Quantifying the regularities between orthography and semantics and their impact on group and individual-level behavior

Siegelman, Noam, Rueckl, Jay G., Lo, Jason Chor Ming, Kearns, Devin M., Morris, Robin D. and Compton, Donald

The final version of record is available via its DOI: https://dx.doi.org/10.1037/xlm0001109

Quantifying the regularities between orthography and semantics and their impact on group- and individual-level behavior

Noam Siegelman[1], Jay G. Rueckl[1,2], Jason Chor Ming Lo[1],

Devin M. Kearns[1,2], Robin D. Morris[3], & Donald L. Compton[4]

[1]Haskins Laboratories [2]University of Connecticut

[3]Georgia State University [4]Florida State University

Corresponding Author:
Noam Siegelman
Haskins Laboratories
300 George St., New Haven, CT, USA, 06511
E-mail: noam.siegelman@yale.edu

Abstract

Statistical views of reading highlight the link between proficient literacy and the assimilation of various regularities embedded in writing systems, including those in the mapping between print and meaning. Still, orthographic-semantic (O-S) regularities remain relatively understudied, with open questions regarding three issues: (a) how O-S regularities should be quantified, (b) how they impact the behavior of proficient readers, and (c) whether individual differences in sensitivity to these regularities predict reading skills. The goal of the current paper is to address these questions. We start by reviewing previous studies estimating print-meaning regularities, where orthography-to-semantics consistency (OSC) is defined as the mean semantic similarity between a word and its orthographic neighbors. While we adopt this general strategy, we identify a potential confound in previous operational definitions. We therefore offer a modified measure, which we use to examine group-level OSC effects in available datasets of single word recognition and reading for comprehension. Our findings validate the existence of OSC effects but reveal variation across tasks, with OSC effects emerging more strongly in tasks involving a direct mapping of print to meaning. Next, we present a re-analysis of word naming data from 399 second through fifth graders, where we examine individual differences in reliance on O-S regularities and their relation to participants' reading skills. We show that early readers whose naming accuracy is more influenced by OSC (i.e., those who rely more on O-S) have better passage comprehension abilities. We conclude by discussing the role of O-S regularities in proficient reading and literacy acquisition.

Keywords: Orthographic-semantic regularities; Print-meaning mapping; Word recognition; Reading; Individual differences.

1. Introduction

Over the course of the past decades reading research has become increasingly grounded in the notion that proficient reading requires the assimilation of statistical regularities present in the writing system. This statistical view of reading maintains that over the course of literacy acquisition children become gradually sensitive to the regularities available to them in the writing system, and that proficient readers utilize these regularities to read more efficiently (see Arciuli, 2018; Frost, 2012; Sawi & Rueckl, 2019 for reviews). This theoretical framework highlights the role of statistical learning mechanisms in reading, and underlies key questions for reading research: What are the available regularities that are embedded in the written input, how can they be quantified, and how do they impact reading behavior?

The first type of regularities to attract the attention of the reading community – and perhaps the best studied case to date – are the regularities between orthography and phonology (O-P). The assimilation of O-P regularities is considered to play a key role in the ability to convert orthographic strings to the spoken forms they represent (phonological decoding), which is taken to be one of the fundamental skills underlying reading (e.g. Ehri, 2005; Share, 1999). Considerable work has been devoted to the question of how exactly to quantify O-P regularities (e.g. how to capture the difference between the English words *mint* and *pint*): Such work resulted in various operational measures, including those that center on whether an individual grapheme-phoneme correspondence is regular or not (e.g. Baron & Strawson, 1976; Forster & Chambers, 1973), the (continuous) degree of consistency of a given O-P pairing (e.g. Jared et al., 1990; Treiman et al., 1995) and information-theoretic metrics that quantify the uncertainty and unpredictability of orthographic and phonological units (e.g. Protopapas & Vlahou, 2009; Siegelman et al., 2020). There is also an abundance of work showing that proficient readers are impacted by O-P regularities such that they recognize faster and more accurately words that are more O-P transparent (e.g. Glushko, 1979; Jared et al., 1990; Seidenberg, 1985). Empirical data further suggest that readers are sensitive to O-P regularities in multiple grain sizes (Steacy et al., 2018; Treiman & Kessler, 2006), and that developmentally, sensitivity to O-P correspondences is gradually acquired over the course of typical reading acquisition (e.g., Sénéchal et al., 2016; Weekes et al., 2006), with greater reliance on larger grain sizes (e.g., body-rime correspondences) later in development (Treiman et al., 2006). Sensitivity to O-P regularities (or lack thereof) also accounts for individual differences in reading development, as children with poorer reading skills exhibit decreased sensitivity to these regularities (e.g., Siegelman, Rueckl, et al., 2020).

Yet despite their central role in literacy acquisition, O-P correspondences are only one of multiple sources of information present in writing systems: In any given written input there is a plethora of different types of regularities that are concurrently available to readers. Some of these regularities exist within the orthographic domain: some letters co-occur more frequently than others (e.g. Gingras & Sénéchal, 2019) and the same holds for printed words (e.g. Fine & Jaeger, 2013). There are also other (and more subtle) types of regularities within the O-P mapping, such as associations between orthographic units and stress patterns (e.g., Arciuli, 2018). In addition to all these orthographic regularities and O-P associations exist the regularities between *orthography and semantics*, regularities that play a central role in word recognition according to computational models of reading but are still relatively understudied compared to O-P regularities (see review and examples below). The current paper focuses on these regularities in the orthographic-semantic (O-S) mapping.

In particular, the goal of the current paper is to address three outstanding questions in regard to O-S regularities. The first question is how to assess O-S regularities at the word-level. The second question is whether O-S regularities impact the reading behavior of the average proficient reader, to what extent, and under what conditions. The third question is whether and how variability in young readers' sensitivity to O-S regularities is related to individual differences in early reading skills.

The structure of the paper follows these three questions. First, we review previous computational and behavioral work on O-S regularities, which leads us to highlight methodological considerations related to their measurement. We build upon recent studies by Marelli, Amenta, and colleagues (Amenta et al., 2017, 2020; Marelli et al., 2015; Marelli & Amenta, 2018) in an attempt to quantify the O-S consistency of a given word relative to another. In this section we deal with a potential confound we identify in older definitions and propose a modified measure of O-S consistency. Second, we use this modified measure in conjunction with available large-scale datasets to examine the impact of O-S consistency on group-level behavior across different reading tasks. In doing so we examine data from both word recognition tasks (i.e., word naming, lexical decision) and naturalistic text reading for comprehension (measured via eye movements). Third, we analyze individual differences in sensitivity to O-S regularities as reflected in the degree to which early readers are impacted by O-S consistency during word naming. We ask whether young readers differ from one another in their reliance on O-S regularities and, if so, whether these individual differences account for inter-individual variability in emerging reading skills.

2. What Are O-S Regularities and How Can They Be Measured?

Historically, O-S regularities have been studied mostly in the context of *morphological* effects on word recognition. The impact of morphological structure on visual word recognition has been demonstrated in a long tradition of behavioral studies dating back to the 1970s when Murrell and Morton (1974) reported that the recognition of a word is facilitated by the prior presentation of a morphologically-related word and Taft and Forster (1975) showed that nonwords that comprise real affixes and stems (e.g., *dejuvenite*) take longer to reject in a lexical decision task than nonwords with no apparent morphological structure. In the years since those seminal studies, the role of morphology in reading has been investigated using a variety of methodologies, in numerous languages, and in both skilled and developing readers (see Feldman, 1994; Frost, 2012; Rastle & Davis, 2008 for reviews).

To a large extent, this research has been driven by a theoretical perspective that takes morphological structure as a primitive in the organization of the lexicon, and maintains that explicit representation of morphological relations is the source of morphological effects in visual word recognition (see Seidenberg & Gonnerman, 2000 for a historical overview). However, an alternative approach takes morphological structure as an emergent property of the structure of the mapping between orthographic (and phonological) forms and meaning (Plaut & Gonnerman, 2000; Rueckl et al., 1997; Seidenberg & Gonnerman, 2000). From this perspective, sensitivity to morphological regularities arise from the same statistical learning process that gives rise to the effects of O-P regularities. This view originated from the Triangle Model of Reading (Plaut et al., 1996; Seidenberg & McClelland, 1989), which considers learning to read as a process in which readers form associations between words' orthography, phonology and semantics. Thus, according to this model, during literacy acquisition readers assimilate not only O-P regularities but also those in the mapping between orthography and semantics (including morphological regularities). The regularities in these two mappings are gradually assimilated by learners, until, at the end of the developmental trajectory, proficient readers achieve an efficient division of labor between reliance on O-P and O-S pathways (Harm & Seidenberg, 2004).

In terms of behavior studies, early research from this perspective focused on priming between morphological relatives. For example, Rueckl et al. (1997) showed that long-term morphological priming (of the sort studied by Murrell & Morton, 1974) is modulated by orthographic similarity (e.g., *came* is a better prime for *come* than *bought* is for *buy).*

Similarly, Gonnerman et al. (2007) demonstrated that short-term cross-modal morphological priming is graded by both semantic (*boldly-bold, lately-late, hardly-hard*) and phonological (*acceptable-accept, criminal-crime, introduction-introduce*) similarity.

Recent research has extended these findings by showing that similarity gradients influence the recognition of unprimed words as well. For example, Ulicheva et al., (2020) showed that readers are impacted by the degree of the systematicity in the mapping between derivational affixes and lexical categories, such that words with morphemes that are more specific and more diagnostic in regard to their lexical category are read faster and more accurately. Most relevant for the current investigation, Marelli et al. (2015) found that reading times of monomorphemic stem words are impacted by whether their meaning is maintained in words that are orthographically similar to them: Participants recognized stems that are similar in meaning to orthographically similar words (e.g. *farm*, which is semantically related to many words from its orthographic family, *farmer, farming,* etc.) faster than stems with semantically distant (*hardly-hard*) or unrelated (*corner-corn*) orthographic neighbors.

Marelli et al. (2015) ascribed this behavioral effect to differences in *orthography-to-semantics consistency* (OSC) across words[1]. Intuitively, words such as *farm* are consistent in terms of O-S because their *orthography is particular (i.e., diagnostic) regarding their semantics* (words that are orthographically related to *farm* are also semantically related to it); in contrast, the orthographic form *corn* (e.g., in *corner* or *acorn*) is not necessarily particular with regard to semantics. Indeed, Marelli et al. (2015) showed that a measure of OSC (see definition below) predicted lexical decision latencies of morphologically simple words above and beyond their frequency, length, and morphological family size. Note that the

[1] Throughout the paper, we use the abbreviation OSC to refer to operational measures that tap into orthographic-semantic regularities (and the corresponding term PSC to refer to parallel measures of phonology-phonology regularities). We use the abbreviation O-S (and P-S) to refer to the mapping between orthography and semantics (and phonology and semantics) more broadly.

conceptualization provided by Marelli et al. emphasizes how the OSC of a word is not a consequence of morphological relations only. Rather, it is a cumulative measure of whether the orthography of a word is diagnostic of its semantics, a feature that is impacted both by the links between a word and words that are orthographically *and* morphologically related to it (e.g., the relation between *farm* and *farmer*), and by the degree of semantic associations between a word and other orthographically-related words (e.g., the semantic (dis)similarity between *corn* and *corner*). In other words, OSC encompasses both morphological regularities (which form "islands of regularity" in the O-S mapping, Rastle et al., 2000; Rueckl, 2010) and other O-S associations (or, in many cases, their absence).

At the operational level, in a series of recent studies Marelli, Amenta and colleagues (Amenta et al., 2017, 2020; Marelli et al., 2015; Marelli & Amenta, 2018) quantified the OSC of a word as a function of the semantic similarity of that word and each of its orthographic neighbors. Various possible definitions of orthographic neighborhood have been proposed (see, e.g., Yarkoni et al., 2008). Amenta et al., (2020) and Marelli & Amenta (2018) defined neighbors of a given word as the set of words in which that word is embedded (e.g., *corns*, *corner*, *acorn*, etc. are neighbors of *corn*; again note that not all words in a word's orthographic neighborhood are necessarily morphologically related to it, and that depending on the exact definition of neighborhood, not all morphological neighbors of a word are included in the OSC calculation, e.g., *run-ran*). There are also multiple ways to quantify semantic similarity. In their work, Marelli, Amenta, and colleagues used representations of word meaning derived from a Latent Semantic Analysis (LSA) model and computed the semantic similarity between a word and each of its neighbors by calculating the cosine similarity between the vectors representing the word and the neighbor (values ranging from -1 to 1; where -1 represents opposite meanings; 1 represents identical meanings; and 0 reflects no semantic relationship). Then, Marelli, Amenta, et al. quantified a word's OSC as the mean cosine similarity between that word and each of its orthographic neighbors. Thus, the greater the semantic similarity of a word and its orthographic neighbors, the higher its OSC score.

More formally, Marelli and colleagues defined a measure of OSC of a word as "the frequency-weighted average semantic similarity between the vector of a word and the vectors of all words that contain that very same word" (Marelli & Amenta, 2018; p. 1484), a definition that is expressed by the formula:

$$OSC(t) = \frac{\sum_{x=1}^{k} f_{r_x} * \cos(\vec{t}, \vec{r_x})}{\sum_{x=1}^{k} f_{r_x}}$$

where t is the target word ($\vec{t}$ being its LSA representation), r_x is each of its neighbors (with their corresponding $\vec{r_x}$ representations in the semantic space), and f_{r_x} is the frequency of each neighbor.

In their recent work, Marelli and Amenta (2018) calculated OSC estimates per this definition for a large number of English words. They then examined the relation between these estimates and item-level behavior in word recognition tasks, finding that OSC predicted naming and lexical decision RT in data from the English and British Lexicon Projects, even when controlling for word frequency, mean neighbor frequency, family size, and length (Marelli & Amenta, 2018). These findings suggest that Marelli and Amenta's definition of OSC captures properties of the structure of the O-S mapping that are relevant to the processes underlying word recognition.

It is important to note that although the mapping from orthography to semantics is of obvious relevance to visual word recognition, readers can also access semantics by way of phonology. This raises the possibility that the consistency of the phonology-to-semantics (P-S) mapping could also influence word recognition. Indeed, Amenta and colleagues further showed that both OSC and *phonology*-to-semantics consistency (henceforth: PSC) account for inter-item variability in visual word recognition (where PSC is defined as the frequency-weighted average semantic similarity of a word to its *phonological* neighbors; Amenta et al., 2017). This observation demonstrates the impact of the regularities in both O-S and P-S mappings even when controlling for the other source of information, which is important as in alphabetic languages (where graphemes represent sounds with at least some consistency) the two are correlated by definition.

2.1 Marelli, Amenta, et al.'s Definition of OSC: A Potential Confound

Although in the current paper we adopt the general strategy proposed by Marelli, Amenta, and colleagues, we argue that two specific properties of their operational definition may lead to a confound. The first property is that in their definition the word itself is included in its neighborhood; this property is reflected by the fact that words that do not have any neighbors other than the word itself have an estimate of OSC=1 (Marelli & Amenta, 2018). The second property is that Marelli and Amenta's calculation is frequency-weighted (i.e., token-based).

Needless to say, each of these two methodological decisions is justifiable. Thus, there is no a priori reason why one would use a type- or a token-based calculation for computing OSC, and it is likely that either mode of calculation provides important information on O-S associations that the other does not (see, e.g., Chee et al., 2020, for a discussion in the context of O-P regularities). Similarly, the decision to include information about the word itself in the calculation of OSC can be defended on a theoretical ground (e.g., if a stem X is more frequent than a set of semantically-unrelated neighbors that include this stem – say Xy and Xz - in most cases where the orthographic form X is encountered the same meaning is present, and the original OSC definition captures that). We do not claim here that token-based calculations are faulty, or that the frequency of a word itself is by-definition irrelevant for its OSC.

Instead, our claim here is that the combination of these two properties may lead to a confound in prior measures of OSC.

Namely, given that the word itself is included in its neighborhood, and given that the calculation of mean semantic similarity is frequency-weighted, Marelli and Amenta's *OSC is affected by the ratio between the frequency of the word and the frequency of its neighbors*. Importantly, this frequency ratio is not a property specific to the O-S mapping: Frequency is an index of how often the representation of an orthographic word form is associated with representations of how that word is pronounced and what it means, and hence this frequency ratio potentially taps into various types of statistical structure across the O-P, O-S, and P-S mappings (and within the O, P, and S representations). In other words, our concern here is that given this frequency-ratio confound, previous definition of OSC may not capture O-S regularities exclusively. This raises the concern that previous results demonstrating the impact of OSC on reading behavior were related to this frequency ratio and the various sources of information it encompasses, rather than to O-S regularities per-se.

Indeed, the results of new analyses we conducted following those reported by Marelli and Amenta using their estimates of OSC strengthen our concern: For brevity, the full results are reported in Supplementary Materials S1. First, they show that as expected given our analysis of the OSC's definition above, OSC per the previous definition has a medium-size correlation (r = 0.45) with the log-transformed ratio between a word frequency and the mean frequency of its neighbors. Second, and importantly, an analyses of behavioral data from the English Lexicon Project (ELP; Balota et al., 2007) suggest that the effects of OSC (per Marelli & Amenta's definition) on word recognition behavior are no longer observed when controlling for this frequency ratio[2]. Overall, then, a need arises for other definitions of OSC that reflect the structure of the O-S mapping specifically.

2.2 A Modified Definition of OSC

Concretely, in the current paper we suggest a modified measure of OSC, which follows the general analytical strategy by Marelli, Amenta, and colleagues, but applies two methodological modifications. First, it uses a type-based rather than frequency-weighted calculation (i.e., each word was counted once regardless of its frequency). Second, the word itself is not included in its neighborhood. These two modifications are meant to de-confound the OSC measure from the ratio between a word frequency and the frequency of its neighbors. Operationally, then, our measure of OSC is defined as the mean (type-based) cosine similarity between a word and its orthographic neighbors, which is expressed by the formula:

$$OSC(t) = \frac{\sum_{x=1}^{k} \cos(\vec{t}, \vec{r_x})}{k}$$

where t is the target word ($\vec{t}$ being its LSA representation), r_x is each of its neighbors (with their corresponding $\vec{r_x}$ representations), and k is its neighborhood size (i.e., number of neighbors of the target word).

Note that in the analyses below we define orthographic neighborhood as all words in the corpus with an orthographic Levenshtein distance of 1. We use this definition of orthographic neighborhood (rather than Marelli and Amenta's definition, i.e., all words that include a word) to maximize the number of words in the corpus that have at least one

[2] This holds also for gaze durations during reading for comprehension. Mixed-effect models (similar in specification to those reported below) showed no significant effect of Marelli & Amenta's OSC estimates on log-transformed Gaze Duration when controlling for the log-transformed frequency ratio (p > .1). Models and outputs for this analysis are available in the project's OSF page. This null result stands in contrast to the analysis using our alternative definition of OSC (see 'Impact on Eye Movements' section, below), which shows that under this definition, O-S regularities do predict Gaze Durations during reading for comprehension.

neighbor (specifically, many longer do not have neighbors based on Marelli and Amenta's definition, but are more likely to have neighbors under a Levenshtein-distance-based definition). Yet we also note already here that the change in the neighborhood definition likely affected the types of relations that are captured by our OSC estimates (e.g., it should be more sensitive to irregular inflections, e.g., *run-ran*; but less to other forms of suffixation with multiple letters, e.g., *run-runner*). More broadly, we stress that our definition of OSC is only one possible measure of O-S regularities and it should not be taken as an 'optimal' measure that captures the full O-S mapping. Thus, much like any other operational definition, our measure is based on a series of decisions that went into its quantification (including: the use of type-based calculation; the exact definition of neighborhood; the type of semantic space used; etc.), each of them inevitably resulting in a measure that is more sensitive to some forms of O-S relations than others. Our argument in the rest of the paper is simply that our OSC measure is a valid measure of O-S regularities and therefore that the findings below regarding its impact on group- and individual-level behavior reflect the effects of O-S on reading. In the *General Discussion* we return to a lengthy analysis of the properties of the O-S mapping that our OSC measure does and does not capture.

3. Validating the Modified OSC Measure Using Behavioral Data

The goal of this section is to validate our OSC measure by examining whether it predicts reading behavior. We do so by investigating the relation between our estimates of OSC and item-level reading behavior in two sets of analyses. The first focuses on word recognition data from the ELP, with data from lexical decision and word naming tasks. The second examines eye-movements, using data from the Ghent Eye-Tracking Corpus (GECO), a book reading eye-movement database (Cop et al., 2017). The combination of the two datasets enables us to examine whether O-S regularities indeed affects reading behavior, and whether such effects vary across reading tasks (and in what ways).

Note that in addition to our metric of OSC (and control variables), in the analyses below we also include as predictors measures of O-P regularities as well as a measure of PSC. As noted above, the rationale behind this inclusion is that in alphabetic languages, OSC and PSC are correlated by definition, and the degree of this correlation depends on O-P consistency (see also Amenta et al., 2017). Thus, whenever possible, we consider all these regularities (as well as their interactions) in the same models.

3.1 Methods: Estimation of Word Properties

Our estimation of orthographic and phonological neighborhood used a corpus of 117,574 English words and their General American pronunciations (Kearns, 2020; based on data from Fitt, 2001). Target words were all items from the ELP. For each target word, we extracted all neighbors (orthographic and phonological) from the larger corpus, by extracting the list of words with a Levenshtein distance of 1 from the word. Orthographic neighbors were defined based on orthographic forms (i.e., Levenshtein distance of one letter) and phonological neighbors were based on phonological transcriptions (i.e., Levenshtein distance of one phoneme). Distances were calculated using the *stringdist* function from the R package *stringdist* (van der Loo, 2014). Following the work by Marelli, Amenta, and colleagues, in the analysis below we use LSA spaces to quantify semantic similarity (but see *Methodological Considerations* for estimates based on alternative semantic representations). Specifically, we used pre-trained LSA semantic spaces (from Günther et al., 2014), with vectors of 300 dimensions, trained on the TASA corpus, to calculate the mean cosine similarities (using the *LSAfun* R package, Günther et al., 2014). As noted above, OSC was defined as the mean cosine similarity between a word and its orthographic neighbors. In addition, the mean cosine similarity between a word and its *phonological* neighbors served as a measure of *PSC*. In all models below we also included as control variables measures of log-transformed word frequency, word length, log-transformed orthographic and phonological neighborhood sizes, and manner and place of articulation (15 dummy-coded variables). Note that in contrast to the OSC measure by Marelli, Amenta, and colleagues, under our definition OSC was uncorrelated with the log-transformed frequency ratio between a word and its neighbors (r = 0.01), as expected given the modifications applied to its definition. Please refer to the project's OSF page (https://osf.io/3aczx/) for a full list of words with their OSC and PSC estimates, along with data and analysis scripts for all models below.

3.2 Impact on Word Recognition

Here, we used data from all monosyllabic words in the ELP. We focus on monosyllabic words so we can include not only measures of OSC and PSC but also a proxy of O-P regularities (as measures of O-P regularities are non-trivial to define in words with more than one syllable; see e.g. Chateau & Jared, 2003) – but see Supplementary Materials S2 for an alternative strategy[3]. The surprisal (-log(p(phoneme | grapheme))) of the vowel unit (estimates from Siegelman, Kearns, et al., 2020) was used as a measure of O-P regularities. Note that words for which OSC and PSC could not be calculated (i.e., words that did not appear in LSA space, or words with no orthographic or phonological neighbors in the space) were excluded from the analysis. Overall, our analyses included k=5355 items for lexical decision and k=5541 for word naming. Figure 1 presents the distribution of OSC values across monosyllabic words included in the word recognition models, and Table 1 presents the pairwise correlations between the various predictors and control variables. The table shows

[3] An alternative approach is to include *all* words in the analysis and not to include a measure of O-P regularities (but only OSC and PSC). In the Supplementary Materials S2 we report such analysis, conducted on all words in the ELP for which OSC and PSC can be calculated (22680 words for word naming; 21702 for lexical decision). Importantly, in line with the results of the monosyllabic analysis, we found effects of OSC on lexical decision (in both accuracy and RT) but not in word naming. This further validates our conclusions below.

that as expected, OSC and PSC were positively (yet imperfectly) correlated, and that longer words tended to have higher OSC and PSC estimates. In addition, words with smaller orthographic and phonological neighborhoods tended to have lower OSC and PSC values.

Figure 1*. Distribution of OSC Across Monosyllabic Words from the ELP (k=5541 words).*

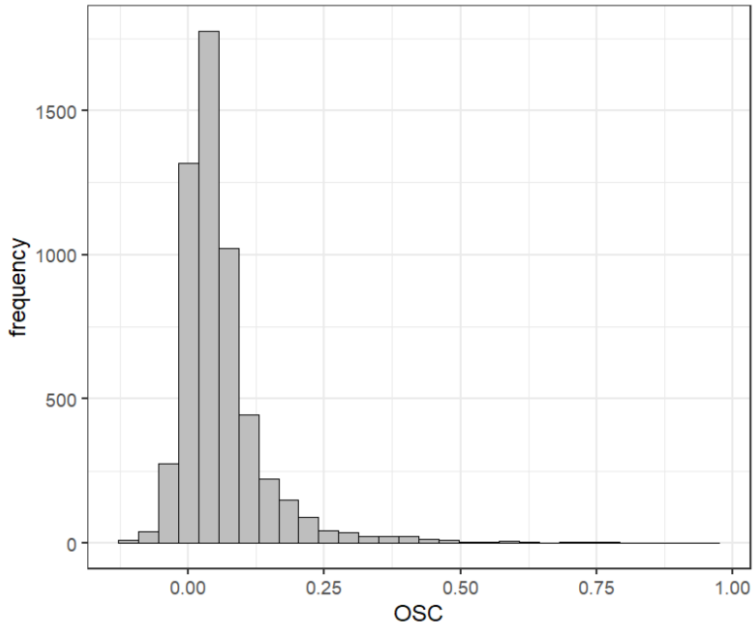


Table 1*: Correlations Between Predictors in Word Recognition Models (k = 5541 words).*

Measure	2	3	4	5	6	7
1) Length	-0.36	0.22	0.24	0.02	-0.66	-0.57
2) Log-transformed frequency		0.09	0.04	0.12	0.29	0.27
3) OSC			0.58	0.01	-0.31	-0.24
4) PSC				0.05	-0.22	-0.41
5) O-P					-0.06	-0.10
6) Log O neigh size						0.66
7) Log P neigh size						

Note: OSC = O-S consistency (modified measure of OSC); PSC = P-S consistency; O-P = O-P regularities (surprisal); Log O Neigh Size = log-transformed orthographic neighborhood size; Log P Neigh Size = log-transformed phonological neighborhood size.

Models on the ELP data reported below use as their basic unit of analysis mean behavior in each item (across subjects): That is, models predicting RT use mean item-level log-transformed RT as the dependent variable, and models predicting accuracy use mean item accuracy (i.e., proportion correct). We refrained from using (generalized) mixed-effect models in the ELP data due to non-convergence of even the most simplified models predicting accuracy. Instead, (generalized) multiple regression models were run to predict the item-level means from the predictors and control variables shown in Table 1, with standard linear models for RT models, and generalized models with a logit link function and a quasi-binomial family for accuracy models. For simplicity, we did not include all possible interactions in our models (e.g., we omitted interactions between frequency and the consistency measures). Instead, we focused on the interactions between OSC, PSC, and O-P surprisal in light of previous studies that highlight the interactive nature of the O-P, P-S, and O-S components of the reading system (e.g., Amenta et al., 2017; Chang & Monaghan, 2019; Harm & Seidenberg, 2004; Strain et al., 1995). We also note that despite the inter-correlations between predictors (see Table 1), in all models reported below multicollinearity was reasonably low, with VIF < 3 for all predictors of interest (i.e., all measures except for the articulation parameters which are by-definition correlated with one another).

The results of the models are presented in Tables 2 and 3 for lexical decision and naming data, respectively. In lexical decision (Table 2), there were significant main effects of all three consistency measures - OSC, PSC, and O-P surprisal - on both response latencies and accuracy. In all cases, words with more consistent mappings (higher OSC, higher PSC, and lower O-P surprisal reflecting higher O-P consistency) had faster and more accurate responses. In addition to these main effects, we found in both RT and accuracy a significant sub-additive two-way interaction between O-S and P-S regularities, with larger effects of OSC in lower levels of PSC – see Figure 2.

Table 2*: Effect of the Modified OSC Measure, O-P and PSC on Lexical Decision RT and Accuracy in ELP Data.*

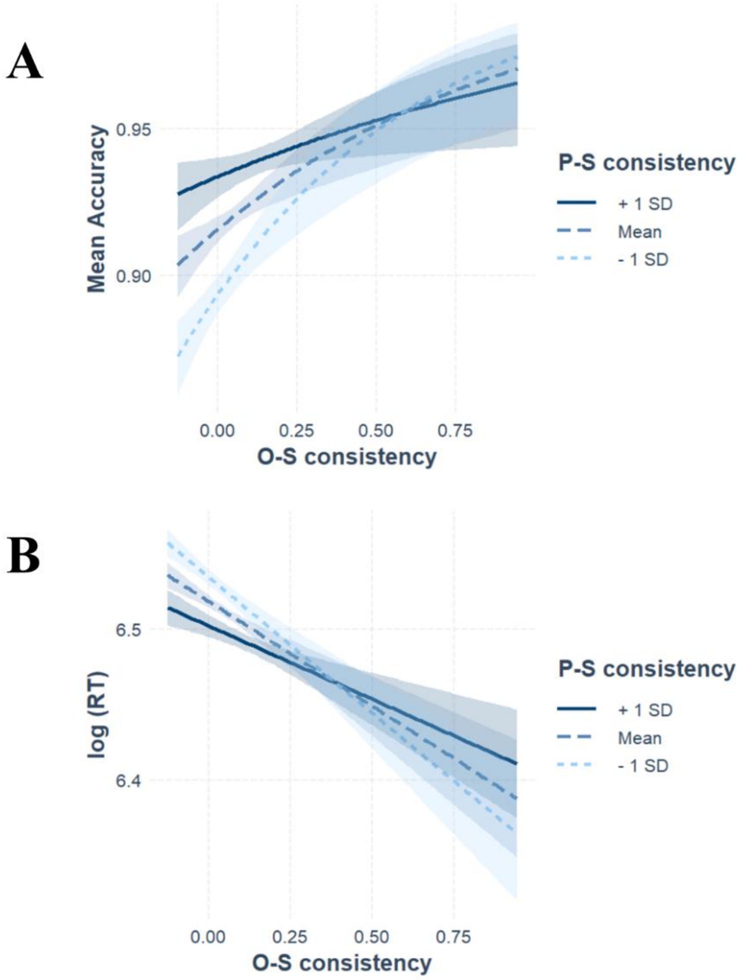
Model	DV	Predictor	β	SE	*t*	*p*	*Partial-R^2*
1	Acc.	OSC[a]	0.100	0.0267	3.759	< .001	0.3%
		O-P[a]	-0.106	0.0180	-5.902	< .001	0.7%
		PSC[a]	0.231	0.0333	6.948	< .001	0.1%
		Log Freq.	0.473	0.0100	46.873	< .001	37.0%
		Word Length	0.547	0.0219	25.027	< .001	13.3%
		Log. O Neigh Size	0.391	0.0396	9.887	< .001	2.2%
		Log. P Neigh Size	-0.012	0.0324	-0.376	.707	<0.1%
		OSC×O-P	-0.066	0.0207	-3.200	.001	0.3%
		PSC×O-P	-0.002	0.0226	-0.093	.926	<0.1%
		OSC×PSC	-0.004	0.0008	-5.626	<.001	0.4%
		OSC×O-P×PSC	0.002	0.0101	0.212	.832	<0.1%
		Articulation[b]					
2	Log-RT	OSC[a]	-0.012	0.0019	-6.230	< .001	0.7%
		O-P[a]	0.006	0.0015	3.851	< .001	0.2%
		PSC[a]	-0.014	0.0023	-6.022	< .001	0.7%
		Log Freq.	-0.034	0.0007	-46.941	< .001	29.2%
		Word Length	-0.011	0.0018	-6.068	< .001	0.6%
		Log. O Neigh Size	-0.020	0.0036	-5.483	< .001	0.6%
		Log. P Neigh Size	-0.011	0.0029	-3.851	< .001	0.3%
		OSC×O-P	0.001	0.0018	0.607	.544	<0.1%
		PSC×O-P	0.002	0.0016	0.992	.321	<0.1%
		OSC×PSC	0.004	0.0006	6.072	<.001	0.7%
		OSC×O-P×PSC	-0.0006	0.0008	-0.776	.438	<0.1%
		Articulation[b]					

Note: Models fitted to data including *k*=5355 monosyllabic words. Acc. = mean accuracy; DV = dependent variable; log-RT = log-transformed response time; Log Freq. = log-transformed frequency; OSC = O-S consistency (modified measure of OSC); O-P = measure of O-P regularities (surprisal); PSC = P-S consistency; Log. O Neigh Size = log-transformed orthographic neighborhood size; Log. P Neigh Size = log-transformed phonological neighborhood size. Models with mean accuracy as a DV use a logit link with a quasi-binomial family. Partial-R^2 values were computed using the rsq package in R (Zhang, 2021).

[a] OSC, O-P, and PSC were scaled and centered to reduce collinearity with the estimated interactions.

[b] Articulation parameters do not have estimates, SE, and t/p-values because these are a set of 15 dummy-coded variables (reflecting manner/place of articulation of first consonant, see also Siegelman, Kearns, et al., 2020).

Figure 2. *Estimated Effect of the Interaction Between O-S and P-S Consistency on Lexical Decision Accuracy (Top Panel, A) and Response Latencies (Bottom Panel, B). Figure plotted using the interactions package in R* (Long, 2019). Shaded areas show 95% confidence intervals.



For word naming RT and accuracy, we found significant effects of O-P surprisal and PSC (Table 3): Words that were more consistent in terms of P-S (higher PSC) and O-P (lower O-P surprisal) had faster and more accurate naming responses. OSC, however, did not have a significant main-effect in predicting naming accuracy or latencies – in contrast to the results of the lexical decision models above. Note that the naming RT analyses did reveal a significant OSC by PSC interaction (again, a sub-additive interaction mirroring that in lexical decision), pointing at OSC effects specifically in P-S inconsistent words. We return to discussing this finding below.

Table 3*: Effect of the Modified OSC Measure, O-P and PSC on Word Naming RT and Accuracy in ELP Data.*

Model	DV	Predictor	β	SE	*t*	*p*	*Partial-R^2*
1	Acc.	OSC[a]	-0.044	0.0270	-1.618	.101	0.1%
		O-P[a]	-0.245	0.0174	-14.096	<.001	4.1%
		PSC[a]	0.125	0.0365	3.414	<.001	0.3%
		Log Freq.	0.028	0.0110	25.788	<.001	14.1%
		Word Length	0.186	0.0253	7.326	<.001	1.6%
		Log. O Neigh Size	0.313	0.0466	6.715	<.001	1.0%
		Log. P Neigh Size	-0.087	0.0379	-2.294	.022	1.3%
		OSC×O-P	-0.024	0.0187	-1.265	.206	<0.1%
		PSC×O-P	0.019	0.0208	0.895	.371	<0.1%
		OSC×PSC	-0.009	0.0073	-1.166	.244	<0.1%
		OSC×O-P×PSC	-0.013	0.0094	-1.393	.164	<0.1%
		Articulation[b]					
2	Log-RT	OSC[a]	-0.002	0.0017	-1.021	.308	<0.1%
		O-P[a]	0.013	0.0013	9.482	<.001	1.6%
		PSC[a]	-0.006	0.0020	-3.059	.002	0.1%
		Log Freq.	-0.014	0.0006	-21.407	<.001	7.7%
		Word Length	0.010	0.0016	6.324	<.001	0.7%
		Log. O Neigh Size	-0.027	0.0032	-8.632	<.001	1.3%
		Log. P Neigh Size	0.012	0.0026	4.796	<.001	0.4%
		OSC×O-P	0.003	0.0016	1.666	.096	<0.1%
		PSC×O-P	-0.001	0.0015	-0.393	.694	<0.1%
		OSC×PSC	0.001	0.0005	2.602	.009	0.1%
		OSC×O-P×PSC	-0.001	0.0007	-1.471	.141	<0.1%
		Articulation[b]					

Note: Models fitted to data including *k*=5541 monosyllabic words. Acc. = mean accuracy; DV = dependent variable; Log-RT = log-transformed response time; Log Freq. = log-transformed frequency; OSC = O-S consistency (modified measure of OSC); O-P = measure of O-P regularities (surprisal); PSC = P-S consistency; Log. O Neigh Size = log-transformed orthographic neighborhood size; Log. P Neigh Size = log-transformed phonological neighborhood size. Models with mean accuracy as a DV use a logit link with a quasi-binomial family. Partial-R^2 values were computed using the rsq package in R (Zhang, 2021).

[a] OSC, O-P, and PSC were scaled and centered to reduce collinearity with the estimated interactions.

[b] Articulation parameters do not have estimates, SE, and t/p-values because these are a set of 15 dummy-coded variables (reflecting manner/place of articulation of first consonant, see also Siegelman, Kearns, et al., 2020).

3.3 Impact on Eye Movements.

Here we examined whether the effects of OSC can be generalized to text reading for comprehension. To do so we used data from the GECO book reading corpus (Cop et al., 2017), which includes eye-tracking data recorded as participants read a full book (*The Mysterious Affair at Styles* by Agatha Christie) for comprehension. The corpus includes data from two groups of participants: A Dutch-English bilingual group and an English monolingual group; here we only examine data from the latter group of n=14 L1 English college-level readers. GECO includes various word-level eye-tracking measures (e.g., first fixation duration, total reading time, skipping, etc.) – for brevity we focus only on words' gaze duration (i.e., sum of fixation durations per word before moving away from it). The full database includes about 774,000 data points (i.e., word occurrences). Because our analysis focuses on fixation times, we removed all data points where no fixation was made (i.e., skips), which amounted to 38.9% of trials (~301,000 data points). We also excluded words for which we did not have estimates of OSC and PSC. That is, we only included words that appear in the ELP, which exist in the LSA semantic space, and have at least one orthographic and one phonological neighbor in this space, leading to a loss of additional ~87,000 trials. Lastly, we excluded outliers where gaze duration was smaller than 80msec or in the top percentile across words and subjects. Overall, the results reported below are based on data including ~360,000 data points (a full breakdown of the number of word types and tokens per subject is available through the code in the project's OSF page).

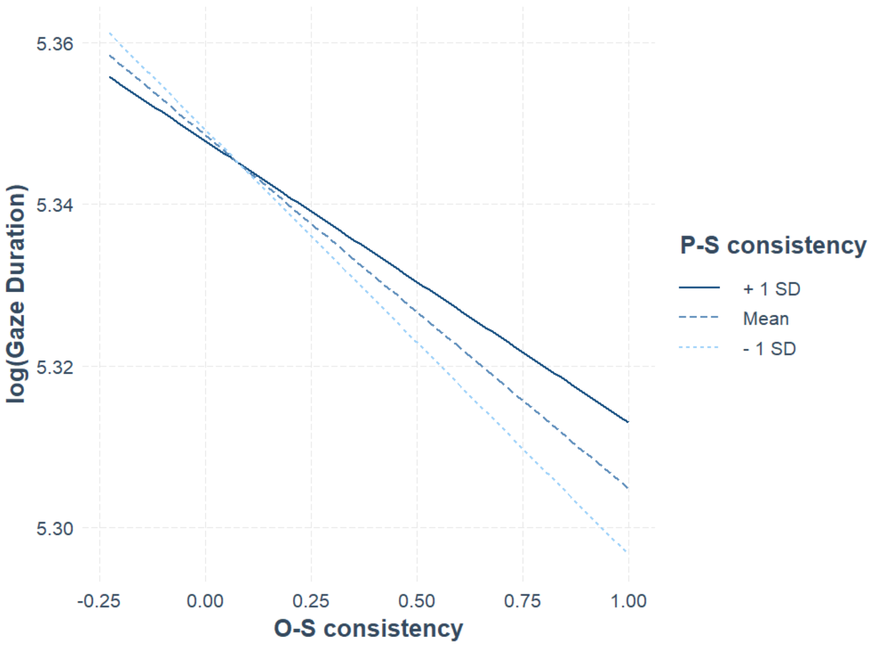
Our analysis of the eye-tracking data was conducted using a linear mixed-effect model, where we predicted log-transformed gaze duration from OSC, PSC, and their interaction as independent variables of interest. We additionally controlled for log-transformed frequency, orthographic word length, orthographic and phonological log-transformed neighborhood, and trial number. Models also included by-subject and by-word random intercepts (the maximal random effect structure that converged; Barr et al., 2013). Models were fitted using the R package *lme4* (Bates et al., 2015), with *p*-value approximation using *lmerTest* (Kuznetsova et al., 2017). The results are presented in Table 4. As can be seen, there was a significant main effect of OSC on reading times, with shorter gaze durations for words with higher values of OSC. We did not find a main effect of PSC, but did observe a significant sub-additive interaction between OSC and PSC, such that the effect of OSC was increasingly more pronounced in words with that are less P-S consistent, mirroring the interactions observed in the word recognition analysis above (Figure 3).

Table 4*: Effect of the Modified OSC Measure and PSC on Log-transformed Word Reading Times in the GECO Eye-movement Database.*

Predictor	β	SE	*t*	*p*	*Partial R^2*
OSC	-0.005	0.0019	-2.707	.007	0.4%
PSC	~0.000	0.0026	0.004	.997	<0.1%
Log Freq.	-0.030	0.0021	-13.997	<.001	11.2%
Word Length	0.049	0.0025	19.768	<.001	16.1%
Trial Number	0.003	0.0006	5.342	<.001	0.1%
Log. O Neigh Size	0.009	0.0028	3.138	.002	0.6%
Log. P Neigh Size	-0.002	0.0027	-0.772	.441	<0.1%
OSC×PSC	-0.001	0.0005	2.012	.044	0.2%

Note: Log-RT = Log-transformed response time; Log Freq. = Log-transformed frequency; OSC = O-S consistency (modified measure of OSC); PSC = P-S consistency; Log. O Neigh Size = Log-transformed orthographic neighborhood size; Log. P Neigh Size = Log-transformed phonological neighborhood size.
All predictors were scaled and centered to reduce collinearity with the estimated interaction and to improve model convergence. Effect sizes were estimated using Satterthwaite approximation of degrees of freedom and the *t_to_eta2* function in the R package *effectsize* (Ben-Shachar et al., 2020).

Figure 3. *Estimated Effect of the Interaction Between O-S and P-S Consistency on Log-transformed Gaze Duration in the GECO Eye-movement Database*. *Figure plotted using the interactions package in R* (Long, 2019).



3.4 Group-level Effects of OSC: Summary and Discussion

Our analysis of group-level word recognition data showed that indeed OSC affects item-level variability in single word recognition, in line with earlier studies using previous operational definitions (Amenta et al., 2017; Marelli et al., 2015; Marelli & Amenta, 2018). Importantly, our analysis also showed that the effects of OSC are impacted by task demands: We found significant main effects of OSC in lexical decision (on both accuracy and response latencies), but not word naming. The divergent findings in the two word recognition tasks are theoretically plausible and consistent with other findings that lexical decision and naming are differentially sensitive to variables associated with semantic and phonological variables (e.g., Balota et al., 2007; Ferrand et al., 2011). Specifically, there is behavioral and computational evidence that semantic properties of words have a greater impact in lexical decision tasks, suggesting that lexical decision involves more direct O-S processes, and our results regarding the main-effects of OSC are consistent with this conclusion.

Our results also revealed several effects of higher-order interactions between O-S, O-P and P-S regularities on word recognition, including significant OSC by PSC interactions in both accuracy and RT in lexical decision, and in naming RT. We reiterate that this OSC by PSC interaction was sub-additive, with larger effects of OSC in words that are more P-S inconsistent. This interaction is in line with the Triangle Model of Reading: Proficient readers are expected to have an efficient division of labor between O-S and O-P processes, and are therefore expected to rely more on one source of information when the other is unavailable (Harm & Seidenberg, 2004; Strain et al., 1995).

Returning to the general effects of OSC on reading, one of the key findings in our group-level analysis was that OSC effects were not limited to single word recognition tasks: Namely, we observed a significant effect of OSC on reading times measured via eye movements during connected text reading. This suggests that readers are sensitive to O-S regularities also when reading in context for comprehension. Also note that in eye movements we again observed an OSC by PSC sub-additive interaction, similar to the one found in the word recognition data and in line with a division-of-labor view. To the best of our knowledge, this is the first demonstration of effects of O-S regularities (and their interaction with P-S) beyond single word recognition. When considered together with the word recognition results, these results suggest that skilled adult readers show general reliance on O-S regularities in tasks that involve a more direct mapping of orthography to semantics (i.e., lexical decision and reading for comprehension), and that the impact of OSC is further modulated by the extent of PSC. We return to these points in the General Discussion.

4. Individual Differences in Reliance on O-S Regularities During Word Naming Among Early Readers

So far, our investigation has focused on item-level effects, showing that O-S regularities indeed impact the reading outcomes of the average English skilled adult reader. Yet the promise of statistical theories of reading goes beyond accounting for the behavior of the average reader. Rather, such approaches have implications also for understanding individual differences in reading skills. Computational models suggest that given individual differences in learning abilities and/or learning experiences, readers are expected to differently rely on the regularities that are available to them in the written input, including in particular those between orthography, phonology and semantics (see, e.g., Plaut, McClelland, Seidenberg, & Patterson, 1996; Rueckl, 2016; Rueckl, Zevin, & Wolf VII, 2019; Woollams, Ralph, Plaut, & Patterson, 2007; Zevin & Seidenberg, 2006). This computational work gives rise to two predictions: (1) not all individuals will display similar reliance on O-P and O-S pathways during reading, and (2) these individual differences will account for reading skills across different developmental stages.

Behavioral studies on individual differences in reliance on O-P and O-S pathways provide some support for these claims. Thus, for example, in our recent study we investigated the word naming behavior of 399 primary school-aged children (Siegelman, Rueckl, et al., 2020). We focused on the individual-level impact of two factors on word naming accuracy: (1) O-P regularities - which in English (and in other alphabetic languages) provides a generally valid cue to access a word's phonology from its orthographic form, and (2) imageability (i.e., the ease of eliciting a mental image), a semantic property that has been used to index the degree of semantic involvement in word naming (see also Woollams et al., 2016). Concretely, we used a word reading task with a manipulation of items' O-P consistency and imageability (modeled after the group-level work by Strain, Patterson, & Seidenberg, 1995 and adapted for children). Then, to extract individual-level measures of reliance on these two sources of information, we ran logistic models on the data of each reader predicting item-level naming accuracy from the two predictors. These models resulted in two slope scores for each child, one reflecting extent of reliance on O-P regularities, and the other reflecting reliance on imageability. The results showed that individuals who were more greatly influenced by O-P regularities during the word naming task, and less influenced by imageability, had better reading skills as reflected in standardized reading test scores (see Pugh et al., 2008; Strain & Herdman, 1999; Woollams et al., 2016 for earlier related findings with adults). These findings were interpreted as suggesting that better early readers are those who rely more greatly on efficient sources of information (e.g., O-P regularities) and less on arbitrary cues (e.g., words' imageability).

Yet note that our prior study and others on adult populations (e.g., Woollams et al., 2016; but see Ulicheva et al., 2020) examined individual differences in the extent of semantic involvement during word naming by focusing on reliance on imageability. In the present study, we extend this analysis to focus instead on individual differences in reliance on the *actual regularities between orthography and semantics*, as reflected by the effect of OSC on individuals' naming accuracy. This enables us to examine two questions: (1) whether early readers differ from one another in their sensitivity to O-S regularities, and (2) what characterizes readers who display greater sensitivity to O-S regularities (i.e., is reliance on actual O-S regularities associated with better or poorer reading skills?). Our predictions are that (1) there will be individual differences in the impact of OSC, and that (2) in contrast to reliance on imageability, reliance on the actual regularities between orthography and semantics would be positively correlated with reading skills. This is because O-S regularities can serve as an efficient cue for the mapping of an orthographic input into semantics (and phonology, through indirect O-S(-P) computations), in contrast to imageability which is an inefficient semantic cue that does not reflect the O-S regularities in the writing system (i.e., does not provide a reliable way to map orthographic units into semantics; and see General

Discussion below, for a discussion of how such result is plausible and indeed observed despite the lack of OSC main effect in adults' naming in the ELP analysis above).

4.1 Individual-Differences Analysis: Methods

This section is based on a re-analysis of the word naming data (and measures of reading skills) from Siegelman, Rueckl, et al., (2020). The sample includes 399 children in the second-fifth grade from two large-scale studies. The first study includes 121 third and fourth graders from public and charter schools in a large urban community in the U.S, and comprises mostly children with reading disabilities (see Siegelman, Rueckl, et al., 2020, for details). The second study includes 278 children in the second-fifth grades from private and public schools, with a wider range of reading skills. For brevity and to maximize statistical power we report the results of an analysis including the aggregated data of both samples (see also aggregated analysis in Siegelman, Rueckl, et al., 2020).

Within a larger battery of behavioral and neural assessments, each of the children participated in a word naming task, including 160 trials presented in a fixed order. Participants were asked to read out loud each word as accurately and quickly as possible, and their responses were coded for accuracy by an experimenter who sat in the experiment room. Words were selected so they would be generally familiar to children in the second grade and up. Originally, words were chosen such that they would vary along three independent variables (following Strain et al., 1995): frequency, imageability, and O-P regularity. Log-transformed Frequency was estimated based on the Zeno corpus, grades 1-8 (Zeno et al., 1995); imageability was based on standard ratings (Paivio et al., 1968); and O-P regularity was operationalized as the surprisal of the vowel pronunciation (Siegelman, Kearns, et al., 2020). Here we also consider the impact of additional independent variables, including OSC, along with PSC and additional control variables (see below). Note that the final sample of

399 does not include 6 additional children who read more than 98% of the words correctly because such near-ceiling accuracy makes it difficult to extract reliable slope scores (see Siegelman, Rueckl, et al., 2020, for details and discussion). In addition to data from the word naming task, all children participated in three sub-tests of the Woodcock-Johnson III (Woodcock et al., 2001): Word Attack (pseudoword reading), Letter Word Identification, and Passage Comprehension. We used the raw scores on these three sub-tests as outcome measures of reading component skills.

Our re-analysis approach of the word naming task followed the strategy by Siegelman, Rueckl, et al. (2020): We ran a series of logistic regression models, each examining the impact of one predictor on word naming accuracy. That is, for each child, we ran a series of logistic models with word naming accuracy as the dependent variable, each including one of the following variables as independent variables: OSC, PSC, log-transformed orthographic neighborhood size, log-transformed phonological neighborhood size, O-P surprisal, and imageability. These analyses resulted in a series of individual-level slope scores for each child, reflecting her/his reliance on each of these factors during word naming. In analyses below we examine the links between these slope scores (reflecting reliance on each of these factors) and reading skills using correlations and linear models (see below). As mentioned above, the original study by Siegelman, Rueckl, et al. (2020) already included the slope measures reflecting reliance on O-P regularities and imageability[4]. Note that the slope scores reflecting reliance on O-P surprisal were "flipped" such that higher slope scores reflect stronger reliance on O-P regularities (in the same direction as the other slope scores). Our decision to extract measures of reliance on orthographic and phonological neighborhood size was due to the positive correlations between our OSC and PSC measures

[4] The task also has a manipulation of word's frequency, yet the analysis by Siegelman, Rueckl, and colleagues show that individual-level reliance on word frequency does not predict reading skills beyond reliance on O-P regularities and imageability at this sample. We therefore do not include frequency in the analysis reported in the main text. Follow-up analysis confirmed that the inclusion of slope scores for reliance on frequency does not alter the results qualitatively.

of interest and these metrics at the item-level (i.e., words with a smaller orthographic neighborhood tend to have higher OSC levels; and words with a smaller phonological neighborhood tend to have higher PSC values; see Table 1 above); we therefore used these slope scores as control variables in the analysis predicting reading skills below. We refer readers to the project's OSF page for a list of the 160 words used in the task along with their item-level properties.

4.2 Individual-Differences Analysis: Results

The distribution of sensitivity to O-S regularities across subjects (i.e., slope scores reflecting the impact of OSC on naming accuracy) is shown in Figure 4. On average, participants' accuracy in the word naming task was not significantly influenced by O-S regularities (mean OSC slope score: -0.02, SD = 0.28: $p > 0.05$)[5]. Nevertheless, the substantial individual differences in the impact of OSC on naming accuracy raises the question of whether the variability around the mean is meaningful and whether it is predictive of reading skills (see a related discussion in Siegelman, Rueckl, et al., 2020 regarding imageabillty effects, which were not significant at the group-level in Study 2 but reflected meaningful individual differences). Table 5 shows the correlations between the various individual-level slope scores, as well as between the slope scores and reading outcomes (i.e., Woodcock-Johnson scores). It shows that the pairwise correlations between reliance on O-S regularities and reading outcomes were positive and significant, while the correlations between reliance on P-S regularities and the same outcomes were negative. Yet these raw correlations with reading component skills should be interpreted with caution, as the correlation table also shows that there were considerable correlations between the various

[5] The null group-level effect of OSC on naming accuracy in this sample was corroborated by a logistic mixed-effect model predicting accuracy from OSC as well as frequency, O-P surprisal, imageability, P-S regularities, and orthographic and phonological log-transformed neighborhood size (with by-subject and by-item random intercepts): Estimated effect of OSC was insignificant, B=-0.08, SE=0.10, Z=-0.75, p=0.45.

predictors: Among other correlations, individuals whose naming accuracy was more strongly impacted by OSC showed smaller effects of orthographic neighborhood size, and the same was true for the impact of PSC and phonological neighborhood size (a result which is not surprising given the negative correlation across the *items* in the task between OSC/PSC and neighborhood size). Also note the various significant correlations of reliance on O-S (and P-S) with reliance on O-P and imageability (e.g., the positive correlation between O-S and O-P, $r = 0.23$, $p < .001$). The key question therefore is whether the relation between reliance on O-S (and/or P-S) regularities and reading skills holds also when controlling for all other slope scores.

Figure 4. *Individual Differences in Reliance on O-S: Distribution of OSC Slope Scores (N=399).*

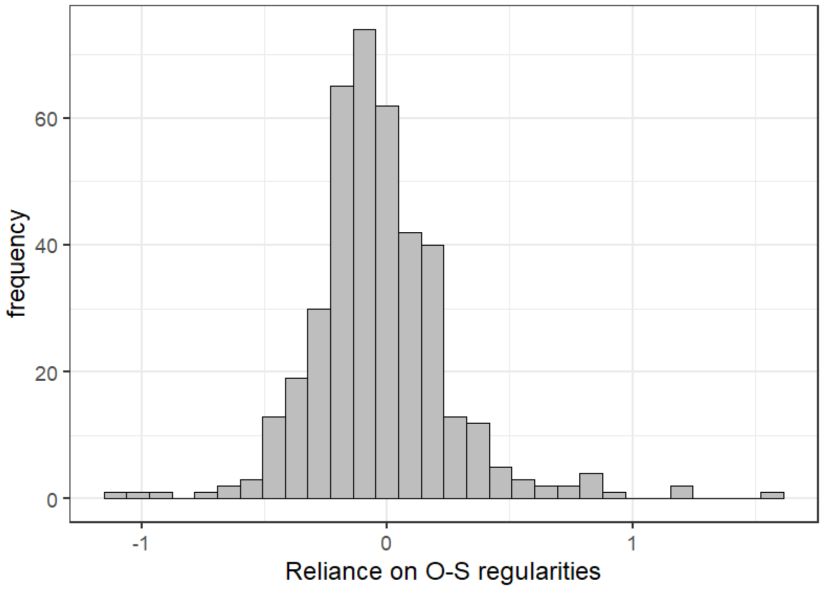


__Table 5__. Correlations Between Slope Scores Extracted from the Word Naming Task.

	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1) O-P	**-.15**	**.23**	**-.50**	-.01	**.46**	**.63**	**.63**	**.53**
2) Imageability		.01	**.17**	.06	.01	**-.33**	**-.26**	**-.28**
3) OSC			**.16**	**-.60**	**-.14**	**.23**	**.31**	**.31**
4) PSC				**-.25**	**-.65**	**-.29**	**-.34**	**-.27**
5) O neigh size					**.52**	**-.14**	**-.21**	**-.17**
6) P neigh size						**.26**	**.31**	**.27**
7) Word Attack							**.78**	**.69**
8) Letter-Word ID								**.79**
9) Passage Comp								

Notes: O-P: O-P regularity (surprisal); OSC: O-S consistency; PSC: P-S consistency; O neigh size: log-transformed orthographic neighborhood size; P neigh size: log-transformed phonological neighborhood size. Significant correlations ($p < .05$; N=399) are in bold.

To examine this question, we next ran three regression models. Each of these models had one component from the Woodcock-Johnson as a dependent variable (i.e., Word Attack, Letter-Word Identification, and Passage Comprehension). All models included as central independent variables all slope scores calculated from the word naming task (i.e., O-P, OSC, PSC, imageability, orthographic neighborhood size, and phonological neighborhood size). We also included in the same models the interaction between reliance on O-P and O-S regularities. The rationale for including this interaction is that in a system with two pathways from a word orthography to phonology - direct O-P and indirect O-S(-P) - the strength of the association between skill and an individual's reliance on the regularities in one pathway may depend on their reliance on the regularities in the other. The results of these models are presented in Table 6.

In Word Attack and Letter-Word Identification there were no significant effects of reliance on O-S or P-S regularities beyond the previously reported contributions of O-P regularities and imageability and the impact of the control variables (nor there was a significant O-P by O-S interaction). However, in the model predicting Passage Comprehension scores we did observe a significant effect of sensitivity to O-S regularities (i.e., impact of OSC on word naming): Individuals who were more greatly impacted by OSC during the word naming task had better passage comprehension scores (see Figure 5 for visualization). We emphasize that this was a positive association: In the same direction as the association between reading skill and reliance on O-P regularities, and in an opposite direction from the outcome's association with reliance on imageability (an issue we return to in the General Discussion). Also note that the impact of PSC did not have significant effects on reading skills in these models when controlling for other predictors.[6] In addition to this main effect, there was also a significant effect of the interaction between reliance on O-P and O-S regularities on Passage Comprehension scores: Thus, the association between the impact of OSC and comprehension skills was stronger among participants who showed lesser reliance on O-P regularities (Figure 6).

[6] An interesting finding is that in all models better reading skill was associated with greater impact of phonological neighborhood size and lesser impact of orthographic neighborhood size. While unexpected, this result is line with the general notion that better early readers are those who more greatly utilize the phonological code during reading, here reflected in the magnitude of their phonological neighborhood size effect. It also suggests that poorer readers show relatively greater impact by the same orthographic information.

Table 6*: Regression Models Predicting Reading Skill from Reliance on Imageabiliity, O-P, O-S, and P-S Regularities.*

Predictor	β (coefficient)	*SE*	*t* value	*p*-value	*Partial-R^2*
Dependent variable: Word attack (adj-R^2=46.7%)					
IMG	-1.65	0.25	-6.69	**<.001**	10.3%
O-P	3.70	0.31	12.00	**<.001**	26.9%
OSC	0.10	0.33	0.29	.753	<0.1%
PSC	0.62	0.35	1.77	.077	0.8%
O neigh size	-1.05	0.37	-2.86	**.004**	2.0%
P neigh size	0.95	0.39	2.41	**.015**	1.5%
O-P×OSC	0.02	0.20	0.09	.928	<0.1%
Dependent variable: Letter-word identification (adj-R^2=50.3%)					
IMG	-1.43	0.29	-4.90	**<.001**	5.8%
O-P	3.81	0.36	10.47	**<.001**	21.9%
OSC	0.59	0.39	1.51	.132	0.6%
PSC	0.18	0.41	0.45	.654	<0.1%
O neigh size	-2.38	0.43	-5.48	**<.001**	7.1%
P neigh size	2.20	0.47	4.71	**<.001**	5.4%
O-P×OSC	-0.46	0.24	-1.90	.058	0.9%
Dependent variable: Passage Comprehension (adj-R^2=39.0%)					
IMG	-1.23	0.22	-5.60	**<.001**	7.4%
O-P	1.95	0.27	7.15	**<.001**	11.6%
OSC	0.97	0.29	3.30	**.001**	2.7%
PSC	0.22	0.31	0.70	.487	0.1%
O neigh size	-0.97	0.33	-2.97	**.003**	2.2%
P neigh size	1.38	0.35	3.94	**<.001**	3.8%
O-P×OSC	-0.49	0.18	-2.69	**.007**	1.8%

Notes: IMG: imageability; O-P: O-P regularity (surprisal); OSC: O-S consistency (modified measure of OSC); PSC: P-S consistency; O neigh size: log-transformed orthographic neighborhood size; P neigh size: log-transformed phonological neighborhood size; Predictors are centered and scaled.

Figure 5. *Variability Among Participants in Reliance on O-P (x-axis) and O-S Regularities (y-axis) and its Relation to Passage Comprehension Scores. Dashed Trend Line Shows the Correlation Between the Two Slope Scores. Color Scale Presents Scores on Passage Comprehension.*

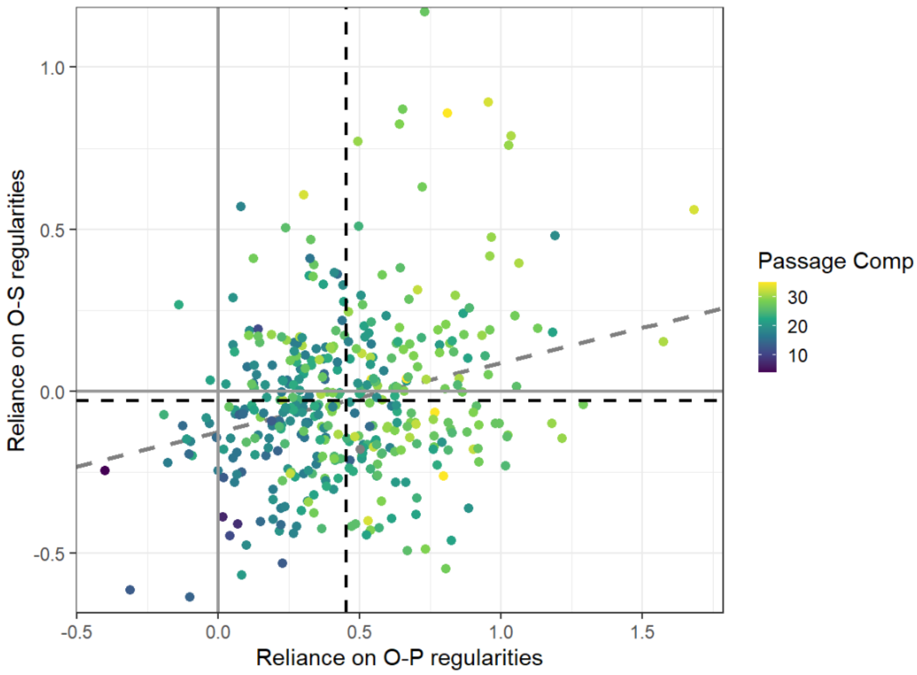
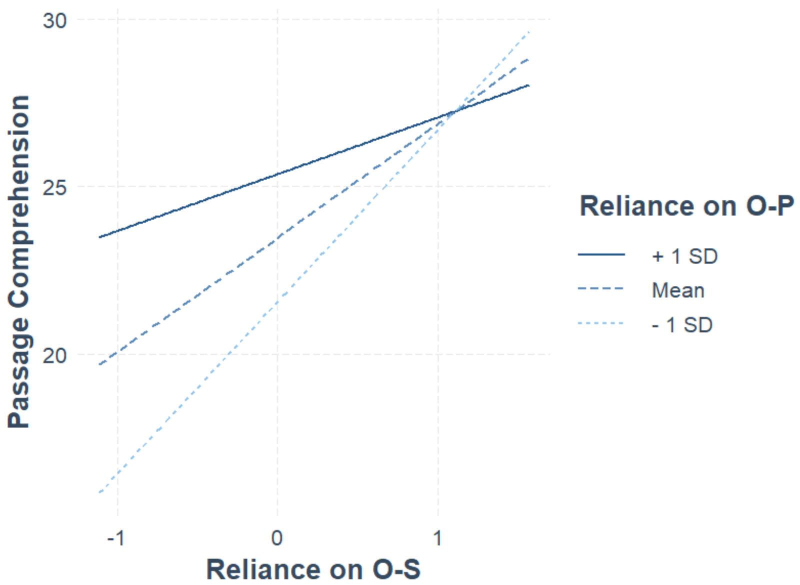


Figure 6. *Estimated Effect of the Interaction Between Reliance on O-P and O-S on Passage Comprehension Scores. Figure plotted using the interactions package in R (Long, 2019).*



5. General Discussion

The goal of the current paper is to advance our understanding of how to measure O-S regularities and how they impact group- and individual-level reading behavior. In terms of measurement, while we adopted the strategy offered by previous studies (Amenta et al., 2017; Marelli et al., 2015; Marelli & Amenta, 2018), we identified a potential confound in previous definitions which led us to offer a modified measure of OSC. Then, we used this modified measure to investigate group-level effects using available datasets of word recognition (lexical decision and word naming) and eye-movements during reading for comprehension. Results indicated that O-S regularities affect the reading behavior of adult readers of English, but that effects of OSC emerge more clearly (at least in the form of main effects) in tasks that place a greater emphasis on semantics compared to phonology (i.e., lexical decision and reading for comprehension; but not word naming). These findings are consistent with the idea that proficient readers rely more heavily on regularities that provide an efficient source of information for solving a particular task (Harm & Seidenberg, 2004): That is, proficient readers have knowledge of various types of regularities in their writing system, and exhibit differential sensitivity to them depending on the reading task's demands (e.g., by showing a general effect of OSC in tasks that involve greater semantic involvement such as lexical decision and reading for comprehension). Our results also reveal interactions between different regularities, including in particular OSC by PSC interactions consistently observed in the lexical decision task. As noted above, these interactions are consistent with the notion of an efficient division of labor in the Triangle Model of Reading: Proficient readers rely more on one source of information when the other is less available (Harm & Seidenberg, 2004). This observation is also in line with previous behavioral work on interactions between O-P and other aspects of O-S processes (interactions between O-P consistency and imageability, Strain et al., 1995).

In the last empirical section of the paper, we presented an investigation of individual differences in reliance on O-S regularities and their relation to emerging reading skills. This analysis revealed that early readers who show greater reliance on O-S regularities during a word naming task (i.e., greater OSC effect) have better comprehension skills. We wish to emphasize two important aspects of these findings.

The first is the difference between reliance on O-S regularities and reliance on imageability: We observed a *positive* association between reading skill and reliance on O-S regularities (in the same direction as reliance on O-P regularities), in contrast to the *negative* correlations between reading skill and reliance on imageability (Siegelman, Rueckl, et al., 2020, see also Table 6 above). The positive correlation between reliance on O-S regularities and skill aligns with a recent study showing that adults who are more sensitive to correspondences between orthographic cues and lexical categories (e.g., better at classifying a pseudoword ending with *–ful* as an adjective rather than a noun) show better performance in a variety of language tasks including reading, spelling, and author recognition (Ulicheva et al., 2020). We claim that the dissociation between reliance on O-S regularities versus reliance on imageability suggests that better early readers are those who are most able to capitalize on the statistical regularities in their writing system, and that these regularities include both O-P and O-S correspondences. By contrast, poor early readers rely less on these statistical properties, and instead rely more on imageability, a semantic property that does not provide a valid cue for "translating" the orthographic code into semantic or phonological representations (i.e., not an inherent property of the regularities between orthography and semantics). Most likely, poor readers' differential reliance on these various cues is a consequence of a combination of factors, including endogenous variables (e.g., individual differences in statistical learning skills, see Sawi & Rueckl, 2019; Siegelman et al., 2017) as well as experiential ones (poor readers' increased exposure to simpler texts that are more likely to involve more concrete topics, Kearns & Hiebert, 2021).

The second point we wish to emphasize is that the effect of individual differences in reliance on O-S regularities was specific to passage comprehension and was not observed when word or pseudoword reading served as the dependent variable. This stands in contrast to individual differences in reliance on O-P regularities and reliance on imageability, which were associated with all examined component reading skills (word and pseudoword reading as well as passage comprehension; Siegelman, Rueckl, et al., 2020 and Table 6 above). This difference may be due to the task demands of passage comprehension versus reading aloud of words or pseudowords: Thus, we speculate that knowledge of O-S regularities plays a more important role in skills such as comprehension that require more computation of semantics, and a lesser role in tasks involving simple mapping of orthographic forms to phonology (e.g., word or pseudoword naming).

This point also relates to a caveat in our individual-differences analysis: The fact that reliance on O-S was estimated through performance in a word naming task. Yet in our group-level analysis (i.e., ELP analysis), we found that adult skilled readers were not impacted by O-S regularities in word naming, but did show such effects in lexical decision and reading for comprehension. Why did we find then that children who were more impacted by O-S regularities during word naming were better comprehenders, despite the presumed non-optimality of O-S information in reading aloud? We believe that this pattern of results points to a dissociation between *reliance* and *knowledge* of a given code. Thus, we interpret our findings as showing that children who are better comprehenders are those who have greater knowledge of O-S regularities that they rely on during reading aloud. Later on in development, however, good readers may be those who – despite their knowledge of O-S regularities – will show less reliance on the same type of regularities in situations where they do not provide an efficient source of information. Instead, better older readers will use O-S regularities only in tasks where they are beneficial or when reading specific words that encourage the use of this type of information (e.g., P-S inconsistent words – see interactions above; or O-P inconsistent polymorphemic words – see Kearns et al., 2016). In other words, we claim that the good early readers in our developmental sample were characterized by knowledge of both O-P and O-S regularities. Yet, at the end of the developmental trajectory, better readers may be those who not only have knowledge of these different regularities, but also know when relying on each source of information is a beneficial strategy for the reading task at hand.

Beyond these specific points, an overarching question regarding our findings at large has to do with the relation between OSC measure(s) and morphological structure, and concretely, whether the impact of OSC on reading behavior goes beyond 'simple' morphological effects. As noted in the introduction, OSC is a cumulative measure of O-S regularities that is impacted, but not fully determined by, morphological structure. As such, it captures various forms of (ir)regularities in the O-S mapping (of both morphological and non-morphological nature) that typical morphological measures do not. Thus, OSC is impacted (among other factors) by: the relations between forms and meanings that are not expressed through overt affixation (e.g., the relation between *hotel, hostel,* and *motel*; and see Blasi et al., 2016 for evidence of non-arbitrary links between forms and meaning in the context of phonology-semantic associations); the *degree* of semantic (dis)similarity between a word and its neighbors – both morphologically-related and others (e.g., the fact that *baker* is more semantically related to *bake* than *stranger* is to *strange*; or the fact that *laughter* and *slaughter* are strongly negatively associated); and the *ratio* between the number of related and unrelated words in a target's neighborhood as well as the ratio of their degree of semantic similarity (e.g., the word *see* has some semantically-related neighbors such as *sees* and *seen*;

but it has a much larger number of unrelated neighbors such as *sea, bee,* and *seed*, which reduces its OSC value). In light of these differences, we expect OSC to capture reading outcomes beyond measures of morphological structure. Indeed, this claim is supported by an analysis reported in the Supplementary Materials S3 where we repeat the word recognition analysis while also controlling for the ratio of morphologically-related words from the size of a word's orthographic neighborhood. In a nutshell, the results show that although the ratio of morphologically-related words in a word's orthographic neighborhood is correlated with OSC, and is by itself a predictor of reading behavior, OSC still accounts for word recognition performance over and above this control variable. And although a full examination of all possible morphological measures is beyond the scope of this paper, we argue that this finding, combined with the conceptual differences between OSC and measures of morphological structure, suggest that the OSC's effects we revealed in this paper indeed go beyond morphological effects.

5.1 Methodological Considerations

While we do believe that the current work improves on previous investigations of O-S regularities and their impact on reading behavior, it admittedly involved a series of methodological decisions that went into our definition of OSC. We re-iterate our early point that our work should not be taken as claiming that our measure of OSC is the only way to quantify O-S regularities. Much like in the quantification of O-P regularities where multiple measures have been employed (e.g., regularity vs. consistency; type- vs. token-based quantification; vowel vs. body-rime measures; etc.), and appear to capture at least partially distinct aspects of O-P structure, our specific quantification does not cover all (potentially psychologically-relevant) aspects of the O-S mapping. In fact, we doubt whether there is a single measure that can capture the full O-S mapping, and this paper does not attempt to offer such an 'optimal' measure. The purpose of the discussion that follows is to highlight specific properties of our definition and discuss their potential impact on our findings, while providing pointers to further analyses (reported in the Supplementary Materials) and future directions that can help explore additional dimensions of the O-S mapping.

The first methodological decision has to do with the definition of neighborhood. We defined a word's neighborhood as all words with a Levenshtein distance of 1 (either orthographic or phonological distance). As noted above, this definition diverges from the one used in previous works by Marelli, Amenta, and colleagues, who defined the neighborhood as all the words that contain the target word. Note that our choice was driven primarily (but not only – see below) by practical considerations: By using this definition we were able to minimize the number of words with a non-defined OSC (i.e., words without any neighbors in the semantic vector-space). This applies especially to longer, morphologically complex words, that do not have neighbors per Marelli, Amenta, et al.'s definition (i.e., are not contained within other words). Indeed, out of the 40,481 words in the ELP, 57.7% had an undefined value for the OSC measure based on Marelli and Amenta's definition of neighborhood using our database (a roughly similar estimate to the proportion of words with an OSC=1 – that have only themselves as neighbors – in Marelli & Amenta, 2018). In contrast, 39.2% of words had an undefined O-S value when defining neighbors according to a Levenshtein distance of 1. Even among monosyllabic words, a substantial proportion of words had an undefined O-S value when using Marelli and Amenta's definition of neighborhood (e.g., from the 5,968 monosyllabic words in the ELP with O-P estimates in Siegelman, Kearns, et al. 2020, 32.3% had an undefined O-S estimate when using the previous definition of neighborhood, versus 8.5% using the Levenshtein-distance based one).

It is important to note, however, that the definition of neighborhood is likely to affect the types of relations that are captured by OSC estimates. Any definition of neighborhood incorporates some arbitrary boundary that divides words to those that are included in the OSC calculation and those that are not. Thus, whereas a definition based on a Levenshtein distance of 1 may be more sensitive to non-morphological orthographic similarity (i.e., grow – crow) and irregular morphological relations (e.g., *grow* – *grew*), a definition based on embedded words will capture morphological relations with affixes longer than one letter (e.g., *big* – *bigger*). Another consequence is that under Marelli and Amenta's definition shorter words often have a very large number of neighbors, with the vast majority of them seemingly weakly related to them both orthographically and semantically (e.g., the word *rat* is embedded in 1,874 other words in our database – including words such as *luc**rat**ive* and *refrige**rat**or*; in contrast, the same word has a neighborhood size of 38 under the Levenshtein-based definition[7]). Also, our definition of neighborhood is bidirectional (i.e., if word A is a neighbor of B then B is a neighbor of A), which is not the case in Marelli, Amenta, et al.'s definition. It is interesting to note that despite these differences, re-calculating OSC (and P-S consistency) with Marelli and Amenta's definition of neighborhood and our modified calculation (i.e., type-based calculation without including the word itself in its neighborhood) revealed strong correlations between the estimates based on the two types of neighbors: r = 0.737 (and r = 0.744 for P-S consistency). Still, the imperfect correlation suggests that as expected the definition of neighborhood affects the exact types of relations that are captured by the OSC measure, and future work is left with further examining the impact of neighborhood definitions on OSC estimates and their predictive value.

Another methodological choice that went into our analysis is the particular semantic vector space used to index semantic similarity. Here we used an LSA model – following the work by Marelli, Amenta and colleagues and in line with many others in the

[7] The original paper presenting the OSC measure (Marelli et al., 2015) avoided this issue by using a position-specific approach where neighbors are defined only as words that start with the target word (i.e., *rats* and *ratio* are neighbors of rat; but not *lucrative*).

cognitive/psycholinguistic literature. Yet LSA is a relatively old approach (originally described by Landauer & Dumais, 1997). In the years following its publication other Distributional Semantic Models (DSMs) were offered and applied in various domains in cognitive sciences (see Günther et al., 2019 for review). One of these models is GloVe (Global Vectors for Word Representation), which arguably outperforms LSA in capturing semantic similarity (Pennington et al., 2014). In the present context, this raises the question of whether our results are contingent on the use of LSA in the estimation of OSC or whether the results generalize also to more recent DSMs such as GloVE. To examine this, in a follow-up analysis we re-calculated OSC values based on pre-trained GLoVe vectors, compared these estimates to the ones using LSA, and re-examined the relations between OSC and word recognition accuracy and latencies: We report the results of this analysis in the Supplementary Materials S4. In a nutshell, this analysis showed that (1) there is a strong (albeit imperfect) correlation between OSC estimates using LSA and GloVe (r = 0.56), (2) our measure of OSC based on GloVe again predicts accuracy and response latencies in lexical decision (but not naming), although (3) some of the higher-order interactions between OSC, PSC, and O-P vary across the estimates using GLoVe and LSA, while others remain stable (specifically, the O-S by P-S interactions in lexical decision are replicated). This suggests that overall, our results are not contingent on the semantic vector-space used. We stress, however, that we do not have any claims regarding the superiority of one DSM or another in estimating O-S regularities – this should be the focus of directed future investigations. To facilitate such work, in the project's OSF page we provide OSC estimates for all words based on both LSA and GloVe.

Lastly, we stress two additional aspects of OSC that impact the types of associations it does and does not capture. The first is our use of a type-based calculation – discussed already in the Introduction. We again emphasize that our choice of a type-based calculation should not be taken as a claim against future uses of frequency-weighted measures – the rate of occurrence of semantically-related vs. unrelated neighbors may very well be a psychologically-relevant factor. Type- and token-based measures should both be explored (much like type- and token-based O-P measures that are both theoretically and empirically justifiable). Lastly, we stress that there are multiple forms of O-S and P-S regularities that OSC/PSC definitions do not capture (both the OSC measure used here and the one used by Marelli, Amenta, and colleagues). These include, for example, some irregular morphological regularities (e.g., *catch* and *caught*), phonostemes (e.g., *sneeze* and *snort*), and systematic links between grammatical categories and various phonological (and hence also, potentially, orthographic) cues (e.g., the fact that in English nouns have on average more syllables and a larger proportion of vowel units than verbs, Monaghan et al., 2007; see Dingemanse et al., 2015 for review). How to capture these regularities and whether and how they play a role in reading are far from trivial questions; at present, we simply highlight them to re-iterate that current operationalizations of OSC do not provide a full coverage of the rich O-S mapping.

5.2. Conclusions and Future Directions

The current paper highlights the importance of considering various types of regularities available in writing systems and the unique explanatory power each of these regularities have in accounting for group- and individual-level reading performance. Within this plethora of regularities, O-S associations have a presumably fundamental role in both literacy development and proficient reading (Harm & Seidenberg, 2004; Nation & Snowling, 1998; Seidenberg & McClelland, 1989), yet to date these regularities are still somewhat understudied empirically. The current paper, we believe, provides a step forward in understanding how to estimate O-S regularities and in unveiling their influence on reading behavior across items (i.e., inter-item variability), reading tasks, and individuals. At the theoretical level, our results demonstrate how readers show differential reliance on different regularities across reading situations given the computations required by the task at hand, and that better early readers are those who exhibit greater knowledge of the statistical regularities in their writing system – not only O-P information, but also O-S – along with lesser reliance on arbitrary cues (e.g., imageability). As such, our findings strengthen general notions of statistical learning views according to which reading involves extraction of regularities from the input, while further demonstrating how proficient reading requires finding a balance between multiple sources of information that are present in one's writing system.

Despite these insights, however, it is clear that many open questions remain regarding the full scope of the impact of O-S regularities on reading, and here we highlight two outstanding issues. The first has to do with the developmental trajectory of sensitivity to O-S regularities: Here we only provided a snapshot into behavior in either the beginning or the end of the reading acquisition trajectory, yet a crucial question is how readers gradually acquire knowledge of different types of regularities until they reach adult-level proficiency (or not, in cases of atypical development). Future studies are therefore left with exploring how O-S effects change over the course of reading acquisition, and their role in accounting for individual differences at different time points along this developmental trajectory (and see Davies et al., 2017 for relevant work in the context of O-P and imageability effects). A second outstanding question has to do with cross-linguistic differences: Our work focused on one writing system (English), which presents to its readers a particular mix of regularities. In other writing systems the reliability of different types of regularities varies (e.g., as a function of orthographic depth, Katz & Frost, 1992), and such variability is thought to impact the relative contribution of different types of regularities to reading (Seidenberg, 2011), including that of O-S regularities. We are hopeful that the current study will provide blueprints for future work exploring these important issues, towards a full understanding of how variability in reading - across individuals, tasks, developmental stages, and languages – is determined by the statistical information that writing systems encompass and the mechanisms that assimilate this information.

Acknowledgements

We thank Kevin S. Brown and Kenneth R. Pugh for helpful discussions. This work was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under award P20HD091013. The data included in the individual-differences analysis was collected under support by this grant as well as by award P01HD070837. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. N.S. received funding from the Israel Science Foundation (ISF), grant number 48/20.

Data availability statement

The project's OSF page includes data and analysis code used in this paper. It also includes a list with estimates of OSC (and PSC) for all words in the ELP, as well as a list of items and their properties used in the individual-differences analysis:

https://osf.io/3aczx/

References

Amenta, S., Crepaldi, D., & Marelli, M. (2020). Consistency measures individuate dissociating semantic modulations in priming paradigms: A new look on semantics in the processing of (complex) words. *Quarterly Journal of Experimental Psychology*, *73*(10), 1546–1563. https://doi.org/10.1177/1747021820927663

Amenta, S., Marelli, M., & Sulpizio, S. (2017). From sound to meaning: Phonology-to-Semantics mapping in visual word recognition. *Psychonomic Bulletin and Review*, *24*(3), 887–893. https://doi.org/10.3758/s13423-016-1152-0

Arciuli, J. (2018). Reading as Statistical Learning. *Language, Speech, and Hearing Services in Schools*, *49*(3), 634–643. https://doi.org/10.1044/2018_lshss-stlt1-17-0135

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The english lexicon project. *Behavior Research Methods*, *39*(3), 445–459. https://doi.org/10.3758/BF03193014

Baron, J., & Strawson, C. (1976). Use of orthographic and word-specific knowledge in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(3), 386–393. https://doi.org/10.1037/0096-1523.2.3.386

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). *lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-8*. http://cran.r-project.org/package=lme4

Ben-Shachar, M. S., Makowski, D., & Lüdecke, D. (2020). Compute and interpret indices of effect size. In *CRAN*.

Blasi, D. E., Wichmann, S., Hammarström, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound-meaning association biases evidenced across thousands of languages.

Proceedings of the National Academy of Sciences of the United States of America, *113*(39), 10818–10823. https://doi.org/10.1073/pnas.1605782113

Bouchet-Valat, M. (2019). *SnowballC: Snowball Stemmers Based on the C 'libstemmer' UTF-8 Library*. R package version 0.6.0. https://CRAN.R-project.org/package=SnowballC.

Chang, Y. N., & Monaghan, P. (2019). Quantity and Diversity of Preliteracy Language Exposure Both Affect Literacy Development: Evidence from a Computational Model of Reading. *Scientific Studies of Reading*, *23*(3), 235–253. https://doi.org/10.1080/10888438.2018.1529177

Chateau, D., & Jared, D. (2003). Spelling-sound consistency effects in disyllabic word naming. *Journal of Memory and Language*, *48*(2), 255–280. https://doi.org/10.1016/S0749-596X(02)00521-1

Chee, Q. W., Chow, K. J., Yap, M. J., & Goh, W. D. (2020). Consistency norms for 37,677 english words. *Behavior Research Methods, 52,* 2535–2555. https://doi.org/10.3758/s13428-020-01391-7

Cop, U., Dirix, N., Drieghe, D., & Duyck, W. (2017). Presenting GECO: An eyetracking corpus of monolingual and bilingual sentence reading. *Behavior Research Methods*, *49*, 602–615. https://doi.org/10.3758/s13428-016-0734-0

Davies, R. A. I., Birchenough, J. M. H., Arnell, R., Grimmond, D., & Houlson, S. (2017). Reading through the life span: Individual differences in psycholinguistic effects. *Journal of Experimental Psychology: Learning Memory and Cognition*, *43*(8), 1298–1338. https://doi.org/10.1037/xlm0000366

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, Iconicity, and Systematicity in Language. *Trends in Cognitive Sciences*, *19*(10), 603–615. https://doi.org/10.1016/j.tics.2015.07.013

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, *9*(2), 167–188. https://doi.org/10.1207/s1532799xssr0902_4

Feldman, L. B. (1994). Beyond orthography and phonology: Differences between inflections and derivations. *Journal of Memory and Language*, *33*(4), 442–470. https://doi.org/10.1006/jmla.1994.1021

Ferrand, L., Brysbaert, M., Keuleers, E., New, B., Bonin, P., Méot, A., Augustinova, M., & Pallier, C. (2011). Comparing word processing times in naming, lexical decision, and progressive demasking: Evidence from Chronolex. *Frontiers in Psychology*, *2*, 306. https://doi.org/10.3389/fpsyg.2011.00306

Fine, A. B., & Florian Jaeger, T. (2013). Evidence for Implicit Learning in Syntactic Comprehension. *Cognitive Science*, *37*(3), 578–591. https://doi.org/10.1111/cogs.12022

Fitt, S. (2001). *Unisyn Lexicon Release (Version 1.3) [Datafile and codebook]*. Centre for Speech Technology Research at the University of Edinburgh.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, *12*(6), 627–635. https://doi.org/10.1016/S0022-5371(73)80042-8

Frost, R. (2012). Towards a universal model of reading. *Behavioral and Brain Sciences*, *35*(5), 263–279. https://doi.org/10.1017/S0140525X11001841

Gingras, M., & Sénéchal, M. (2019). Evidence of Statistical Learning of Orthographic Representations in Grades 1–5: The Case of Silent Letters and Double Consonants in French. *Scientific Studies of Reading*, *23*(1), 37–48. https://doi.org/10.1080/10888438.2018.1482303

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(4), 674–691. https://doi.org/10.1037/0096-1523.5.4.674

Gonnerman, L. M., Seidenberg, M. S., & Andersen, E. S. (2007). Graded semantic and phonological similarity effects in priming: Evidence for a distributed connectionist approach to morphology. *Journal of Experimental Psychology: General, 136*(2), 323–345. https://doi.org/10.1037/0096-3445.136.2.323

Günther, F., Dudschig, C., & Kaup, B. (2014). LSAfun - An R package for computations based on Latent Semantic Analysis. *Behavior Research Methods*, *47*, 930–944. https://doi.org/10.3758/s13428-014-0529-0

Günther, F., Rinaldi, L., & Marelli, M. (2019). Vector-Space Models of Semantic Representation From a Cognitive Perspective: A Discussion of Common Misconceptions. *Perspectives on Psychological Science*, *14*(6), 1006 –1033. https://doi.org/10.1177/1745691619861372

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*, 662–720. https://doi.org/10.1037/0033-295X.111.3.662

Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language*, *29*(6), 687–715. https://doi.org/10.1016/0749-596X(90)90044-Z

Katz, L., & Frost, R. (1992). The Reading Process is Different for Different Orthographies: The Orthographic Depth Hypothesis. *Advances in Psychology*, *94*, 67–84. https://doi.org/10.1016/S0166-4115(08)62789-2

Kearns, D. M. (2020). Does English Have Useful Syllable Division Patterns? *Reading Research Quarterly*, *55*(S1), S145–S160. https://doi.org/10.1002/rrq.342

Kearns, D. M., & Hiebert, E. H. (2021). The Word Complexity of Primary-Level Texts: Differences Between First and Third Grade in Widely Used Curricula. *Reading Research Quarterly*. https://doi.org/10.1002/rrq.429

Kearns, D. M., Steacy, L. M., Compton, D. L., Gilbert, J. K., Goodwin, A. P., Cho, E., Lindstrom, E. R., & Collins, A. A. (2016). Modeling Polymorphemic Word Recognition: Exploring Differences Among Children With Early-Emerging and Late-Emerging Word Reading Difficulty. *Journal of Learning Disabilities*, *49*(4), 368–394. https://doi.org/10.1177/0022219414554229

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Landauer, T. K., & Dumais, S. T. (1997). A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. *Psychological Review*, *104*(2), 211–240. https://doi.org/10.1037/0033-295X.104.2.211

Long, J. A. (2019). *Interactions: Comprehensive, user-friendly toolkit for probing interactions (R package version 1.1.0)*. https://cran.r-project.org/package=interactions

Marelli, M., & Amenta, S. (2018). A database of orthography-semantics consistency (OSC) estimates for 15,017 English words. *Behavior Research Methods*, *50*(4), 1482–1495. https://doi.org/10.3758/s13428-018-1017-8

Marelli, M., Amenta, S., & Crepaldi, D. (2015). Semantic transparency in free stems: The effect of Orthography-Semantics Consistency on word recognition. *Quarterly Journal of Experimental Psychology*, *68*(8), 1571–1583. https://doi.org/10.1080/17470218.2014.959709

Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The phonological-distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology*, *55*, 259–305. https://doi.org/10.1016/j.cogpsych.2006.12.001

Murrell, G. A., & Morton, J. (1974). Word recognition and morphemic structure. *Journal of Experimental Psychology*, *102*(6), 963–968. https://doi.org/10.1037/h0036551

Nation, K., & Snowling, M. J. (1998). Semantic processing and the development of word-recognition skills: Evidence from children with reading comprehension difficulties. *Journal of Memory and Language*, *39*, 85–101. https://doi.org/10.1006/jmla.1998.2564

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, *76*(1), 1–25. https://doi.org/10.1037/h0025327

Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global vectors for word representation. *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*. https://doi.org/10.3115/v1/d14-1162

Plaut, D. C., & Gonnerman, L. M. (2000). Are non-semantic morphological effects icompatible with a distributed connectionist approach to lexical processing? *Language and Cognitive Processes*, *15*(4), 445–485. https://doi.org/10.1080/01690960050119661

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

Protopapas, A., & Vlahou, E. L. (2009). A comparative quantitative analysis of Greek orthographic transparency. *Behavior Research Methods*, *41*(4), 991–1008. https://doi.org/10.3758/BRM.41.4.991

Pugh, K. R., Frost, S. J., Sandak, R., Landi, N., Rueckl, J. G., Constable, R. T., Seidenberg, M. S., Fulbright, R. K., Katz, L., & Mencl, W. E. (2008). Effects of stimulus difficulty and repetition on printed word identification: An fMRI comparison of nonimpaired and reading-disabled adolescent cohorts. *Journal of Cognitive Neuroscience*, *20*(7), 1146–1160. https://doi.org/10.1162/jocn.2008.20079

Rastle, K., & Davis, M. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes*, *23*(7), 942–971. https://doi.org/10.1080/01690960802069730

Rastle, K., Davis, M. H., Marslen-Wilson, W. D., & Tyler, L. K. (2000). Morphological and semantic effects in visual word recognition: A time-course study. *Language and Cognitive Processes*, *15*(4), 507–537. https://doi.org/10.1080/01690960050119689

Rueckl, J. G. (2010). Connectionism and the role of morphology in visual word recognition. *The Mental Lexicon*, *5*(3), 371–400. https://doi.org/10.1075/ml.5.3.07rue

Rueckl, J. G. (2016). Toward a theory of variation in the organization of the word reading system. *Scientific Studies of Reading*, *20*(1), 86–97. https://doi.org/10.1080/10888438.2015.1103741

Rueckl, J. G., Mikolinski, M., Raveh, M., Miner, C. S., & Mars, F. (1997). Morphological priming, fragment completion, and connectionist networks. *Journal of Memory and Language*, *36*(3), 382–405. https://doi.org/10.1006/jmla.1996.2489

Rueckl, J. G., Zevin, J. D., & Wolf VII, H. (2019). Using computational techniques to model and better understand developmental word-reading disorders (i.e., dyslexia). In J. Washington, D. Compton, & P. McCardle (Eds.), *Dyslexia: Revisiting Etiology, Diagnosis, Treatment, and Policy*. Brookes Publishing Co.

Sawi, O. M., & Rueckl, J. (2019). Reading and the Neurocognitive Bases of Statistical Learning. *Scientific Studies of Reading*, *23*(1), 8–23. https://doi.org/10.1080/10888438.2018.1457681

Seidenberg, M. S. (1985). The time course of phonological code activation in two writing systems. *Cognition*, *19*(1), 1–30. https://doi.org/10.1016/0010-0277(85)90029-0

Seidenberg, M. S. (2011). Reading in different writing systems: One architecture, multiple solutions. In P. McCardle, B. Miller, J. R. Lee, & O. J. L. Tzeng (Eds.), *Dyslexia across languages: Orthography and the brain–gene–behavior link* (pp. 146–168). Paul H

Brookes Publishing.

Seidenberg, M. S., & Gonnerman, L. M. (2000). Explaining derivational morphology as the convergence of codes. *Trends in Cognitive Sciences*, *4*(9), 353–361. https://doi.org/10.1016/S1364-6613(00)01515-1

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

Sénéchal, M., Gingras, M., & L'Heureux, L. (2016). Modeling Spelling Acquisition: The Effect of Orthographic Regularities on Silent-Letter Representations. *Scientific Studies of Reading*, *20*(2), 155–162. https://doi.org/10.1080/10888438.2015.1098650

Share, D. L. (1999). Phonological Recoding and Orthographic Learning: A Direct Test of the Self-Teaching Hypothesis. *Journal of Experimental Child Psychology*, *72*(2), 95–129. https://doi.org/10.1006/jecp.1998.2481

Siegelman, N. (2021, November). Quantifying the regularities between orthography and semantics and their impact on group- and individual-level behavior. Retrieved from osf.io/3aczx

Siegelman, N., Bogaerts, L., Christiansen, M. H., & Frost, R. (2017). Towards a theory of individual differences in statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160059. https://doi.org/10.1098/rstb.2016.0059

Siegelman, N., Kearns, D. M., & Rueckl, J. G. (2020). Using information-theoretic measures to characterize the structure of the writing system: the case of orthographic-phonological regularities in English. *Behavior Research Methods*, *52*, 1292–1312. https://doi.org/10.3758/s13428-019-01317-y

Siegelman, N., Rueckl, J. G., Steacy, L. M., Frost, S. J., van den Bunt, M., Zevin, J. D., Seidenberg, M. S., Pugh, K. R., Compton, D. L., & Morris, R. D. (2020). Individual differences in learning the regularities between orthography, phonology and semantics predict early reading skills. *Journal of Memory and Language*, *114*, 104145. https://doi.org/10.1016/j.jml.2020.104145

Steacy, L. M., Compton, D. L., Petscher, Y., Elliott, J. D., Smith, K., Rueckl, J. G., Sawi, O., Frost, S. J., & Pugh, K. R. (2018). Development and prediction of context-dependent vowel pronunciation in elementary readers. *Scientific Studies of Reading*, *23*(1). https://doi.org/10.1080/10888438.2018.1466303

Strain, E., & Herdman, C. M. (1999). Imageability effects in word naming: An individual differences analysis. *Canadian Journal of Experimental Psychology*, *53*(4), 347–359. https://doi.org/10.1037/h0087322

Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic Effects in Single-Word Naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1140–1154. https://doi.org/10.1037/0278-7393.21.5.1140

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, *14*(6), 638–647. https://doi.org/10.1016/S0022-5371(75)80051-X

Treiman, R., & Kessler, B. (2006). Spelling as Statistical Learning: Using Consonantal Context to Spell Vowels. *Journal of Educational Psychology*, *98*(3), 642–652. https://doi.org/10.1037/0022-0663.98.3.642

Treiman, R., Kessler, B., Zevin, J. D., Bick, S., & Davis, M. (2006). Influence of consonantal context on the reading of vowels: Evidence from children. *Journal of Experimental Child Psychology*, *93*(1), 1–24. https://doi.org/10.1016/j.jecp.2005.06.008

Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The Special Role of Rimes in the Description, Use, and Acquisition of English Orthography. *Journal* *of Experimental Psychology: General*, *124*(2), 107–136. https://doi.org/10.1037/0096-3445.124.2.107

Ulicheva, A., Marelli, M., & Rastle, K. (2020). Sensitivity to meaningful regularities acquired through experience. *Morphology*, 1–22. https://doi.org/10.1007/s11525-020-09363-5

van der Loo, M. P. J. (2014). The stringdist package for approximate string matching. *R Journal*, *6*(1), 111–122. https://doi.org/10.32614/rj-2014-011

Weekes, B. S., Castles, A. E., & Davies, R. A. (2006). Effects of consistency and age of acquisition on reading and spelling among developing readers. *Reading and Writing*, *19*(2), 133–169. https://doi.org/10.1007/s11145-005-2032-6

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Riverside Publishing.

Woollams, A. M., Lambon Ralph, M. A., Madrid, G., & Patterson, K. E. (2016). Do you read how I read? Systematic individual differences in semantic reliance amongst normal readers. *Frontiers in Psychology*, *7*, 1757. https://doi.org/10.3389/fpsyg.2016.01757

Woollams, A. M., Ralph, M. A. L., Plaut, D. C., & Patterson, K. (2007). SD-squared: On the association between semantic dementia and surface dyslexia. *Psychological Review*, *114*(2), 316–339. https://doi.org/10.1037/0033-295X.114.2.316

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

Zeno, S., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide*. Touchstone Applied Science Associates.

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

Zhang, D. (2021). *rsq: R-Squared and Related Measures* (R package version 2.2). https://cran.r-project.org/package=rsq