63(2), 159–162.

Department of Basic Education, 2023, *PIRLS 2021: South African preliminary highlights report*, Department of Basic Education, Pretoria.

Dujardin, E. & Mathey, S., 2022, 'The neighbourhood frequency effect in naming is influenced by substituted-letter confusability and lexical skills', *Journal of Cognitive Psychology* 34(8), 947–961. https://doi.org/10.1080/20445911.2022.2099873

Erten, B., Bozsahin, C. & Zeyrek, D., 2014, 'Turkish resources for visual word recognition', in *Proceedings of the 9th international conference on language resources and evaluation*, pp. 2106–2110, European Language Resources Association (ELRA), Reykjavik.

Ferrand, L., Méot, A., Spinelli, E., New, B., Pallier, C., Bonin, P. et al., 2018, 'MEGALEX: A megastudy of visual and auditory word recognition', *Behavior Research Methods* 50(3), 1285–1307. https://doi.org/10.3758/s13428-017-0943-1

Folk, J.R., Rapp, B. & Goldrick, M., 2002, 'The interaction of lexical and sublexical information in spelling: What's the point?', *Cognitive Neuropsychology* 19(7), 653–671. https://doi.org/10.1080/02643290244000184

Grainger, J. & Jacobs, A.M., 1996, 'Orthographic processing in visual word recognition: A multiple read-out model', *Psychological Review* 103(3), 518–565. https://doi.org/10.1037/0033-295X.103.3.518

Houghton, G. & Zorzi, M., 2003, 'Normal and impaired spelling in a connectionist dual-route architecture', *Cognitive Neuropsychology* 20(2), 115–162.

Huntsman, L.A. & Lima, S.D., 2002, 'Orthographic neighbors and visual word recognition', *Journal of Psycholinguistic Research* 31(3), 289–306. https://doi.org/10.1023/A:1015544213366

Kapnoula, E.C., Protopapas, A., Saunders, S.J. & Coltheart, M., 2017, 'Lexical and sublexical effects on visual word recognition in Greek: Comparing human behavior to the Dual Route Cascaded Model', *Language, Cognition and Neuroscience* 32(10), 1290–1304.

Khanna, M.M., Cortese, M.J., Hughes, K. & Anderson, K., 2023, 'Spelling monosyllabic English words', *Quarterly Journal of Experimental Psychology* 77(4), 729–746. https://doi.org/10.1177/17470218231178507

Kuperman, V., Schroeder, S. & Gnetov, D., 2024, 'Word length and frequency effects on text reading are highly similar in 12 alphabetic languages', *Journal of Memory and Language* 135, 1–15. https://doi.org/10.1016/j.jml.2023.104497

Lété, B., Peereman, R. & Fayol, M., 2008, 'Consistency and word-frequency effects on spelling among first- to fifth-grade French children: A regression-based study', *Journal of Memory and Language* 58(4), 952–977.

Lim, S.W.H., 2016, 'The influence of orthographic neighborhood density and word frequency on visual word recognition: Insights from RT distributional analyses', *Frontiers in Psychology* 7(March), 1–7. https://doi.org/10.3389/fpsyg.2016.00401

Lo, S. & Andrews, S., 2015, 'To transform or not to transform: Using generalized linear mixed models to analyse reaction time data', *Frontiers in Psychology* 6(1171) 1–16. https://doi.org/10.3389/Fpsyg.2015.01171

Maggio, S., Lété, B., Chenu, F., Jisa, H. & Fayol, M., 2012, 'Tracking the mind during writing: immediacy, delayed, and anticipatory effects on pauses and writing rate', *Reading and Writing* 25(9) 2131–2151. https://doi.org/10.1007/S11145-011-9348-1

Marian, V., 2017, 'Orthographic and phonological neighborhood databases across multiple languages', *Written Language & Literacy* 20(1), 6–26.

Meade, G., Grainger, J. & Declerck, M., 2021, 'Friend or foe? Flankers reverse the direction of orthographic neighbourhood effects', *Language, Cognition and Neuroscience* 36(5), 535–542.

Morton, J. & Patterson, K.E., 1980, 'A new attempt at an interpretation, or an attempt at a new interpretation', in M. Coltheart, K.E. Patterson & J.C. Marshall (eds.), *Deep dyslexia*, pp. 91–118, Routledge & Kegan Paul, London.

Newman, R.S., German, D.J. & Jagielko, J.R., 2017, 'Influence of lexical factors on word-finding accuracy, error patterns, and substitution types', *Communication Disorders Quarterly* 39(2), 356–366. https://doi.org/10.1177/1525740117712205

Oflazer, K., Göçmen, E. & Bozsahin, C., 1994, *An outline of Turkish morphology*, 20th year edn., Report to NATO Science Division SfS III (TU-LANGUAGE), Brussels.

Parker, A.J., Egan, C., Grant, J.H., Harte, S., Hudson, B.T. & Woodhead, Z.V., 2021, 'The role of orthographic neighbourhood effects in lateralized lexical decision: A replication study and meta-analysis', *PeerJ* 9, 1–25. https://doi.org/10.7717/peerj.11266

Pierce, J.W., Gray, J.R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H. et al., 2019, 'Psychopy2: Experiments in behavior made easy', *Behavior Research Methods* 51, 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pretorius, E., 2018, *Reading in the African languages: An annotated bibliography 2004–2007*, viewed 23 October 2023, from https://www.jet.org.za/clearinghouse/projects/primted/resources/language-and-literacy-resources-repository/reading-in-african-languages-an-annotated-bibliography-pretorius-2018.pdf/view

Pretorius, E.J. & Spaull, N., 2016, 'Exploring relationships between oral reading fluency and reading comprehension amongst English second language readers in South Africa', *Reading and Writing: An Interdisciplinary Journal* 29(7), 1449–1471.

Probert, T. & De Vos, M., 2016, 'Word recognition strategies amongst isiXhosa/English bilingual learners: The interaction of orthography and language of learning and teaching', *Reading & Writing* 7(1), 1–10.

Rapcsak, S.Z., Henry, M.L., Teague, S.L., Carnahan, S.D. & Beeson, P.M., 2007, 'Do dual-route models accurately predict reading and spelling performance in individuals with acquired alexia and agraphia?', *Neuropsychologia* 45(11), 2519–2524. https://doi.org/10.1016/j.neuropsychologia.2007.03.019

Rapp, B., Epstein, C. & Tainturier, M.J., 2002, 'The integration of information across lexical and sublexical processes in spelling', *Cognitive Neuropsychology* 19(1), 1–29.

Rees, S. & Randera, A., 2017, *IsiXhosa foundation phase corpus*, Molteno.

Roux, S. & Bonin, P., 2009, 'Neighborhood effects in spelling in Adults', *Psychonomic Bulletin and Review* 16(2), 369–373.

Scaltritti, M., Arfé, B., Torrance, M. & Peressotti, F., 2016, 'Typing pictures: Linguistic processing cascades into finger movements', *Cognition* 156, 16–29. https://doi.org/10.1016/j.cognition.2016.07.006

Schaars, M.M.H., Segers, E. & Verhoeven, L., 2017, 'Predicting the integrated development of word reading and spelling in the early primary grades', *Learning and Individual Differences* 9(September), 127–140. https://doi.org/10.1016/j.lindif.2017.09.006

Schaefer, M., Probert, T. & Rees, S., 2020, 'The roles of phonological awareness, rapid automatised naming and morphological awareness in isiXhosa', *Per Linguam* 36(1), 90–111.

Sears, C.R., Hino, Y. & Lupker, S.J., 1999, 'Orthographic neighbourhood effects in parallel distributed processing models', *Canadian Journal of Experimental Psychology* 53(3), 220–229. https://doi.org/10.1037/h0087311

Siakaluk, P.D., Sears, C.R. & Lupker, S.J., 2002, 'Orthographic neighborhood effects in lexical decision: The effects of nonword orthographic neighborhood size', *Journal of Experimental Psychology: Human Perception and Performance* 28(3), 661–681.

Slattery, T.J., 2009, 'Word misperception, the neighbor frequency effect, and the role of sentence context: evidence from eye movements', *Journal of Experimental Psychology: Human Perception and Performance* 35(6) 1969–1975. https://doi.org/10.1037/A0016894

Sonderegger, M., 2022, *Regression modeling for linguistic data* (Preprint), viewed 01 February 2023, from osf.io/pnumg

Wilsenach, C., 2013, 'Phonological skills as predictor of reading success: An investigation of emergent bilingual Northern Sotho/English learners', *Per Linguam* 29(2), 17–32. https://doi.org/10.5785/29-2-554

Wilsenach, C., 2019, 'Phonological awareness and reading in Northern Sotho: Understanding the contribution of phonemes and syllables in grade 3 reading attainment', *South African Journal of Childhood Education* 9(1), 1–10. https://doi.org/10.4102/sajce.v9i1.647

Yap, M.J., Liow, S.J.R., Jalil, S.B. & Faizal, S.S.B., 2010, 'The Malay lexicon project: A database of lexical statistics for 9,592 words', *Behavior Research Methods* 42(4), 992–1003. https://doi.org/10.3758/BRM.42.4.992

Yarkoni, T., Balota, D. & Yap, M., 2008, 'Moving beyond Coltheart's N: A new measure of orthographic similarity', *Psychonomic Bulletin and Review* 15(5), 971–979. https://doi.org/10.3758/PBR.15.5.971

Ziegler, J.C., Perry, C. & Coltheart, M., 2003, 'Speed of lexical and nonlexical processing in French: The case of the regularity effect', *Psychonomic Bulletin and Review* 10(4), 947–953. https://doi.org/10.3758/BF03196556

Zulu, P.N., Botha, G. & Barnard, E., 2008, 'Orthographic measures of language distances between the official South African languages', *Literator* 29(1) 185–204. https://doi.org/10.4102/lit.v29i1.106