Individual Differences in Spelling Ability Influence Phonological Processing during Visual Word Recognition

Mark Yates

University of South Alabama

Timothy J Slattery

Bournemouth University

Author Note

Address correspondence to Mark Yates, Psychology Department, 75 S. University Blvd., Mobile, AL 36688 Phone: (251) 460-7872 Email: myates@southalabama.edu

Abstract

In the research reported here, we investigated how phonological processing in the lexical decision task is influenced by individual differences in the reading and spelling abilities of participants. We used phonological neighborhood spread as a measure of phonological processing. Spread refers to the number of phoneme positions in a word that can be changed to form a phonological neighbor. Replicating previous research, we found that words forming neighbors across three positions (P3) were recognized more rapidly than those forming neighbors across only two positions (P2). Importantly, we found that this spread effect interacted with spelling ability. The difference between P3 and P2 was largest when spelling recognition was high and spelling production low. These opposing effects of spelling ability are explained in terms of a language system that consists of separate orthographic systems for reading and spelling. Although these two orthographic systems are separate, they share information through a shared response buffer (Jones & Rawson, 2016). Within this framework, it is argued that lexical decisions are made once the information in the response buffer reaches threshold and that time to reach this threshold is influenced by two sources. One is the quality of the orthographic connections in the reading system and is measured by spelling recognition. The other is the quality of the orthographic connections in the spelling system and is measured by spelling production.

Keywords: Visual word recognition; Phonological processing; Spelling ability

1. Introduction

There is a rich history of word recognition research in cognitive psychology. A wealth of studies have examined how orthography, phonology, and semantics interact to influence the word recognition process. To explain this research different models of word recognition have been proposed (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Harm & Seidenberg, 2004; Perry, Ziegler, & Zorzi, 2007; Seidenberg & McClelland, 1989). A hidden assumption that permeates the research and modeling efforts on word recognition is that most skilled readers have similar word recognition systems. For example, most models assume that the representations for a word (e.g., orthographic and phonological) are the same between readers. Consider two of the most successful computational models, the dual-route cascaded (DRC) model (Coltheart et al., 2001) and connectionist dual process (CDP+) model (Perry et al., 2007). Both models employ interactive activation, and in both, the orthographic and phonological lexical entries are hard wired such that they fully specify their letters and phonemes. Thus, these models have been built assuming that both the orthographic and phonological representations are precise, and as such, what these models really simulate is the word recognition performance of readers with precise and fully specified lexical representations. Part of the reason the models have been designed this way is because they were intended to simulate item level data that has been averaged across subjects. This averaging across subjects is the norm within word recognition research and has been useful in elucidating the nature of word recognition within skilled readers (Andrews & Lo, 2012). Nevertheless, averaging across subjects obfuscates the potential role of individual differences in word recognition.

1.1 Lexical Quality Hypothesis

Perfetti's lexical quality hypothesis (Perfetti & Hart, 2002) predicts that there should be individual differences in the quality of lexical representations and that these differences should map onto differences in how the word recognition process unfolds. According to the lexical quality hypothesis, every word consists of orthographic, phonological, and semantic constituents, and the quality of these can vary. In a precisely specified lexical representation, these constituents are fully specified and tightly bound to one another such that the visual input will activate the appropriate representation quickly and with minimal competition from other representations. When relating lexical quality to individual differences one must consider the overall quality of the lexical representations an individual has. Readers whose representations tend to be underspecified and who lack many high-quality representations would be considered less skilled relative to readers with many high-quality representations that are fully specified. Recent research by Andrews and colleagues (Andrews & Hersch, 2010; Andrews & Lo, 2012, 2013) used the predictions of the lexical quality hypothesis to frame the results of a series of experiments highlighting the impact of individual differences on visual word recognition. In support of the lexical quality hypothesis, this research has clearly demonstrated that individual differences between skilled readers interacts with orthographic masked priming. As an example, using four letter primes and targets, Andrews and Hersch (2010) found that better spellers showed inhibitory priming, but only for targets with large orthographic neighborhoods. Using five letter primes and targets also showed larger inhibitory priming effects for better spellers, but orthographic neighborhood did not interact with spelling. Subsequent research has shown individual differences also interact with morphological priming. Andrews and Lo (2013) found that participants with higher vocabulary than spelling showed strong priming effects for transparent prime-target pairs (e.g., *worker-WORK*), but diminished priming for opaque pairs (e.g., *corner-CORN*). Participants that were better at spelling than vocabulary showed as strong a priming effect for opaque pairs as transparent pairs. Taken as a whole, Andrews and colleagues' research has highlighted the importance of understanding how individual differences in the lexical quality of skilled reader's lexical representations influences word recognition.

1.2 Phonological processing in visual word recognition

One of the most debated issues in visual word recognition is the degree to which phonology influences recognition. The idea that phonological effects depend on reading ability has been around for over 30 years (Waters & Seidenberg, 1985), and there is considerable interest in phonological processing during visual word recognition. Despite these facts, there has been little research looking at how individual differences within readers influences phonological effects. According to the time-course model of Waters and Seidenberg, phonological activation lags behind orthographic activation. If recognition can be made based on orthographic information before phonological activation, then phonology should not play a role in recognition. On the other hand, when recognition based on orthographic activation does not occur before phonological activation, then one would expect phonological effects to emerge. They state that this should happen for lower frequency words and for poorer readers. The reason in both cases is that orthographic activation is not fast enough to lead to recognition before phonological activation comes online. To support this prediction, they reported that slower subjects, who are putatively poorer readers, showed a regularity effect (i.e., slower responses to irregular words than to regular words) in lexical decision, but that faster subjects did not. Later research by Unsworth and Pexman (2003) using a lexical decision task evaluated how individual differences influence three phonological effects: the homophone effect (i.e., longer latencies to homophones than to nonhomophones), the homograph effect, (i.e., longer latencies to homographs than to nonhomographs), and regularity effects. To determine reading skill, they used a measure of print exposure, the Author Recognition Test (Stanovich & West, 1989), and a measure of vocabulary and reading comprehension, the Nelson Denny Reading Test. Their results revealed that both less and more skilled readers failed to show a homograph effect, suggesting that less skilled readers did not rely more on phonological processing than did more skilled readers. The homophone effect was found in the latency data of both less and more skilled readers, indicating that phonology was activated to a certain extent for both groups of readers. Less skilled readers made more errors to homophones than to controls compared with more skilled readers. The authors attribute this to less efficient phonological-orthographic connections and/or less efficient orthographic-phonological connections. Congruent with the claim that less skilled readers have less reliable orthographic-phonological mappings, they found that only the less skilled readers produced a regularity effect. Taken together these results indicate that less and more skilled readers differ in the quality of their mappings between orthography and phonology. More recent research by Burt and Jared (2015) revealed a difference in the size of the homophone effect for more versus less skilled readers. It is not clear why Burt and Jared found a difference in the latency data whereas Unsworth and Pexman did not. The two studies differ in terms of stimuli and how reading skill is defined, making direct comparisons difficult.

Given the interest in phonological processing and the growing body of work demonstrating the importance of understanding how individual differences impact visual word recognition, the purpose of the current research was to apply this individual difference approach to phonological processing during visual word recognition. Specifically, we investigated whether individual differences in orthographic quality (measured by spelling production and spelling recognition), semantic quality (measured by vocabulary), and reading ability (measured by reading comprehension and reading speed) interact with the phonological neighborhood spread effect in the lexical decision task. Phonological neighborhood spread is a measure of how a word's phonological neighbors are distributed across phoneme positions and is denoted with the letter *P*. The *P* value of a word is the number of phonemes that can be changed to form a neighbor. For example, for the word *chill* that has a phonological CVC structure, there is no phonological neighbor that can be formed by changing the /I/ phoneme. Thus, there are only two phoneme positions within the word *chill* that can be changed to yield a neighbor making it a *P* = 2 word. An example of a *P* = 3 word is *dish* where neighbors can be formed across all three positions (e.g., *fish, dash,* and *dill*). Yates (2009a) compared responses to words that were either *P* = 2 or *P* = 3 and found that *P* = 3 words were responded to more rapidly than were *P* = 2 words in a lexical decision task. The effect was explained within the context of a distributed framework like that used in parallel distributed processing (PDP) models (e.g., Harm & Seidenberg, 2004; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989) where a set of orthographic units is connected to a set of phonological units through a set of hidden units. Within the distributed framework, as phonological neighbors of a target word share many of the same units and connections, each time a word's neighbor is encountered the connections for the target word would be strengthened. The end result is faster recognition for words with many phonological neighbors (for a similar account of orthographic neighborhood effects see Andrews, 1992; Sears, Hino, & Lupker, 1999). With regards to spread, the important consideration is overlap in terms of phoneme pairs. As an example, for a CVC word every time the model encounters a CV_ neighbor the connections coding the initial consonant and the vowel will be strengthened in the target word. For *P* = 3 words, all three phoneme pairings (i.e., CV_, C_C, and _VC) are available in its neighborhood allowing the target word to receive additional strengthening across all possible phoneme pairs. On the other hand, for *P* = 2 words, one of the pairings is not found in the neighborhood, and only occurs in the target word itself. This means the model will only get exposure to that phoneme pairing in the target word, leading to weaker phonological connections. The net result is that *P* = 2 words will have weaker phonological connections relative to *P* = 3 words, and phonological processing will be delayed for *P* = 2 words. As phonological activation feeds back to the orthographic units to facilitate the lexical decision, *P* = 2 words will be responded to more slowly than *P* = 3 words that benefit from increased phonological feedback. Yates (2009a) referred to this as the *impoverished phonological connections account*. This account is similar to the accounts given for the facilitative effect of orthographic neighborhood density during word recognition (Andrews, 1992; Sears, Hino, & Lupker, 1995). It is also similar to our account of the effect of phonological neighborhood density on visual word recognition given in our previous work (Yates, 2005; Yates, Locker, & Simpson, 2004). Our reason for choosing phonological neighborhood spread in the current work is that the stimuli are controlled across a much larger number of potential confounding variables than word lists we or others have used in studies looking at phonological neighborhood density. It is also important to state that the words used in the current study were controlled on many other neighborhood characteristics (e.g., phonological and orthographic neighborhood density). For a complete list, please see Table 1.

1.3 Do spelling and reading share the same orthographic representations?

Before turning to the issue of whether the same orthographic representations are used for spelling and reading, it is worth noting that similar debates exist in other areas. For example, a similar debate in the literature is whether there are separate or shared input and output phonological systems. Some have argued for separate phonological input and output systems (Howard & Nickels, 2005; Martin, Lesch, & Bartha, 1999), whereas others have argued for the same representations being used for input and output phonology (Allport & Funnell, 1981). It is also worth noting that the debate between separate or shared systems extends beyond language. For example, in terms of object versus face processing, researchers have shown that brain regions such as the fusiform face area are responsive to faces but not objects, whereas other regions are responsive to objects but not faces (Grill-Spector, Knouf, & Kanwisher, 2004). This seems to indicate that faces and objects are processed separately. However, others have shown that the fusiform face area is not only involved in face processing but is also used in object processing (Haxby et al., 2001).

For the current study, we were interested in how participants' orthographic quality interacted with phonological processing in the lexical decision task. We did not have any specific predictions with regards to semantic quality or reading ability but felt it was prudent to include these to rule out alternative interpretations of the impact of orthographic quality on the phonological neighborhood spread effect. With regards to orthographic quality, there are three possible predictions that can be made, but to understand the competing predictions, it is first necessary to discuss the debate in the literature regarding whether spelling and reading share a common orthographic lexicon. By one account, spelling and reading share the same lexicons (e.g., Allport & Funnell, 1981; Holmes & Carruthers, 1998). For instance, when reading, the orthographic lexicon would serve as an orthographic input lexicon that would be used in recognizing and reading words. When spelling, the orthographic lexicon would serve as an output lexicon used to produce the spelling of a word. Holmes and Carruthers (1998) have argued for a single orthographic lexicon based on their finding that when participants were unable to correctly spell a word they could not correctly distinguish between the correct spelling of a word and their own misspellings in a spelling recognition task. If there were separate orthographic lexica, one would expect participants to be able to correctly identify words they misspelled as the recognition task would be dependent upon the reading orthographic representations. Similarly, Burt and Tate (2002) showed that participants performance in the lexical decision task (a reading task) were dependent on their spelling accuracy in a spelling production task. Others have argued that these item-specific relationships between reading and spelling can be explained by models with separate orthographic lexicons if one assumes that each lexicon has similar properties (Jones & Rawson, 2016). For example, if a word has a weak representation in both orthographic lexicons then performance on the spelling and reading tasks would be similar. We will call the view that there is one orthographic lexicon used for both reading and spelling the *completely shared model*. The opposing view is that the reading and spelling orthographic lexicons are separate (e.g., Weekes & Coltheart, 1996). We will call this the *completely separate model*. Most of the debate in the literature has been framed around whether spelling and reading share the same lexicons (e.g., orthographic lexicon), but the same arguments would apply to the distributed approach. From the distributed perspective, the question would be whether spelling and reading share the same orthographic units.

In recent work, an interesting theoretical middle ground has been proposed and empirically verified. Specifically, Jones and Rawson (2016) argue that there are separate cognitive mechanisms that subserve reading and spelling, but that they share activation. According to their account there are separate sublexical conversion systems and lexica for spelling and reading. That is, for spelling there is a unique phonological lexicon that is independent from the phonological lexicon that is used for reading just as there are unique orthographic lexicons for both reading and spelling. Additionally, there is a sublexical orthography-phonology conversion system for reading and a separate sublexical phonology-orthography system used for spelling. An important addition to this model is that both the orthographic spelling and orthographic reading lexicons pass activation to a shared response buffer. Activation in the response buffer can in turn activate units in both orthographic lexicons. In this way, activation in one system (e.g., spelling) is shared with the other system (e.g., reading) and vice versa. Jones and Rawson refer to this model as the *separate-but-shared model* to differentiate it from the completely shared and the completely separate models. The separate-but-shared model was devised to explain discrepancies in the literature that neither the completely separate nor completely shared models could explain. Additionally, Jones and Rawson show across three experiments where they trained participants in spelling and reading different very low frequency words that the data from reading and spelling tests were better explained by the separate-but-shared account than by either of the other approaches.

In the current study, we chose to use two measures of spelling ability to allow for the possibility that reading and spelling may use different orthographic representations. These two measures are based on the work of Andrews and Hersch (2010) and detailed below it the methods section, but it is necessary to introduce them briefly here. The first is a spelling production task where the participant listened to a spoken word and was asked to type the word. If the reading and spelling systems rely on different orthographic representations, then this test should gauge the quality of the spelling orthographic representations. The other test was a spelling recognition test where the participant was presented with a letter string and had to decide if the letter string spelled a word. The nonwords were formed by changing one to three letters from a real word while keeping the pronunciation intact for most (Andrews & Hersch, 2010). For the purposes of the current work, it is also worth noting that the spelling recognition test is an untimed lexical decision task. In both tasks, the participants are asked to decide if a letter string spells a word. As such, the spelling recognition task has more in common with word recognition than with spelling. A fact that has led some researchers to conclude that classifying spelling recognition as a spelling task instead of a reading task is arguable (Ehri, 1997). For this reason, we believe that the spelling recognition test is a measure of the reading orthographic representations. Using the two measures allows us to measure the quality of the orthographic representations within the reading and spelling systems. If, on the other hand, reading and spelling use the same orthographic representations, then the two measures should be assessing the same thing.

1.4 Predictions from the models

1.4.1 Completely separate model predictions

Now that the three competing models have been delineated, we can consider the prediction each makes with regards to the effect of orthographic quality on phonological processing. For the completely separate model, the orthographic quality of the reading lexicon should influence phonological processing in visual word recognition. The nature of the influence, though, is less clear. At first blush, as orthographic quality increases phonological effects should diminish because the lexical decision can be made rapidly before phonological effects emerge. However, it is important to understand that the spelling recognition test is not a pure measure of reading orthographic representations. Indeed, no such test exists due to the interactivity of the word recognition system. For example, one way of deciding whether a stimulus is a nonword, particularly a pseudohomophone, or a real word is in terms of a spell check. For example, the orthographic input can be compared with an orthographic representation generated from phonology or semantics (Harm & Seidenberg, 2004; Ziegler, Jacobs, & Klüppel, 2001). If the feedback orthographic representation and the input agree, then the input is a word, and if they differ, the input is deemed a nonword. The key is that phonology is used to generate the orthographic comparison, either directly or by way of semantics. For example, the pseudohomophone *drane* will activate the phonological representation /dreɪn/ that will in turn activate *drain*. As *drane* and *drain* do not match, *drane* can be classified as a nonword. It follows that participants with high quality phonological connections will be able to generate a specific orthographic comparison allowing them to correctly reject the nonword. For participants with less well specified phonological connections, the feedback from phonology will not settle on a precise orthographic representation making the spell check less efficient. For example, /dreɪn/ may partially activate *drane* because of the /-eɪn/ overlap in words such as *cane* and *lane*. For words, in addition to relying on phonology to generate an orthographic comparison, participants can also rely on the direct connections between orthography and semantics to generate an orthographic comparison. That is, for words, the input specifies a precise semantic representation that can in turn feedback to specify a precise orthographic pattern to use for the spelling verification. Although the time to achieve coherence between orthographic and semantic nodes is longer than the time to achieve coherence between orthographic and phonological nodes (Van Orden & Goldinger, 1994), the spelling recognition test is untimed, and therefore, time to coherence is less of an issue than it is in a timed lexical decision task where there is pressure to respond as rapidly as possible. Given that word spelling recognition may also be supported by orthographic-semantic connections, the recognition performance to nonwords may be more sensitive to the quality of participants' reading phonological representations. Based on this account of the processes used during the spelling recognition test, an alternative prediction is that individuals with better spelling recognition scores should show larger phonological effects and that this should be especially true for their nonword performance. To foreshadow the results, we found that this was indeed the case. With regards to the quality of the spelling orthographic representations, the predictions of the completely separate model are more straightforward. Any effects should be minimal as the orthographic spelling units do not directly pass activation to the representations within the reading system.

1.4.2 Separate-but-shared model predictions

In the separate-but-shared model, the lexical decision would be made once the orthographic information in the shared response buffer reaches threshold. Thus, orthographic information coming from the reading and/or spelling systems that boosts the activation level in the response buffer will facilitate responding. The separate-but-shared model holds that spelling recognition should be a measure of the orthographic and phonological representations in the reading system and makes the same predictions as the completely separate model as the orthographic information from the reading system would directly affect the activation levels in the response buffer. However, according to the separate-but-shared model, performance on the spelling production task gives an indication of the quality of the connections in the spelling system. Participants with less well specified phonological and orthographic connections in the spelling system have more difficulty converting a phonological input to an orthographic output as is required in the spelling production task. Conversely, participants that have better specified connections in the spelling system will be able to generate the correct spelling for a given auditory input. In terms of performance in the lexical decision task, the initial input is to the reading system, but assuming the reading system and spelling system share activation as the separate-but-shared account holds, activation will begin to accrue in the spelling system that can in turn influence the lexical decision by feeding back to the shared response buffer. Participants with better orthographic spelling connections will benefit more from the spelling system activation, and this activation may allow the lexical decision response to proceed before the phonological spread effect emerges. That is, the orthographic information in the shared response buffer reaches threshold before the phonological connections in the reading system increase the activation of the orthographic information in the reading system that ultimately feeds activation to the shared response buffer. Conversely, for poor spellers, the activation from the spelling system will not be sufficient to allow a response before phonological effects emerge. This account predicts that the phonological spread effect should emerge relatively late in processing. To test this, we will use reaction time distributional analyses.

1.4.3 Completely shared model predictions

It is now time to consider the predictions from the completely shared model. The one thing that seems to be certain is that that two spelling measures should both be measures of the same orthographic representations and their connections and should interact similarly with phonological processing. That is, the spelling production task measures the orthographic representations as they are accessed from the phonological representations, and the spelling recognition test measures the orthographic representations, possibly as facilitated by a phonological spell check as discussed above. Importantly, the same phonological and orthographic representations would be employed for both spelling measures, leading to the prediction that both should interact with phonological processing in a similar fashion. This seems to be the implicit assumption in much of the past research as the two spelling measures are always combined to give an overall spelling score (e.g., Andrews & Hersch, 2010; Andrews & Lo, 2012, 2013).

As stated above with the separate-but-shared model, poor spellers, as measured by production, could be expected to show larger effects of phonological spread. The problem with this account as applied to the completely shared model is that poor spelling production means the phonological code is not able to fully specify the orthographic code. If the reason people perform poorly on spelling production is because the coupling between phonology and orthography is impoverished, it is not clear how these impoverished connections could in turn support the lexical decision and show up as a phonological spread effect. In fact, one would expect impoverished phonological connections to be associated with a diminished spread effect. This is not a problem for the separate-but-shared account as the reading and spelling system have separate orthographic and phonological representations. So, the phonological representations that underlie the spread effect in word recognition and the phonological representations that support the spelling production task are different. Assuming the phonological spread effect relies on phonology feeding activation to the orthographic level as explained above, the prediction from the completely shared model would be that participants with well specified phonological-orthographic connections should show the larger *P* effect. These would be the participants that score well on the spelling production measure. Similarly, if spelling recognition is affected by a spell check from phonology, then those subjects scoring higher should have better phonological-orthographic connections and show a larger *P* effect.

1.4.4 Predictions summary

To summarize, the completely separate model predicts that only the spelling recognition test performance should interact with the phonological spread effect. Provided spelling recognition is influenced by feedback from phonology, better recognizers should show a larger effect. The completely shared model predicts that both spelling recognition and spelling production should interact in a similar fashion with the phonological spread effect such that good performance on both should show larger effects. The separate-but-shared model predicts that better spellers on recognition and poorer spellers on production should show the largest phonological spread effects.

2. Method

2.1 Participants

Data collection began late in one semester and continued the next semester. Before the experiment, we decided to obtain data until the end of the second semester with the criterion that at least 50 participants would be needed for final analysis. Sixty-six undergraduates at the University of South Alabama served as participants. All participants were native English speakers and reported having normal or corrected to normal vision. Five of these participants were used to train researcher assistants on how to administer the protocol. The data from these five participants were not analyzed. This left

61 participants for potential data analysis. In total, 10 of these participants were excluded from further analysis (six had LDT error rates greater than 20%; three experienced recording errors on one of the tasks; one answered only 15% of the reading comprehension questions correctly). This left 51 participants for the final analyses. For these 51 participants, 29 were female and 22 were male. The age range was 18 – 48 with a mean age or 21.4 years and a median of 20.

2.2 Lexical Decision Stimuli

The stimuli used in the current experiment were originally used by Vitevitch (2007) in a series of auditory word recognition experiments and were subsequently used to investigate phonological processing within visual word recognition (Yates, 2009a; Yates, Friend, & Ploetz, 2008). The stimuli consisted of 92 words that have a CVC phonological structure. For half of the words, a phonological neighbor can be formed across two of the three phoneme positions (P = 2). For the other half, a phonological neighbor can be formed across all three phoneme positions (P = 3). As reported in previous studies (Vitevitch, 2007; Yates, 2009a; Yates et al., 2008), the words are controlled on many variables known to influence visual word recognition (see Table 1). In addition, we verified that the two groups of words did not differ significantly on the more recent subtitle word frequency and subtitle contextual diversity norms of Brysbaert and New (2009), both p > .55. The nonword foils were 92 pseudohomophones (e.g., *darc*) selected from the ARC nonword database (Rastle, Harrington, & Coltheart, 2002). The pseudohomophones did not differ from the words in terms of length. Pseudohomophones were chosen to ensure that participants made the lexical decision by settling on a specific orthographic representation, rather than basing it on familiarity (Yap, Balota, Tse, & Besner, 2008). We also note that previous research has shown that the spread effect is found in the lexical decision task when pseudohomophones are used as foils (Yates, 2009b).

Table 1: *Means and (standard deviations) for the control variables of the experimental stimuli.*

Control Variables	P = 2	P = 3
Frequency of Occurrence[1]		
Kučera and Francis[2]	1.00 (.76)	1.10 (.62)
CELEX[3]	1.14 (.69)	1.15 (.53)
Subtitle[5]	2.65 (.86)	2.71 (.62)
Neighborhood Density		
Phonlogical[2]	8.7 (3.5)	9.2 (1.9)
Orthographic[2]	4.9 (4.2)	5.5 (4.3)
Phonographic[4]	3.1 (3.1)	3.0 (2.8)
Neighborhood Frequency[1]		
Phonological[2]	1.23 (.37)	1.24 (.31)
Orthographic[3]	1.26 (.76)	1.52 (.75)
Phonographic[4]	6.42 (3.39)	6.71 (3.64)
Phonotactic Probability		
Sum of the phones[2]	.116 (.05)	.113 (.04)
Sum of the biphones[2]	.004 (.004)	.004 (.003)
Number of Letters[2]	4.5 (.81)	4.4 (.77)
Familiarity[2]	6.86 (.28)	6.88 (.20)
Orthographic Neighborhood Spread[3]	1.8 (1.1)	2.1 (1.0)
Feedforward Rime Consistency[3]	.83 (.27)	.84 (.25)
Feedback Rime Consistency[3]	.75 (.31)	.66 (.34)
Age of Acquisition[3]	4.1 (1.0)	4.0 (0.9)
Imageability[3]	4.6 (1.4)	5.1 (1.3)
Contextual Diversity[1,4]	.87 (.66)	.93 (.55)
Subtitle Contextual Diversity[1,5]	2.41 (.71)	2.49 (.57)
OLD20[4]	1.7 (.3)	1.6 (.2)

1. Reported in common log values.

2. Based on values reported in Vitevitch (2007).

3. Based on values reported in Yates et al. (2008).

4. Based on the values reported in Yates (2009).

5. Based on the Brysbaert and New (2009) norms.

2.3 Individual Difference Measures

The current study utilized five measures designed to evaluate spelling and reading ability. Spelling ability was measured by a production test and recognition test. Both spelling measures were previously used by Andrews and Hersch (2010). The production test consisted of 20 words selected by Andrews and Hersch (2010) from a study by Burt and Tate (2002), and were correctly spelled by 30% – 90% of their participants. For the production test, participants were required to spell (i.e., produce) the word after hearing it. The spelling recognition test consisted of 44 correctly spelled words and 44 that were incorrectly spelled. Those that were misspelled had one to three of their letters changed, although the pronunciation was often preserved (Andrews & Hersch, 2010). Participants had to decide which words were spelled correctly or incorrectly.

In addition to the spelling measures, participants also read passages and answered comprehension questions about the passages. The passages were taken from (Taylor, 1997). After reading each passage the participants answered 10 true or false questions to test their comprehension of the passage. The total time to read the passage was also recorded, giving an indication of each participant’s reading speed. The participants were also given the Shipley vocabulary test (Shipley, 1940). The Shipley consists of 40 target words that are each presented with four other words. The participants were asked to choose the word most closely related in meaning to the target word.

2.4 Procedure

Participants first completed the lexical decision task. Stimuli were displayed on an IBM compatible computer running E-Prime experimental software (Schneider, Eschman, & Zuccolotto, 2002) and outfitted with a response box to collect participants' lexical decisions. Responses were made with the index finger (word responses) and middle finger (nonwords responses) using the participants' right hand. Participants were instructed to respond as rapidly as possible, but to avoid making mistakes. Before making lexical decisions to the experimental stimuli, the participants first responded to 10 practice stimuli consisting of an equal number of words and nonwords. None of the practice items were included in the experimental trials. Each trial began with a 250ms blank screen followed by a fixation point (a plus sign) that was presented for 750ms. Immediately after the fixation point a letter string was shown and remained on the screen until the participant pressed one of two buttons indicating whether they thought the letter string was a word or nonword. Reaction times were measured from the onset of the stimulus until the participant made a response. The order of the stimuli was randomly determined for each participant.

Following the lexical decision task, the participants completed the individual difference measures. They first read the passages that were used to calculate reading speed and comprehension. Each passage was presented and remained on the screen until the participant pressed a button indicating they were finished reading. Following the passage, 10 comprehension questions were presented one at a time. For each question, the participants pressed one of two keys on the keyboard to indicate whether they thought the sentence was correct or incorrect. After completing the final question, the next passage was shown. The participants were given one practice passage followed by the two passages used to calculate comprehension and reading speed. Next, the participants were given the spelling production test. For each word, the experimenter first read the word to them and then read a sentence containing the word. Following this, the participant typed the word using the computer. For the spelling recognition test, participants were shown either a word spelled correctly or incorrectly and asked to decide if the word was spelled correctly. At the end, the participants were given the Shipley Vocabulary test on a sheet of paper. Participants were instructed to circle the word most closely related to each target word.

3. Results

3.1 Individual Difference Measures

For spelling production, the Levenshtein distance (Levenshtein, 1966) between the participant's spelling and the correct spelling was computed for each word. Levenshtein distance is defined as the minimum number of insertions, deletions, or substitutions required to covert one letter string to another. Using Levenshtein distance provides a more fine-grained measure of spelling production ability than does a simple correct/incorrect distinction. For example, one of the words that participants were asked to spell was *euphoric*. One participant incorrectly spelled it as *euphorec* while another spelled it incorrectly as *uforic*. Simply recording both as incorrect obscures the fact that the first participant is much closer to the correct spelling than is the second participant. A person's accuracy on a word was calculated as 1/(1+Levenshtein distance), hereafter referred to as Levenshtein accuracy. A participant's spelling production score was the average of their Levenshtein accuracy. Spelling recognition was calculated as the sum of the correct responses on the spelling recognition test. Scores could range from 0 – 88. For both reading speed and comprehension, we treated the first passage as practice, and used the last two to calculate participants' speed and comprehension. Reading comprehension was calculated as the sum of the correct responses. Scores could range from 0 – 20. For reading speed, we calculated the words per minute for each passage and took the average. The Shipley vocabulary score was calculated as the total correct with the maximum score being 40. After calculating each of the individual difference measures, we converted each measure to Z scores and used these in all subsequent analyses. Table 2 contains the correlations between each of the individual difference measures.

Table 2: *Correlations between individual difference measures.*

Measures	1	2	3	4	5
1. Spelling Recognition	_				
2. Spelling Production	.729*	_			
3. Reading Speed	.199	.274	_		
4. Reading Comprehension	.567*	.544*	.121	_	
5. Vocabulary	.474*	.582*	.147	.506*	_

Notes: * $p < .001$

3.2 Latency Analyses

3.2.1 Linear Mixed Effects Models

Latencies shorter than 250 ms or longer than 2,500 ms were treated as outliers and removed from all analyses (0.5% of the data). For the latency data, only correct responses were analyzed. Looking at the item latencies and error rates revealed that the word *guise* had an accuracy rate of 27%, and it was not used in any of the analyses. The accuracy rate to all other words was greater than 50%. The word *lurch* had a mean reaction time that was over 3 standard deviations from the mean of its condition and was not used for any analyses. The mean reaction time to $P = 2$ words was 709 ms with an error rate of 8.7%. For the $P = 3$ words, the mean reaction time was 677 ms with an error rate of 4.3%. These facilitative effects for spread were significant for both reaction times $F(1, 50) = 27.17$, $p < .001$ and error rates $F(1, 50) = 37.21$, $p < .001$, replicating previous work with phonological neighborhood spread in the visual lexical decision task (Yates, 2009a).

To evaluate the relationship between the spread effect and each of the individual difference measures we used linear mixed effects models (Baayen, Davidson, & Bates, 2008). Before running the model, we transformed the reaction times using the inverse transformation multiplied by -1000 (i.e., -

1000/RT). Multiplying by -1000 removes decimals and reflects the inverse reaction times so that larger inverse reaction times indicate slower responding. The model included spread and the interaction between spread and each of the individual difference measures as fixed effects. Additionally, we included the inverse of the previous reaction time as a fixed effect as previous research has shown a strong effect of previous reaction time on responding (Baayen & Milin, 2010). As there is no previous reaction time for the first trial, we eliminated these from the linear mixed effects models. The analyses reported are based on random intercepts for both subjects and items, and by subject random slopes for Spread. All variables were mean centered. The spread factor was coded P = 2 (.5) and P = 3 (-.5). This led to the following model.

*invRT ~ Spread*Spelling Production + Spread*Spelling Recognition + Spread*Reading Comprehension +Spread*Reading Speed + Spread*Vocabulary + Inverse Previous RT + (1+Spread|Subject) + (1|Word)*[1]

Effects with t > 2 are considered significant at the $p < .05$ level (Baayen & Milin, 2010). All analyses were conducted using the lme4 package (Version 1.1-9; Bates, Mächler, Bolker, & Walker, 2015) in the R programming environment (Version 3.2.1). We first ran the data through the model and removed any data points with absolute standardized residuals exceeding 2.5 standard deviations (Baayen & Milin, 2010), resulting in the removal of 76 data points (1.8% of the data). We then ran the model again. The parameter estimates, standard errors, and t-values are in Table 3. The main effect of spread was again significant. This replicates previous work by indicating that there is a facilitative effect of spread such that P3 words are responded to more rapidly than P2 words. The effect of previous reaction time was

[1] We also ran the model with the addition of by item random slopes for each of the individual difference measures. The conclusions were the same based on the more complex model. In all other models reported in this paper, the inclusion of by item random slopes for the individual difference measures resulted in a failure to converge.

also significant. For the individual difference measures, reading speed was significant, indicating that as reading speed increased lexical decision latencies decreased. None of the other main effects of the individual difference measures proved significant. Importantly, however, both spelling measures interacted with the spread effect. The facilitative spread effect was maximal when spelling recognition was large. Conversely, the facilitative spread effect was minimal when spelling production was large. None of the other interactions were significant.

Table 3: LMM estimates of the fixed effects with all individual difference measures included

	Estimate	SE	t
(Intercept)	-1.611	0.036	-44.68
Spread	0.083	0.036	2.28
Spelling Production	0.033	0.052	0.64
Spelling Recognition	-0.065	0.049	-1.34
Reading Comprehension	0.081	0.041	1.97
Reading Speed	-0.099	0.033	-2.98
Vocabulary	-0.027	0.041	-0.66
Inverse Previous RT	0.154	0.013	12.11
Spread:Spelling Production	-0.046	0.017	-2.72
Spread:Spelling Recognition	0.036	0.016	2.31
Spread:Reading Comprehension	-0.012	0.013	-0.92
Spread:Reading Speed	0.004	0.011	0.45
Spread:Vocabulary	0.016	0.013	1.24

To better understand why the two spelling measures interact in opposite directions with the spread effect, we conducted principal components analysis to generate two orthogonal measures of spelling. Both spelling measures had high positive correlations with the first principal component (PC1; both $r = .93$), and this component accounted for 86% of the variance. The second principal component (PC2) correlated positively with production ($r = .368$) and negatively with recognition ($r = -.368$) and accounted for the remaining 14% of the variance. This component captures the difference between spelling recognition and spelling production when general spelling ability is partialled out. Individuals with low PC2 scores do worse on production than they do on recognition. That is, they are better at spelling recognition than spelling production. We used the following linear mixed effect model to test whether these components interacted with spread.

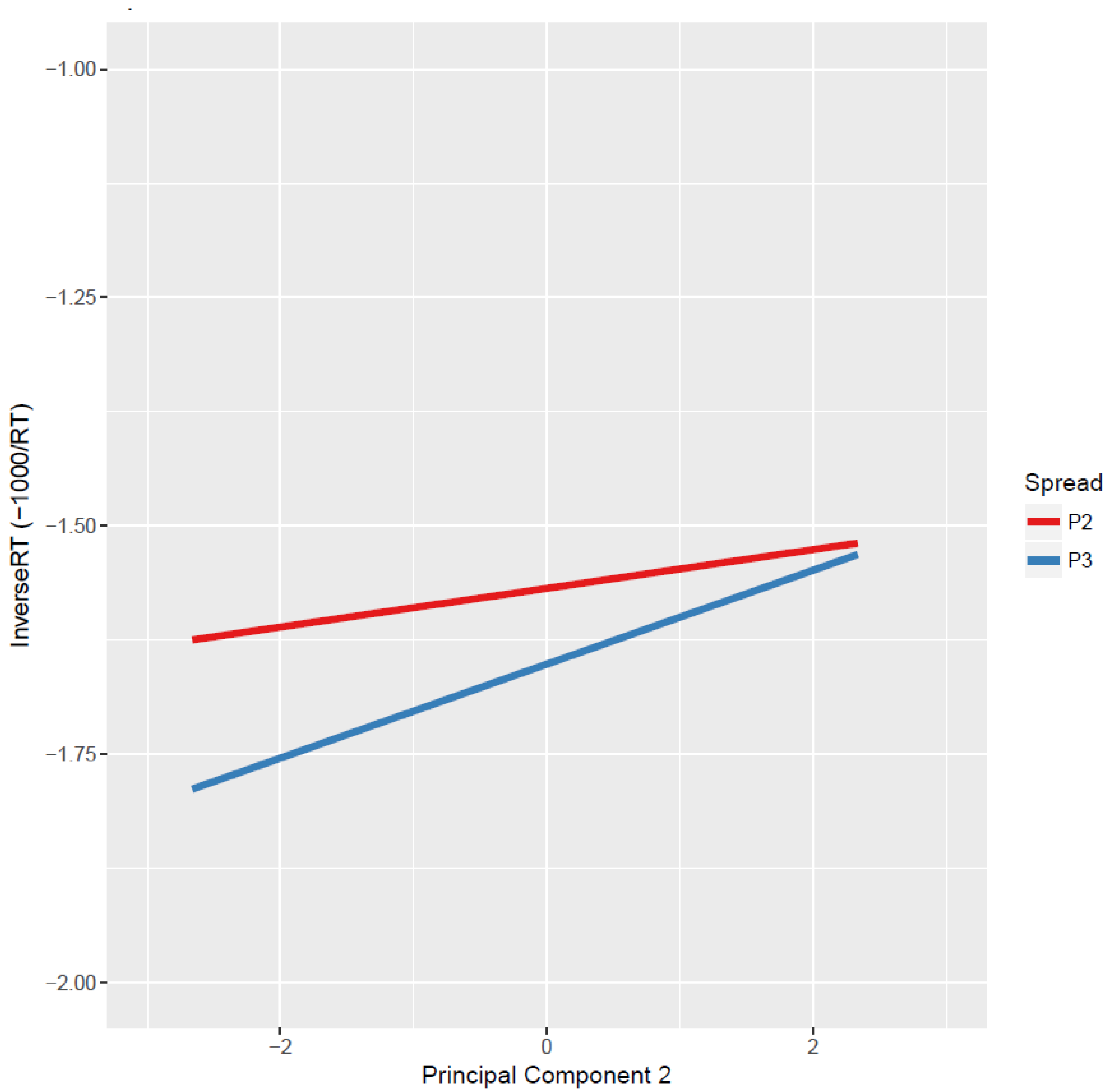
*invRT ~ Spread*PC1 + Spread*PC2 + Spread*Reading Comprehension +Spread*Reading Speed + Spread*Vocabulary + Inverse Previous RT + (1+Spread|Subject) + (1|Word)*

We first ran the model and removed 77 data points (1.8% of the data) that had standardized residuals greater than 2.5 standard deviations and then ran the model on the trimmed data. Table 4 contains the parameter estimates, standard errors, and t-values when the model was run on the trimmed data. As can be seen, PC2 interacted with spread and indicates that participants low on production and high on recognition showed a larger spread effect. Figure 1 plots the spread by PC2 interaction. To produce the plot we used the R package sjPlot (Version 2.4.1, Lüdecke, 2018).

Table 4: LMM estimates of the fixed effects with principal components of the spelling measures

	Estimate	SE	t
(Intercept)	-1.611	0.036	-44.68
Spread	0.083	0.036	2.28
PC1	-0.030	0.044	-0.67
PC2	0.036	0.033	1.11
Reading Comprehension	0.081	0.041	1.97
Reading Speed	-0.099	0.033	-2.98
Vocabulary	-0.027	0.041	-0.66
Inverse Previous RT	0.154	0.013	12.11
Spread:PC1	-0.009	0.014	-0.63
Spread:PC2	-0.030	0.011	-2.85
Spread:Reading Comprehension	-0.012	0.013	-0.92
Spread:Reading Speed	0.005	0.011	0.46
Spread:Vocabulary	0.016	0.013	1.24

Figure 1. Spread X PC2 interaction



3.2.2 Testing for influential cases

To make sure that the effects reported are not due to outliers on the predictors, we calculated the centered leverage values for each participant. To conduct the analysis, we calculated a mean difference score (P = 2) – (P = 3) for each participant and regressed this difference score on the two principal components, reading comprehension, reading speed, and vocabulary. Various suggestions have been given for the appropriate point above which a leverage value is considered high. Stevens (2012) suggested 3p/n. A more conservative estimate is 2p/n (Hoaglin & Welsch, 1978). Using the more conservative criterion, three participants could be considered to have high leverage. We removed these three participants and reran the linear mixed effect model above that contains the two principal components. Importantly, the spread X PC2 interaction was still significant (b = -.028, t = -2.16). Although assessing leverage is important, it does not tell you whether a subject unduly influenced the model as a whole. To determine influence, we used R package influence.ME (version 0.9.9; Nieuwenhuis, Pelzer, & te Grotenhuis, 2017) to calculate Cook's D for the linear mixed effect model containing the principal components. As with leverage, different cutoff points have been suggested for Cook's D. A common one is anything greater than 1 should be cause for concern (Cook & Weisberg, 1982). None of the subjects had a Cook's D close to 1. A more conservative approach is to look at any cases with a Cook's D greater than 4/n. Based on this criterion, two subjects could be considered influential. We removed these two subjects and reran the linear mixed effects model containing the two principal components. Again, the spread X PC2 interaction remained significant (b = -.033, t = -3.17). As a final note, we would like to emphasize that in all linear mixed effects models in this research we always fit the model, then removed any data points with absolute standardized residuals exceeding 2.5 standard deviations (Baayen & Milin, 2010), and then refit the model and reported those results.

3.2.3 Distributional analyses

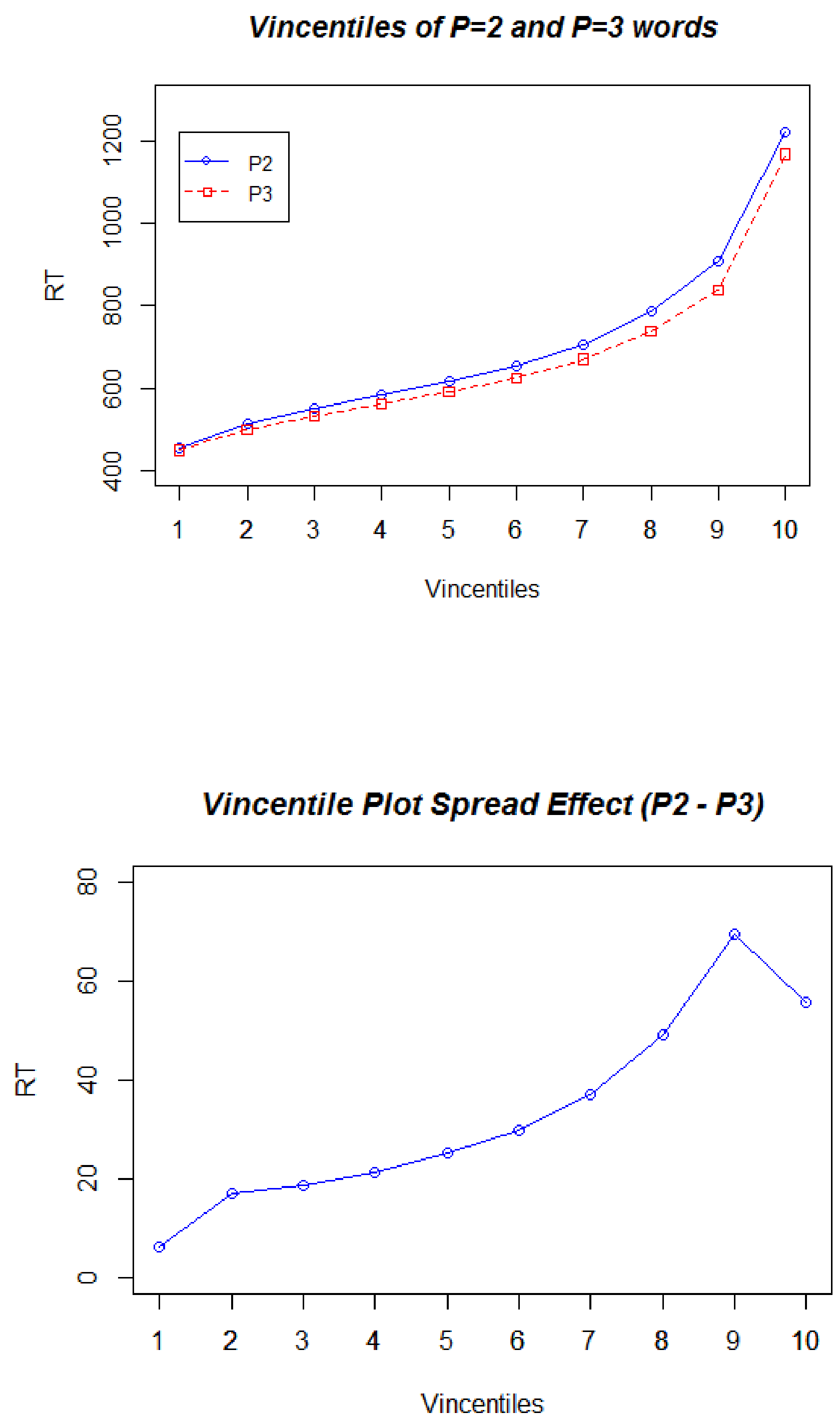
We now turn to the results of the distributional analyses. Each participant's reaction time distribution was fit to an ex-Gaussian distribution. The ex-Gaussian is a convolution of the Gaussian and exponential distributions and has been shown to be useful in understanding reaction time distributions

(Balota & Yap, 2011). For each participant, we obtained the mean and standard deviation (μ and σ) of the Gaussian and the mean and standard deviation of the exponential distribution (τ).

We analyzed the correct reaction times for each participant using the QMPE software (Version 2.18, Cousineau, Brown, & Heathcote, 2004). The parameter estimates were derived using the quantile maximum likelihood estimation procedure as this has been shown to be less biased than continuous maximum likelihood with small samples (Heathcote, Brown, & Mewhort, 2002). For each participant the maximum possible number of quantiles was used to minimize bias and maximize efficiency (Heathcote et al., 2002). We fit the $P = 2$ and $P = 3$ reaction times separately. All fits converged within 250 iterations, but for four of the fits, the Hessian was singular, and the parameter estimates were questionable, so we removed the four participants with questionable fits from further analysis.

First, we ran paired samples t-tests to compare the parameter estimates for $P = 2$ and $P = 3$ responses. For μ, the $P = 2$ distributions ($M = 497$, $SD = 89$) did not differ significantly from $P = 3$ distributions ($M = 485$, $SD = 100$), $t\ (46) = 1.68$, $p = .099$. Likewise, for the σ parameter, the $P = 2$ parameter estimates ($M = 57$, $SD = 34$) did not differ significantly from the $P = 3$ estimates ($M = 49$, $SD = 38$), $t\ (46) = 1.69$, $p = .097$. For the τ parameter estimates, there was a significant difference between $P = 2$ ($M = 203$, $SD = 95$) and the $P = 3$ ($M = 182$, $SD = 100$), $t\ (46) = 2.24$, $p = .030$. We verified the ex-Gaussian analyses with descriptive Vincentile plots. To produce the Vincentile plots we ordered each participant's reaction times from fastest to slowest for the $P = 2$ and $P = 3$ reactions times. For each participant, we calculated the mean of the fastest 10% (Vincentile 1), the next fastest 10 % (Vincentile 2), etc. We then calculated the mean for the Vincentiles across participants to produce the Vincentile plots found in the top panel of Figure 2. The Vincentiles for the spread effect were calculated by subtracting the $P = 2$ and $P = 3$ Vincentiles and are plotted in the bottom panel. Both plots agree with ex-Gaussian analyses and show that the spread effect is largest in the right tail of the distribution.

Figure 2. Vincentile plot for lexical decision performance as a function of Spread (top panel) and the Spread effect (bottom panel).



To relate the spelling measures to the parameter estimates, we calculated a difference score (*P* = 2*)* – (*P* = 3) for each of the three parameter estimates. A series of regression analyses was conducted with the two spelling measures as predictors and the parameter difference scores as the criterion. For both the regression analysis using the μ difference scores and the σ difference scores neither spelling measure was a significant predictor. When the τ difference score was used as the criterion, spelling production was a significant predictor ($\beta = -.417$, $p = .048$). Spelling recognition was not significant, although it was in the opposite direction to spelling production ($\beta = .251$, $p = .226$) congruent with the previous analyses. We also conducted a series of regression analyses with the difference scores as the criterion and the two orthogonal principal components used in the mixed effects model as predictors. When the τ difference scores were used, the significance of PC2 was just above the traditional .05 cutoff ($\beta = -.254$, $p = .085$), all other $p > .30$.

3.2.4 Spelling recognition of words versus nonwords

Finally, we evaluated how spelling recognition as a function of wordness interacted with the spread effect. We divided the spelling recognition variable into number correct for the words and the number correct for the nonwords (e.g., correctly indicating that a nonword does not spell a word). First, we noted that the correlation between these two is quite low ($r = .14$, $p = .32$) indicating that in the spelling recognition test, recognizing a word and rejecting a nonword may be tapping different mechanisms[2]. We reran the LMM with the individual difference measures, but this time we broke

[2] We note that the correlation between P2 word accuracy and pseudohomophone accuracy in the lexical decision task was $r = .16$, $p = .25$. The correlation between P3 word accuracy and pseudohomophone accuracy was likewise small $r = .15$, $p = .29$. However, the correlation between P2 word and P3 word accuracy was large $r = .58$, $p < .001$. All this taken together indicates that accuracy to words and nonwords appear not to be related, supporting our claim that there are different influences on the recognition of words and nonwords.

spelling recognition into performance for words and performance for nonwords. All the effects significant in the first analysis (see Table 1) remained significant in the new analysis. Importantly, nonword recognition performance interacted with spread $b = .03$, $t = 2.46$, but word recognition performance did not $b = .004$, $t = .30$.

4. Discussion

The results of the experiment replicate previous research showing an effect of phonological spread on lexical decisions. Importantly, the results show that the size of the spread effect depended on spelling ability and how spelling ability was measured. For spelling production, the spread effect was larger for those scoring lower on the measure. Distributional analyses indicated that the spread effect emerged late in processing, and that this late effect was larger for participants with poor spelling production ability. This supports the claim that when the orthographic spelling representations are high quality they can feed back to the shared response buffer in the separate-but-shared model facilitating the lexical decision before the slower phonological effects in the reading system emerge. Interestingly, spelling recognition interacted in the opposite direction with spread. The spread effect was larger for those scoring high on the measure. Additionally, we showed that the spelling recognition performance to the nonwords (e.g., correctly indicating that the nonword was not a word) was responsible for the interaction, supporting our claim that the spelling recognition task is a measure of the orthographic-phonological connections in the reading system. The results of the principal components analysis highlighted the difference between the spelling recognition and the spelling production task with regard to the spread effect. The only component that interacted with spread was PC2 that captured the differences between the two spelling measures.

Historically, there have been two competing models with regard to shared representations between reading and spelling systems. The completely separate model assumes that spelling and reading do not share any representations in common. This model predicts that only the spelling recognition test performance should interact with the phonological spread effect. This was clearly not the case. As such, our results do not support the claims of the completely separate model. The second model, the completely shared model, holds that the spelling and reading systems rely on the same representations and predicts that both spelling recognition and spelling production should interact in a similar fashion with the phonological spread effect. This prediction is also not in agreement with our results. Nevertheless, this view seems to be the assumption in most of the literature as the two spelling measures are often combined to give and overall spelling score (e.g., Andrews & Hersch, 2010). Had that been the approach in the present research we would have missed the differential effects of the two spelling measures. For this reason, we think it is important in future research to consider individual difference measures in isolation rather than combining them. A more recent theoretical bridge between these two antithetical viewpoints is the separate-but-shared model. According to this model, the reading and spelling systems have their own unique representations. In that sense, it is like the completely separate model. However, within the separate-but-shared model the orthographic units in both the spelling and reading lexicons pass activation to a shared response buffer. As activation accrues in the shared response buffer, it in turn passes activation to both the orthographic spelling and orthographic reading lexicons. This means activation in one lexicon affects the activation within another and that both lexicons affect the rate at which orthographic information in the shared response buffer reaches threshold for responding. The separate-but-shared model predicts phonological neighborhood spread effects should be larger for good spellers in terms of recognition and bad spellers in terms of production. This prediction is congruent with our findings. According to the separate-but-shared model the spelling production task would be primarily dependent upon the units in the spelling system whereas spelling recognition and visual lexical decision would be primarily dependent upon the units in the reading system. This distinction is important in explaining why the phonological neighborhood spread effect interacts in opposite ways with the two spelling measures.

One of the contributions of this research is to show that spelling recognition and spelling production interact with phonological processing, as indexed by the spread effect, in very different ways. The spread effect is maximal when spelling recognition is high and spelling production is low. This was highlighted most clearly in the principal components analysis where the component capturing the difference between the two spelling measures interacted with spread. It is worth reiterating that the first principal component accounted for 86% of the variance in the two spelling measures and indicates that, in general, participants that do well on one measure tend to do well on the other measure. This is hardly surprising. In fact, the high correlation between the two measures is predicted by the separate-but-shared model. Although this model allows for different orthographic representations in the reading and spelling systems, the two are connected through a shared response buffer. Because of this, practice in one system will interact with and shape the representations in the other. For example, spelling practice will, of course, influence the orthographic representations in the spelling system, but it will also influence the orthographic representations in the reading system, albeit to a lesser degree. Thus, spelling recognition that measures the quality of the orthographic representations in the reading system and spelling production that measures the quality of the orthographic representations in the spelling system should be correlated. In the current sample, this correlation was $r = .729$. It's important to note, however, that $r^2 = .53$, indicating that half the variance in either measure in unexplained by the other. It is this unexplained variance between the two that allows the measures to interact in opposite ways with spread. The difference between these two measures is captured by the second principal component, and this component interacted with spread. We need to be clear that we are not interpreting the second principal component as being indicative of two different types of readers (i.e., good production/poor recognition vs poor production/good recognition). This can be a problem when one mentally dichotomizes a continuous variable. Instead, the second principal component captures the relative strength in spelling recognition vs spelling production independent of the similarity between the two measures. That is, even though a given participant may be a good speller and have done well on both spelling measures, chances are they will have done better on one spelling measure relative to the other. This difference is what the second principal component captures. The interaction between the second principal component and spread indicates that the phonological spread effect is largest in readers whose recognition system is better than their production system.

Although we have discussed the spread effect in terms of distributed representations, we do not mean to imply that these results could not be explained in terms of a localist architecture such as that used within the dual-route cascaded (DRC) model (Coltheart et al., 2001). In fact, the DRC model has been used to successfully simulate phonological neighborhood effects in the naming task (Mulatti, Reynolds, & Besner, 2006; Yates, 2010; Yates et al., 2008). In terms of the present research, however, it is not clear how the model could simulate the spread effect in lexical decision as the model does not include units that represent the pairing of phonemes (i.e., CV_, _VC, C_C). An additional problem that localist models would seem to have in relation to the current research is how to best model individual differences. The representations within localist models are not learned but are instead instantiated by the modeler with the assumption that the representation is well specified. If localist models are going to simulate individual differences by varying the quality of the representations within the models, a principled account of how these representations came to vary needs to be provided (for a similar argument regarding how GPC rules are formed in DRC, see Zevin & Seidenberg, 2006). Another possible way to simulate individual differences with a localist architecture is to allow the parameters in the model that control activation flow between units to vary (for example see Adelman, Sabatos-DeVito, Marquis, & Estes, 2014). Future work will need to make clear how the representations and/or parameters came to vary within the model as this will be key to fully understanding individual difference effects on word recognition within localist models.

It is worth noting that there are potentially other ways to explain why spelling recognition and spelling production interact in opposite ways with the phonological neighborhood spread effect. For example, in our account we have assumed that the lexical decision is made in a similar fashion by all participants. This is an implicit assumption in virtually all lexical decision research, but this may well be incorrect. We know that nonword environment changes the strategies used in making the lexical decisions (Yap et al., 2008). Likewise, it is possible that different subjects use different strategies based on their individual strengths and weaknesses in terms of word recognition and reading. Maybe some base lexical decisions on orthographic information and others base lexical decisions on phonological information. The common assumption is that lexical decisions are based on orthographic activation, but Rastle and Brysbaert (2006) have challenged this by showing that phonological priming effects can be best understood in the DRC model when lexical decisions are based on phonological activation. This is true even when pseudohomophones serve as the nonwords. The reason is because although the sublexical pathway will activate /dreɪn/ for both *drane* and *drain*, the direct lexical pathway will only activate /dreɪn/ when *drain* is presented. In terms of the current research, one could argue that in the spelling recognition test the initial flow of activation is from orthography to phonology. If the orthographic form is able to specify a phonological form, then the input can be recognized as a word. Thus, those that do well on spelling recognition have orthographic representations that precisely specify the correct phonological representation and because of this they base the lexical decision on phonological information. However, our finding that the nonword spelling recognition performance is really what interacts with spread would argue against this as it indicates the spelling recognition test, and presumably its timed counterpart the lexical decision task, are based on orthographic information that is influenced by phonological feedback. One could also argue that some participants instead base lexical decisions on orthographic activation and that those most likely to do this are those who do well in spelling production as they have well specified orthographic representations within a shared reading/spelling system. The trouble with this interpretation, however, is it is not clear how to reconcile this with the finding that spelling recognition interacts in the opposite direction. Although our results do not support the idea that how the lexical decision task is approached is dependent upon individual participant strategies, we find the idea intriguing and believe it should be a topic of future research. Most researchers find acceptable that the strategies participants use in the lexical decision task are modulated by the types of words and nonwords used. Word recognition is a two-way street, though. We need to consider how participant strategies are modulated by participant characteristics. As a field, we have spent most of our time concerned with the items we use with little regard to participant characteristics. The recent work by others such as Andrews and colleagues (Andrews & Hersch, 2010; Andrews & Lo, 2012, 2013) as well as the work presented here indicates that it is time to start seriously considering how subject characteristics influence word recognition.

References

Adelman, J. S., Sabatos-DeVito, M. G., Marquis, S. J., & Estes, Z. (2014). Individual differences in reading aloud: A mega-study, item effects, and some models. *Cognitive Psychology*, *68*, 113–160. https://doi.org/10.1016/j.cogpsych.2013.11.001

Allport, D. A., & Funnell, E. (1981). Components of the Mental Lexicon. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *295*(1077), 397–410. https://doi.org/10.1098/rstb.1981.0148

Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(2), 234–254. https://doi.org/10.1037/0278-7393.18.2.234

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

Andrews, S., & Lo, S. (2012). Not all skilled readers have cracked the code: Individual differences in masked form priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 152–163. https://doi.org/10.1037/a0024953

Andrews, S., & Lo, S. (2013). Is morphological priming stronger for transparent than opaque words? It depends on individual differences in spelling and vocabulary. *Journal of Memory and Language*, *68*(3), 279–296. https://doi.org/10.1016/j.jml.2012.12.001

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological*

Research, *3*(2), 12–28.

Balota, D. A., & Yap, M. J. (2011). Moving Beyond the Mean in Studies of Mental Chronometry: The Power of Response Time Distributional Analyses. *Current Directions in Psychological Science*, *20*(3), 160–166. https://doi.org/10.1177/0963721411408885

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, Instruments & Computers*, *41*(4), 977–990. https://doi.org/10.3758/brm.41.4.977

Burt, J. S., & Jared, D. (2015). The role of lexical expertise in reading homophones. *The Quarterly Journal of Experimental Psychology*, *0218*(September), 1–45. https://doi.org/10.1080/17470218.2015.1062528

Burt, J. S., & Tate, H. (2002). Does a reading lexicon provide orthographic representations for spelling? *Journal of Memory and Language*, *46*(3), 518–543. https://doi.org/10.1006/jmla.2001.2818

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

Cook, R. D., & Weisberg, S. (1982). *Residuals and influence in regression*. New York: Chapman and Hall.

Cousineau, D., Brown, S., & Heathcote, A. (2004). Fitting distributions using maximum likelihood: methods and packages. *Behavior Research Methods, Instruments, & Computers : A Journal of the Psychonomic Society, Inc*, *36*(4), 742–756. https://doi.org/10.3758/BF03206555

Ehri, L. C. (1997). Learning to read and learning to spell are one and the same, almost. In C. A. Perfetti, L. Rieben, & M. Fayol (Eds.), *Learning to Spell: Research, Theory, and Practice Across Languages* (pp. 237–268). Mahwah, NJ: Erlbaum.

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience*, *7*(5), 555–562. https://doi.org/10.1038/nn1224

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*(3), 662–720. https://doi.org/10.1037/0033-295X.111.3.662

Haxby, J. V, Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science*, *293*(5539), 2425–2430. https://doi.org/10.1126/science.1063736

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2002). Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin & Review*, *9*(2), 394–401. https://doi.org/10.3758/BF03196299

Hoaglin, D. C., & Welsch, R. E. (1978). The hat matrix in regression and ANOVA. *The American Statistician*, *32*(1), 17–22.

Holmes, V. M., & Carruthers, J. (1998). The Relation between Reading and Spelling in Skilled Adult Readers. *Journal of Memory and Language*, *39*(2), 264–289. https://doi.org/10.1006/jmla.1998.2583

Howard, D., & Nickels, L. (2005). Separating input and output phonology: Semantic, phonological, and orthographic effects in short-term memory impairment. *Cognitive Neuropsychology*, *22*(1), 42–77.

https://doi.org/10.1080/02643290342000582

Jones, A. C., & Rawson, K. A. (2016). Do reading and spelling share a lexicon. *Cognitive Psychology*, *86*, 152–184. https://doi.org/10.1016/j.cogpsych.2016.02.003

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady*, *10*, 707–710.

Lüdecke, D. (2018). sjPlot: Data Visualization for Statistics in Social Science. Retrieved from https://cran.r-project.org/package=sjPlot

Martin, R. C., Lesch, M. F., & Bartha, M. C. (1999). Independence of Input and Output Phonology in Word Processing and Short-Term Memory. *Journal of Memory and Language*, *41*(1), 3–29. https://doi.org/10.1006/jmla.1999.2637

Mulatti, C., Reynolds, M. G., & Besner, D. (2006). Neighborhood effects in reading aloud: New findings and new challenges for computational models. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(4), 799–810. https://doi.org/10.1037/0096-1523.32.4.799

Nieuwenhuis, R., Pelzer, B., & te Grotenhuis, M. (2017). Influence. ME: tools for detecting influential data in mixed effects models. Retrieved from https://cran.r-project.org/web/packages/influence.ME/

Perfetti, C. A., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189–213). Amsterdam: John Benjamins.

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review*, *114*(2), 273–315. https://doi.org/10.1037/0033-295X.114.2.273

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, *103*(1), 56–115. https://doi.org/10.1037/0033-295X.103.1.56

Rastle, K., & Brysbaert, M. (2006). Masked phonological priming effects in English: Are they real? Do they matter? *Cognitive Psychology*, *53*(2), 97–145. https://doi.org/10.1016/j.cogpsych.2006.01.002

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC nonword database. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, *55*(4), 1339–1362. https://doi.org/10.1080/02724980244000099

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools Inc.

Sears, C., Hino, Y., & Lupker, S. J. (1995). Neighborhood size and neighborhood frequency effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(4), 876–900. https://doi.org/http://dx.doi.org/10.1037/0096-1523.21.4.876

Sears, C., Hino, Y., & Lupker, S. J. (1999). Orthographic neighbourhood effects in parallel distributed processing models. *Canadian Journal of Experimental Psychology*, *53*(3), 220–230. https://doi.org/10.1037/h0087311

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*(4), 523–568. https://doi.org/10.1037/0033-295X.96.4.523

Shipley, W. C. (1940). A Self-Administering Scale for Measuring Intellectual Impairment and Deterioration. *The Journal of Psychology*, *9*(2), 371–377. https://doi.org/10.1080/00223980.1940.9917704

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly*, *24*(4), 402–433. https://doi.org/10.2307/747605

Stevens, J. P. (2012). *Applied multivariate statistics for the social sciences*. Routledge.

Taylor, S. E. (1997). *Visagraph Eye Movement Recording System*. Huntington City, NY: Taylor Associates, Inc.

Unsworth, S. J., & Pexman, P. M. (2003). The impact of reader skill on phonological processing in visual word recognition. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *1*, 63–81. https://doi.org/10.1080/02724980244000206

Van Orden, G. C., & Goldinger, S. D. (1994). Interdependence of form and function in cognitive systems explains perception of printed words. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(6), 1269–1291. https://doi.org/10.1037/0096-1523.20.6.1269

Vitevitch, M. S. (2007). The spread of the phonological neighborhood influences spoken word recognition. *Memory & Cognition*, *35*(1), 166–175. https://doi.org/10.3758/BF03195952

Waters, G. S., & Seidenberg, M. S. (1985). Spelling-sound effects in reading: Time-course and decision criteria. *Memory & Cognition*, *13*(6), 557–572. https://doi.org/10.3758/BF03198326

Weekes, B., & Coltheart, M. (1996). Surface dyslexia and surface dysgraphia: Treatment studies and their theoretical implications. *Cognitive Neuropsychology*, *13*, 277–315. https://doi.org/10.1080/026432996382033

Yap, M. J., Balota, D. A., Tse, C.-S., & Besner, D. (2008). On the additive effects of stimulus quality and word frequency in lexical decision: Evidence for opposing interactive influences revealed by RT distributional analyses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(3), 495–513. https://doi.org/10.1037/0278-7393.34.3.495

Yates, M. (2005). Phonological neighbors speed visual word processing: Evidence from multiple tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(6), 1385–1397.

https://doi.org/10.1037/0278-7393.31.6.1385

Yates, M. (2009a). Phonological neighborhood spread facilitates lexical decisions. *Quarterly Journal of Experimental Psychology*, *62*(7), 1304 – 1314. https://doi.org/10.1080/17470210902725746

Yates, M. (2009b). Phonological neighbourhood spread facilitates lexical decisions. *Quarterly Journal of Experimental Psychology (2006)*, *62*(7), 1304–1314. https://doi.org/10.1080/17470210902725746

Yates, M. (2010). Investigating the importance of the least supported phoneme on visual word naming. *Cognition*, *115*(1), 197–201. https://doi.org/10.1016/j.cognition.2009.12.002

Yates, M., Friend, J., & Ploetz, D. M. D. M. (2008). Phonological neighbors influence word naming through the least supported phoneme. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(6), 1599–1608. https://doi.org/10.1037/a0011633

Yates, M., Locker, L. J., & Simpson, G. B. (2004). The influence of phonological neighborhood on visual word perception. *Psychonomic Bulletin & Review*, *11*(3), 452–457. https://doi.org/10.3758/BF03196594

Zevin, J. D., & Seidenberg, M. S. (2006). Simulating consistency effects and individual differences in nonword naming: A comparison of current models. *Journal of Memory and Language*, *54*(2), 145–160. https://doi.org/10.1016/j.jml.2005.08.002

Ziegler, J. C., Jacobs, A. M., & Klüppel, D. (2001). Pseudohomophone effects in lexical decision: Still a challenge for current word recognition models. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(3), 547–559. https://doi.org/10.1037/0096-1523.27.3.547