Individual differences in learning the regularities between orthography, phonology and semantics predict early reading skills

Noam Siegelman[a,*], Jay G. Rueckl[a,b], Laura M. Steacy[c], Stephen J. Frost[a], Mark van den Bunt[a], Jason D. Zevin[a,d], Mark S. Seidenberg[e], Kenneth R. Pugh[a,b,f], Donald L. Compton[c], Robin D. Morris[g]

[a] Haskins Laboratories, USA
[b] University of Connecticut, USA
[c] Florida State University, USA
[d] University of Southern California, USA
[e] University of Wisconsin-Madison, USA
[f] Yale University, USA
[g] Georgia State University, USA

1986

ARTICLE INFO

ABSTRACT

Statistical views of literacy development maintain that proficient reading requires the assimilation of myriad statistical regularities present in the writing system. Indeed, previous studies have tied statistical learning (SL) abilities to reading skills, establishing the existence of a link between the two. However, some issues are currently left unanswered, including questions regarding the underlying bases for these associations as well as the types of statistical regularities actually assimilated by developing readers. Here we present an alternative approach to study the role of SL in literacy development, focusing on individual differences among beginning readers. Instead of using an artificial task to estimate SL abilities, our approach identifies individual differences in children's reliance on statistical regularities as reflected by actual reading behavior. We specifically focus on individuals' reliance on regularities in the mapping between print and speech versus associations between print and meaning in a word naming task. We present data from 399 children, showing that those whose oral naming performance is impacted more by print-speech regularities and less by associations between print and meaning have better reading skills. These findings suggest that a key route by which SL mechanisms impact developing reading abilities is via their role in the assimilation of sub-lexical regularities between printed and spoken language - and more generally, in detecting regularities that are more reliable than others. We discuss the implications of our findings to both SL and reading theories.

Introduction

One of the basic motivations to study Statistical Learning (SL) mechanisms is their presumed role across cognition. Thus, researchers study SL processes not just to understand the "SL mechanism(s)" in isolation, but rather as a way to gain new insights as to the processes involved in other cognitive domains. This premise is rooted in two research traditions. The first comprises works of pioneering computational modeling that demonstrate how incorporating statistical learning principles can account for complex linguistic behavior (Rumelhart & McClelland, 1986; Seidenberg & McClelland, 1989). The second are experimental demonstrations showing that SL computations can potentially account for multiple cognitive functions, including segmentation of continuous input (e.g. Aslin, Saffran, & Newport, 1998), categorization (e.g. Maye, Werker, & Gerken, 2002), and prediction of upcoming signal (e.g. Dale, Duran, & Morehead, 2012). Together, these two lines of research led to the view that SL plays a role in multiple cognitive domains, from low level perception (e.g. Barakat, Seitz, & Shams, 2013), to complex cognitive functions such as music appreciation (e.g. Salimpoor, Zald, Zatorre, Dagher, & McIntosh, 2015), face recognition (e.g. Dotsch, Hassin, & Todorov, 2017), and the acquisition and processing of spoken and written language (see Arciuli, 2018; Romberg & Saffran, 2010 for reviews).

In this vein, over the course of the past decades reading research has

---

become increasingly grounded in the idea that proficient reading requires the assimilation of statistical regularities present in the writing system. This statistical view of writing systems had an effect on virtually all sub-domains of reading research. As such, multiple studies document statistical regularities that are available to readers as they acquire and use their writing systems. These include associations between letters and speech sounds (Harm & Seidenberg, 2004; Seidenberg & McClelland, 1989) and between different orthographic units (including the correlations between different letters, e.g. Gingras & Sénéchal, 2019, and between different words, e.g. Fine & Florian Jaeger, 2013), as well as more complex relationships between orthographic units and morphological and semantic representations (Seidenberg & Gonnerman, 2000; Ulicheva, Harvey, Aronoff, & Rastle, 2018), and between orthographic units and stress patterns (Arciuli, 2018), among others (see Sawi & Rueckl, 2019 for review).

Indeed, studies from the reading literature propose that patterns of reading behavior can be best explained when considering this rich array of statistical regularities that are embedded in the written input (see, e.g., Frost, 2012; Sawi & Rueckl, 2019; Seidenberg, 2011 for reviews). This already suggests that SL - the mechanism(s) responsible for extracting regularities from the sensory input - plays an important role in the process of literacy acquisition and later on in guiding efficient reading. Attempts to further solidify this claim come from studies that use separate tasks to estimate individual's SL abilities (using for example the embedded triplet task or artificial grammar learning paradigms) and examine the relation between performance on these tasks and reading behavior. Thus, studies of individual differences have tied SL performance to variation in different aspects of reading skills both in first language (Arciuli & Simpson, 2012; Spencer, Kaschak, Jones, & Lonigan, 2014; Torkildsen, Arciuli, & Wie, 2019), and second language (Frost, Siegelman, Narkiss, & Afek, 2013). Still, the magnitude of the correlations between SL and reading varies across studies (see Arnon, 2019b; Siegelman, Bogaerts, Christiansen, & Frost, 2017), and recent studies suggest that correlations are observed only when using specific SL tasks and/or reading measures (Elleman, Steacy, & Compton, 2019; Qi, Sanchez Araujo, Georgan, Gabrieli, & Arciuli, 2019; Schmalz, Moll, Mulatti, & Schulte-Körne, 2019). Relatedly, studies of clinical populations show that individuals with reading disabilities are generally characterized by lower SL abilities (Gabay, Thiessen, & Holt, 2015; Kahta & Schiff, 2019; Sigurdardottir et al., 2017), although here again, there is some debate regarding the actual magnitude of these effects (see Schmalz, Altoè, & Mulatti, 2017).

Together, these studies provide an important proof of concept regarding the existence of some link between SL computations and reading skills. Importantly, however, questions remain not only in regards to why some studies fail to observe a relationship between SL performance and reading, but also as to the exact SL computations involved in reading and how variation in SL abilities gives rise to variability in literacy acquisition. These questions arise at least in part due to limited external validity of SL studies, which typically deal with learning of a single type of regularity over a short period of time (see Erickson & Thiessen, 2015; Frost, Armstrong, & Christiansen, 2019, for detailed discussion). In the context of reading, it is unclear whether such simplified form of learning can scale-up to account for the complex nature of the statistical knowledge that is required to account for reading behavior, which consists of multiple types of regularities, concurrently presented in various grain sizes and across modalities, and learned over extended periods of time.

In the present paper we report two studies that use an alternative approach to study the link between SL and reading, focusing on individual differences in beginning readers. Namely, instead of documenting individual differences in SL by using a simplified artificial task and then relating the individual differences observed on this task to some linguistic function (e.g. some component skill of reading), our approach focuses on identifying individual differences in children's reliance on statistical regularities as reflected directly in their word

naming behavior. We believe that this approach has the promise of informing SL research as well as the study of reading. On the SL side, it can help inform researchers about the subtle regularities that humans are able to assimilate "in the wild". In parallel, it has implications for understanding reading and reading acquisition by providing insights into the processes involved in typical (or impaired) reading acquisition. In other words, the current study demonstrates the theoretical power of a statistical approach to reading and serves to illustrate the directions that the SL literature should be extended to in order to address the issues of interest to reading researchers.

More concretely, in the experiments reported below we assessed how lexical and sub-lexical statistical regularities impact the word naming behavior of individual children. We then asked whether children whose word naming performance exhibits more sensitivity to the reliable statistical regularities are better readers overall (as reflected in standardized reading measures) compared to those who display lesser sensitivity to this statistical information. In parallel, we examined whether there is a relation between reading skill and individuals' reliance on more arbitrary associations that provide less efficient source of information for word naming. Given variability in SL abilities (e.g. Siegelman & Frost, 2015), and their presumed relation to reading abilities (e.g. Arciuli & Simpson, 2012; Frost et al., 2013), we hypothesized that there would be systematic variability in the degree to which individuals rely on different statistical regularities in word naming, and that those who show greater sensitivity to the reliable regularities, and lesser sensitivity to arbitrary associations, would display better reading skills.

Literacy acquisition and print-speech correspondences

As a starting point, we focus on statistical regularities that are particularly important for the development of skilled reading: the correspondences between letters and sounds. These regularities are considered to play a key role in learning the mappings between an orthographic string to the spoken form it represents (phonological decoding), which is widely considered to be one of the fundamental skills underlying proficient reading (e.g. Ehri, 2005; Share, 1999). This claim is supported by an extensive body of evidence, including correlations between phonological decoding and overall reading skills (e.g., Kearns, Rogers, Koriakin, & Al Ghanem, 2016; Perfetti, Beck, Bell, & Hughes, 1987; Shankweiler et al., 1999), impaired phonological decoding skills in populations with reading disabilities (e.g., Scarborough, 1998; Vellutino, Fletcher, Snowling, & Scanlon, 2004), and the impact of orthographic-phonological (O-P) regularities on word reading in skilled adult readers (e.g. Glushko, 1979; Jared, McRae, & Seidenberg, 1990).

Operationally, the strength of the association between an orthographic form and its phonological realization can be quantified by various measures (see Borleffs, Maassen, Lyytinen, & Zwarts, 2017; Siegelman, Kearns, & Rueckl, 2020, for reviews). A common measure is O-P consistency, which is a function of the number of words with a similar pronunciation of a given orthographic unit ('friends') and the number of words where the same orthographic unit is realized differently ('enemies'). Typically, this measure focuses on the body-rime level (*body-rime consistency*), and is defined as the ratio *friends / (friends + enemies)* of a word body (e.g. *ead* in the word head; e.g. Jared et al., 1990). Alternatively, a consistency measure at a smaller-grain size focuses on regularities at the grapheme-phoneme level (e.g. Chateau & Jared, 2003; Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995). Thus, the consistency of the grapheme *i* in the word *mint* (where $i \rightarrow /\textipa{I}/$) is a function of the number of words where *i* is similarly pronounced as $/\textipa{I}/$ (e.g., *bin*, *sing*, etc.) and the number of words where *i* is pronounced otherwise (e.g., *pint*). Indeed, a variety of findings make clear that skilled readers read faster and more accurately words with O-P mappings that are more "regular" or consistent at multiple grain-sizes, and that adults are particularly impacted by body-rime regularities (e.g. Cortese & Simpson, 2000; Jared, 2002).

Developmentally, sensitivity to O-P correspondences is gradually acquired over the course of typical reading acquisition (e.g., Sénéchal, Gingras, & L'Heureux, 2016; Weekes, Castles, & Davies, 2006), again with increased reliance on larger grain sizes later in development (Treiman & Kessler, 2006).

Additional sources of information for word reading

Importantly, sub-lexical O-P regularities are not the only source of information available to readers. One additional factor is word frequency - the rate of occurrence of an orthographic form. Although measures of word frequency typically ignore phonological and semantic ambiguity, to a first approximation they can be taken as 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. Thus, frequency captures some of the statistical structure in the mappings from orthography to phonology (O-P) and from orthography to semantics (O-S) (as well as the mapping between semantics and phonology, S-P) at the whole-word level. Other factors have to do specifically with the mapping between orthography to semantics. Generally, although O-S mapping is far less systematic than the mapping from orthography to phonology, it does embody statistical regularities and skilled readers assimilate and make use of these regularities. To a large extent, given the relatively arbitrary nature of the mapping, the O-S regularities that underlie word reading largely involve lexical-level correspondences between word forms and word meanings, as manifest in the effects of lexical ambiguity (e.g. Borowsky & Masson, 1996). Note, however, that because some words that are similar in form are also similar in meaning, the O-S mapping also embodies sub-lexical regularities (Marelli & Amenta, 2018; Monaghan, Chater, & Christiansen, 2005). This is perhaps most apparent in the ubiquitous effects of morphological structure on word reading (e.g. Grainger, Colé, & Segui, 1991; Rastle, Davis, Marslen-Wilson, & Tyler, 2000). From a statistical perspective, these effects arise because morphologically related words are typically related in both form and meaning and these form-meaning regularities are assimilated over the course of reading acquisition (Rueckl, 2010; Seidenberg & Gonnerman, 2000). Regardless of these important morphological regularities, however, the structure of the O-S mapping is generally more arbitrary than the mapping between orthography and phonology.

The impact of different sources of information related to O-P and O-S, as well as their interactions, are well demonstrated by the results of the word naming study by Strain, Patterson, and Seidenberg (1995). Strain et al. investigated the effect of imageability (the ease of eliciting a mental image, Paivio, Yuille, & Madigan, 1968), frequency, and O-P consistency on word naming in adult readers of English. Because imageability is a characteristic of word meaning, Strain et al. treated the imageability effect as a marker of the contribution of O-S processes. Strain et al. found that high-imageability words were named faster and more accurately than low-imageability words. Importantly, the magnitude of this effect varied with both word frequency and O-P consistency and was largest for low frequency, inconsistent words. These results illustrate how multiple statistical regularities interact to jointly determine behavior in a given task. In this particular case, the impact of imageability depends on the degree to which a word can be efficiently named on the basis of lexical and sub-lexical O-P regularities.

It should be noted that in contrast to word frequency and O-P consistency, which are regularities related to correspondences in the mappings between orthography, phonology, and semantics *per se*, imageability is related to statistical structure within the semantic domain. Specifically, imageability is thought to be related to degree of inter-correlation among semantic features (Harm & Seidenberg, 2004; Woollams, Lambon Ralph, Madrid, & Patterson, 2016). The degree of intercorrelation (or 'coherent covariation', McClelland & Rogers, 2003) among semantic features determines how strongly and/or quickly a semantic representation can be activated (Woollams et al., 2016).

Therefore, higher imageability of a word leads to more efficient computation of its meaning, and hence to stronger involvement of O-S(-P) processes.

From a developmental perspective, just as readers must learn to balance the influence of O-P regularities at different grain sizes, so too must they learn to make use of both O-S and O-P regularities. The question is then how do children eventually achieve the efficient division of labor between O-P and O-S processes exhibited by skilled readers. Computational models suggest that this developmental change is a consequence of how differences in the structure of the O-P and O-S mappings impact the statistical learning mechanism that attunes the reader to this structure (Harm & Seidenberg, 2004). Behaviorally, then, early readers are expected to exhibit a different weighing of the regularities in O-P and O-S mappings compared to adult readers in the same task.

Individual differences in the sensitivity to O-P and O-S processes

Even though there is an expected developmental trajectory at the group-level, not all individuals are predicted to follow it and similarly rely on the different regularities available to them in the input. As noted above, recent research using SL paradigms has revealed that there is substantial variability in SL ability. This suggests that individuals may differ in the degree to which they rely on different associations in a given task. That is, because not all individuals are similarly adept at learning the statistical regularities embedded in their environment, some may be better than others in finding the optimal balance in the use of different types of regularities (specifically O-P and O-S associations) for solving a particular reading task. In the context of a word naming task, which involves the mapping between orthographic and phonological word forms, we predict that some individuals will be better able to take advantage of the more systematic structure of the O-P mapping, whereas others will be more reliant on relatively arbitrary O-S associations. Furthermore, because individual differences in SL ability are correlated with differences in reading achievement (see above), we might further hypothesize that differences in the relative influence of O-P and O-S associations (i.e. *the division of labor*, Harm & Seidenberg, 2004) will be associated with differences in reading skill in general. Specifically, we hypothesize that readers who display greater sensitivity to O-P regularities in a naming task will have better reading skills than those who display less sensitivity to these regularities. In contrast, readers who rely more heavily on O-S processes are expected to have lower reading skills. This hypothesis is compatible with computational models that simulate reading behavior at the individual level, suggesting that it may explain observed variance in behavioral studies (Plaut, McClelland, Seidenberg, & Patterson, 1996; Rueckl, Zevin, & Wolf, 2019; Zevin & Seidenberg, 2006).

To re-iterate, our approach focuses on estimating individuals' reliance on different types of regularities in the context of a word naming task, where O-P regularities provide a more direct (and hence more efficient) source of information than O-S. Thus, sensitivity to O-P and O-S in word naming should *not* be taken to reflect an individual's sensitivity to these processes in general (see more on this issue in the General Discussion). Our claim is that the reliance on O-P and O-S(-P) in a word naming task reflects an individual's general tendency to rely on reliable (vs. arbitrary) regularities in the writing system, and will thus be related to overall reading skills. This hypothesis is consistent with a number of previous studies documenting variation in individuals' reliance on different cues in word naming as a function of reading skill. In fact, preliminary suggestions regarding individual differences in code utilization among skilled readers can be traced back to the 1970's: In their work on the effects of O-P regularities in the reading of English, Baron and Strawson (1976) distinguished between "Phoenician" readers who are relatively reliant on sub-lexical spelling to sound rules, and "Chinese" readers who are more dependent on lexical-level processing. Although the Phoenician vs. Chinese distinction did not hold

up to further scrutiny (e.g. Brown, Lupker, & Colombo, 1994; Yap, Balota, Sibley, & Ratcliff, 2012), the core intuition underlying this distinction—that readers differ in the degree to which they rely on different types of information—is central to more recent investigations concerned with individual differences in the division of labor. For example, Strain and Herdman (1999) demonstrated that when college students were grouped into tertiles based on phonological decoding skills (as measured by a standardized test), those in the lowest tertile of decoding skill showed the strongest imageability effect. Larger imageability effects were also found for groups of children with lower skill levels compared to groups of better early readers (Coltheart, Laxon, & Keating, 1988; Schwanenflugel & Noyes, 1996), and in parallel, early empirical and theoretical work on the effects of O-P consistency pointed towards differences between groups of adult readers (e.g. slower and faster readers, Seidenberg, 1985; readers with and without reading difficulties, Plaut et al., 1996). Relatedly, children with poorer comprehension skills are worse in reading irregular O-P words compared to children with similar decoding skills but better comprehension ability (Nation & Snowling, 1998). More recently, studies further explored individual differences in O-S processing among adult readers, as reflected in the strength of imageability effects while reading irregular words. Specifically, it was found that larger imageability effects are negatively correlated with overall word reading performance (i.e. readers who rely more on semantic imageability generally read words more slowly; Woollams et al., 2016). In the same vein, a recent paper investigating visual word recognition through the lifespan showed that imageability effects decrease with age (in both naming and lexical decision tasks) and with overall reading skill (in naming; Davies, Birchenough, Arnell, Grimmond, & Houlson, 2017). Lastly, individual differences in the magnitude of the imageability effects were also found to be correlated with patterns of neural activity and structural connectivity (Graves et al., 2014; Hoffman, Lambon Ralph, & Woollams, 2015).

The current study

It is notable that the vast majority of prior studies of individual differences in the division of labor have investigated individual differences among adult readers. To our knowledge, there are no prior studies investigating the joint influences of O-P and O-S regularities on word reading in beginning readers even at the group level. The present study addresses this gap in the literature. By investigating individual differences in children's sensitivity to O-P and O-S regularities in a word naming task and how differences in this division of labor are related to their reading skills, we sought to illuminate the processes by which developing readers become attuned to the structure of a writing system and why some individuals do this more successfully than others. Importantly, from the perspective of SL theory, understanding how sensitivity to different types of information develops can inform us regarding the actual statistics assimilated by readers as they learn to read.

Concretely, below we present data collected using a word naming task from two samples. The first sample includes a group of 3rd and 4th graders, with an over-sampling of children with reading disabilities (RD). The second sample comprises a larger group of 2nd to 5th graders, sampled from the full spectrum of reading abilities. Children in both samples participated in a word naming task in which O-P regularities, imageability, and frequency were manipulated. We used this task to estimate each individual's reliance on O-P and O-S processes using logistic models predicting accuracy from measures of O-P regularities and imageability. In addition, standardized measures of component reading skills were collected for each participant. To preview our findings, we show that in line with our predictions, individuals substantially vary in their sensitivity to O-P and O-S regularities, and that those two processes account for very large portions of individual variation in reading skills among beginning readers of English.

Study 1

Methods

Participants

Children were recruited to a larger study examining response to intervention for reading disabilities (RD), which includes behavioral and neural measures at two time points, before and after an intensive phonics-based intervention. Here we only report baseline behavioral data collected prior to intervention. The sample includes 123 children (69 boys) in the third and fourth grade (mean age in years: 9.2, SD = 0.7, range: 7.8–11.3). Because the larger study focuses on response to intervention, this sample includes over-sampling of children diagnosed with RD: 101 out of the 123 children had a RD diagnosis, whereas the rest of the sample (22) were typically developing (TD) readers. Participants come from public and charter schools from a large urban community (68% African-American, 25% Caucasian, 7% Mixed Race/other). Diagnosis of RD was made using a study-based criterion, defined as scoring at least one standard deviation below age-norm expectations on at least one standard reading assessment (see Arrington et al., 2019 for details). Due to the over-sampling of children with RD, participants had generally low average standard reading scores on Woodcock-Johnson III sub-tests (Woodcock, McGrew, & Mather, 2001): Letter-Word Identification: M = 92.08, SD = 12.24; Word Attack: M = 91.21, SD = 11.71; Passage Comprehension: M = 84.55, SD = 13.19; and Reading Fluency: M = 90.76, SD = 14.41).

Materials, design and procedure

Within a larger battery of behavioral and neural assessments, each child participated in a word naming task, as well as four sub-tests of the Woodcock-Johnson III, which are the focus of the analyses reported here.

The word naming task included 160 trials presented to children in a fixed order. In each trial, a fixation cross appeared for 500 ms, and was then replaced by a monosyllabic word which was presented in the center of the screen until response. Participants were asked to read out loud each word as accurately and quickly as possible, and their responses were coded by an experimenter who sat in the experiment room. In addition, children's responses were recorded and in cases when the experimenter could not decide during the experimental session whether a correct response was made they went back and listen to the recordings. Words were selected so they would be generally familiar to children in the 2nd grade and up, and so they would vary along the three independent variables: frequency, imageability, and O-P regularity. Frequency (log-transformed) was estimated for each word based on the Zeno corpus, grades 1–8 (Zeno, Ivens, Millard, & Duvvuri, 1995), and words' imageability was based on standard ratings (Paivio et al., 1968). O-P regularity was operationalized as vowel surprisal (i.e. $-\log(p(i))$) of the vowel pronunciation, which is a function of the relative likelihood of the pronunciation of a vowel grapheme (see Siegelman et al., 2020). Hence, for example, the word *pint* has a higher surprisal value than *mint*, since $p(i \rightarrow /aɪ/)$ is smaller than $p(i \rightarrow /ɪ/)$[1]. Note that the decision to focus on regularities at vowel level differs from previous investigations of the imageability by consistency interaction, which have generally operationalized O-P consistency in terms

---

[1] We used the values provided in the recent work by Siegelman et al. (2020), which quantifies grapheme-phoneme regularities based on a corpus of 10,093 monosyllabic English words. To ensure that our estimates of individuals' reliance on O-P regularities (see below) were not skewed by utilizing a corpus that is based on adult reading materials, we examined the item-level estimates of O-P surprisal for the 160 items in the word naming task based only on words that appear in the Zeno corpus (Zeno et al., 1995) in grades 1 and 2 (3632 words). This calculation revealed a minimal impact of the type of corpus used with a near-perfect correlation between the values based on the child-directed input and the original values provided by Siegelman et al. (2020), of *r* = 0.985.

of body-rime correspondences. Our choice of focusing on the vowel-level was based on results indicating that young readers rely more on grapheme-phoneme than body-rime regularities (Steacy et al., 2018; Treiman & Kessler, 2006).[2] Items were selected to minimize the correlations between the three independent variables. Thus, the correlations between the three independent variables across the 160 items were small (O-P surprisal and imageability: $r = -0.08$; O-P surprisal and frequency: 0.04; imageability and frequency: 0.16). Accuracy in each trial was used as the dependent variable, and was coded as 1 (correct) or 0 (incorrect) in each trial. Trials with microphone malfunctions or unclear responses were disregarded from further analysis (14.5% of all trials).[3] Please refer to the Supplementary Material for the full list of items and their frequency, imageability, and O-P surprisal values, and mean by-item accuracy levels in Studies 1 and 2.

Each participant also completed four sub-tests of the Woodcock-Johnson III Tests of Achievement battery: Letter-word Identification (measuring letter and word decoding), Word-Attack (pseudo-word reading), Reading Fluency, and Passage Comprehension. We use the raw scores in each of these four sub-tests as outcome measures.

Results and discussion

Group-level analysis

Before turning to the main focus of the paper – analyses of individual differences – we first report the group-level effects in the word naming task. To examine these, we ran a logistic mixed-effect model (Jaeger, 2008) using the *lme4* package in R (Bates, Maechler, Bolker, & Walker, 2015), with correctness in each trial as the dependent variable. The model included frequency, imageability, and O-P surprisal as well as all interactions as fixed effects. The three predictors were scaled and centered to avoid collinearity. The model also included random intercepts for subjects and words, and random slopes for O-P surprisal and frequency (the maximal random effect structure that converged; Barr, Levy, Scheepers, & Tily, 2013). Collinearity in the model was small (all $|r| < 0.25$). Table 1 presents the estimated fixed effects. As can be seen, and in line with previous studies, we found significant effects of all three main-effects on accuracy: words were named more accurately if they were higher on imageability, higher on frequency, and lower on O-P surprisal (i.e. *more* regular in terms of O-P). In addition to the main effects, we found two significant interactions: O-P regularity by frequency, and imageability by frequency. The O-P regularity by frequency interaction resembled that in adult studies, with a stronger O-P effect in infrequent compared to frequent words. In contrast, the imageability by frequency interaction went in an opposite direction compared to adult studies: The effect of imageability was stronger in frequent compared to infrequent words (i.e. super-additive interaction of imageability by frequency). The significant interactions are visually depicted in Fig. 1. We return to the pattern of interactions in the General Discussion.

Individual-differences analysis

Our main analyses focused on the effects of the three independent variables – O-P regularity, imageability, and frequency – on the word

Table 1
Group-level effects on accuracy in the word naming task in Study 1.

Predictor	β (coefficient)	SE	z value	p value
IMG	.234	.081	2.90	**.003**
O-P REG	-.500	.085	−5.86	**< .001**
FREQ	1.03	.085	3.22	**< .001**
IMG*O-P REG	-.135	.083	−1.64	.10
IMG*FREQ	.324	.093	3.48	**< .001**
O-P REG*FREQ	.269	.085	3.15	**.002**
IMG*O-P REG*FREQ	.154	.089	1.72	.09

Notes: IMG: imageability; O-P REG: O-P regularity (surprisal); FREQ: Frequency.

naming behavior of each subject. To estimate these individual-level effects, we ran a set of three logistic regression models on the data of each participant, including the subject's trial-by-trial data from the word naming task. All three models had accuracy in each trial as the dependent variable, while each model included one of the three predictors (centered and scaled) as the independent variable. The output of these models were used to estimate the impact of each of the three predictors on each individual's word naming accuracy. Thus, each child had one slope score that estimates sensitivity to O-P regularity, one slope score that quantifies the effect of semantic imageability, and one slope score that indexes reliance on word frequency. Also note that because our individual-level models were based on predicting accuracy, at least some naming errors are required (i.e., no slopes can be estimated for a subject whose performance is perfect). We therefore excluded from the individual-level analysis data from two subjects who had a mean word naming accuracy of more than 98% (unsurprisingly, both subjects were from the TD group). Our individual-differences analyses below are based on the data of remaining 121 participants, whose mean accuracy varied between 10% and 97.8% (mean = 69.4%, SD = 18.5%). Note that due to inter-individual variance in the number of correct responses, estimates of slope scores are expected to vary in their reliability across subjects: Less reliable estimates are expected for subjects with very high or low accuracy rates, compared to subjects with a large enough number of correct and incorrect trials. Nonetheless, we do not expect this variability in reliability to systematically bias the correlation between the slope scores and reading skill – which are the center of the current analyses. Rather, if anything, the increased measurement error for some subjects should result in an under-estimation of the true correlation due to attenuation.[4] Also note that although in the surprisal measure of O-P regularities higher values represent more surprising (i.e. unpredictable) readings, for simplicity, we 'flipped' the O-P surprisal slope scores in all further analyses. Thus, the slope score of sensitivity to O-P regularity was coded such that higher, positive, values represent *more* sensitivity to O-P regularities, in the same direction as the slope scores reflecting sensitivity to imageability and frequency.

Given our theoretical focus, we first focused on the individual differences in sensitivity to O-P regularity and imageability, and their relation to individual differences in reading skills. These results are summarized in Fig. 2, which depicts the spread of individuals according to each child's sensitivity to O-P regularity (x-axis) and imageability (y-axis). As can be seen, and as expected given the group-level analysis above, readers exhibited sensitivity to both O-P regularity and imageability on average (vertical and horizontal dashed lines, both significantly higher than zero; O-P regularities: one-sample $t(120) = 15.35, p < 0.001$; imageability: $t(120) = 14.01, p < 0.001$). Importantly, however, there was substantial variability along the two

---

[2] Also note that generally there are high correlations between vowel-level and body-level regularities across items (i.e. words that are more regular at one grain size are also more regular at the other; see Siegelman et al., 2020). This was also the case in the current items, which were not selected to isolate sensitivity to vowel-level vs. body-level information. Differentiating between sensitivity to regularities in the two grain-sizes requires selection of items in which the two properties are uncorrelated.

[3] Because participants in this age range produced a large number of unclear or idiosyncratic responses (e.g., 'hmmm', 'when is this done') leading to unclear or inaccurate speech onsets, we a-priori decided not to further analyze latency data from the reading task and focus on accuracy, which tends to be more reliable in this age range.

[4] To account for this issue, future modifications to the task can adopt an adaptive design, where items are chosen online based on each child's performance in the naming task, so a similar mean accuracy rate is achieved for all subjects.
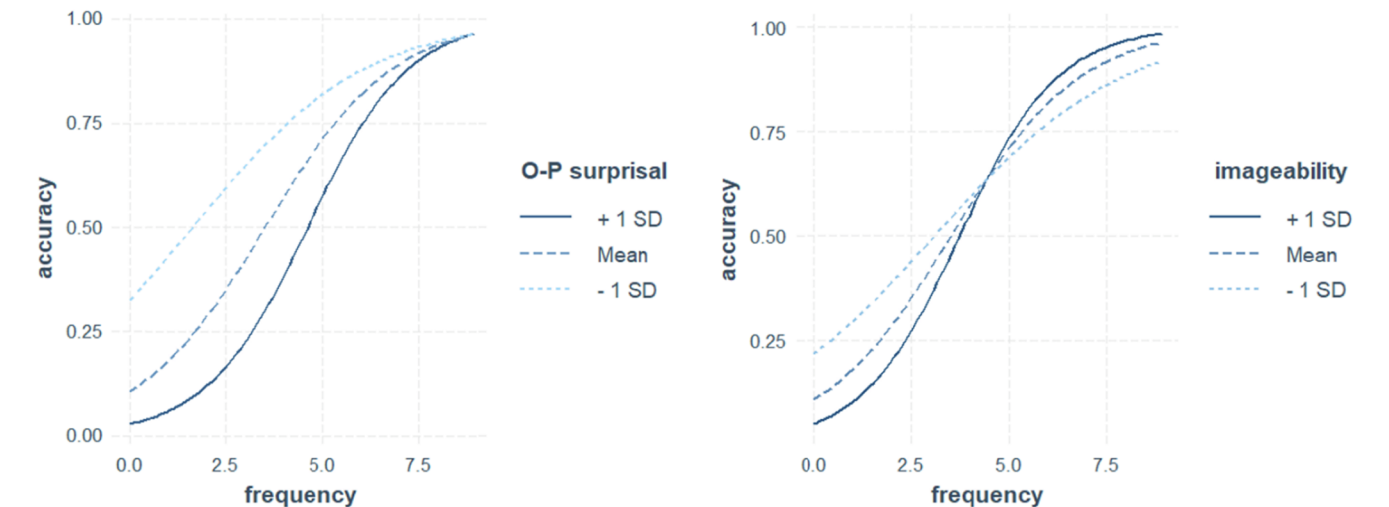
Fig. 1. Significant interactions in the group-level analysis of Study 1. Left panel: greater impact of frequency in higher levels of surprisal (i.e. when O-P regularities are *less* consistent). Right panel: greater impact of frequency in higher levels of imageability. Plots created using the *interact_plot*() function in *interactions* package in R (Long, 2019).
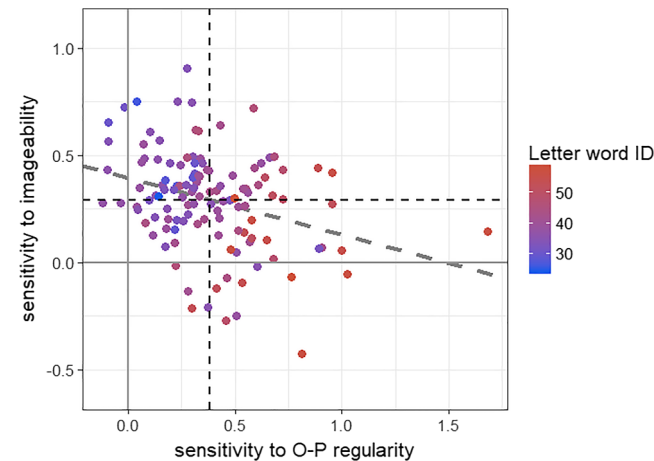

Fig. 2. Variability among participants in Study 1 in sensitivity to imageability (y-axis) and O-P regularities (x-axis). Vertical and horizontal dashed lines show mean sensitivity to O-P regularity and imageability, respectively. Dashed trend line show correlation between the two slope measures. Color scale presents raw scores on Letter-Word Identification task. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

axes, suggesting that not all individuals are similarly impacted by these two properties. Also note that sensitivity to O-P regularity and imageability were significantly negatively correlated ($r = -0.32$, $p < 0.001$), although this correlation was moderate in magnitude, suggesting that the effects of O-P regularity and imageability reflect related, yet separate, processes.

Next, we examined the relations between each individual's reliance on O-P and O-S associations as assessed by the slope scores and their overall reading skills, estimated by the standardized reading measures. We thus ran four multiple regression models, all including the two slope scores as independent variables, and raw scores in one of the Woodcock Johnson sub-tests as a dependent variable. The results of these models are summarized in Table 2. Across all four sub-tests, we found that greater sensitivity to O-P regularity and lower sensitivity to imageability were associated with better reading performance. Note that the effects were not only significant, but also of large magnitude (adj-R^2 values: Word-Attack: 53%; Letter-Word Identification: 46%; Reading Fluency: 44%; Passage Comprehension: 32%). As an illustration of the

Table 2

Regression models predicting reading skill from sensitivity to imageability and O-P regularity in Study 1.

Predictor	β (coefficient)	SE	z value	p value
Dependent variable: Letter-Word Identification (adj-R^2 = 46%)				
O-P REG	4.57	0.55	8.26	**< .001**
IMG	−1.63	0.55	−2.95	**.004**
Dependent variable: Word Attack (adj-R^2 = 53%)				
O-P REG	3.65	0.43	8.55	**< .001**
IMG	−2.15	0.43	−5.03	**< .001**
Dependent variable: Reading Fluency (adj-R^2 = 32%)				
O-P REG	5.92	1.02	5.80	**< .001**
IMG	−3.06	1.02	−2.99	**.003**
Dependent variable: Passage Comprehension (adj-R^2 = 44%)				
O-P REG	2.87	0.39	7.35	**< .001**
IMG	−1.47	0.39	−3.77	**< .001**

Notes: IMG: imageability; O-P REG: O-P regularity (surprisal). Predictors are centered and scaled.

strength of this association, Fig. 2 presents letter-word identification scores in color scale: It can be clearly seen that as individuals' sensitivity to O-P regularity increases (i.e. higher scores along the a-axis), and sensitivity to imageability decreases (i.e. lower scores along the y-axis), Letter-Word Identification scores increases. Next, we also examined whether sensitivity to frequency adds to the predictive value of O-P regularity and imageability, by adding to the multiple regression models the slope scores reflecting reliance on frequency. Notably, in three out of four models (letter-word identification, word attack, passage comprehension) adding frequency to the models did not result in significant improvement in model fit ($p > 0.05$; Δadj-R^2 < 1%). Sensitivity to frequency was significantly associated with Passage Comprehension scores beyond sensitivity to O-P regularity and imageability ($β = 2.32$, SE = 1.01, $p = 0.024$), albeit with a small improvement in model fit (Δadj-R^2 = 2.35%). Together, then, measures of sensitivity to O-P regularity and imageability based on the word naming task were better predictors of participants' reading skills than sensitivity to frequency.

The associations of reading skills with the individual-level effects of O-P regularity and imageability (and the numerically smaller magnitude of the correlations with that of frequency) is also apparent in Table 3 which shows all zero-order correlations between the three slope measures and the reading sub-tests. Whereas reading skills were strongly positively correlated with sensitivity to O-P regularity (all $0.67 \geq r$'s ≥ 0.53), and negatively correlated with imageability (all

Table 3

Bi-variate Pearson correlations of sensitivity to O-P regularity, imageability, and frequency with standardized reading measures in Study 1. Significant correlations (p < 0.05) are shown in bold.

Measure	2	3	4	5	6	7
1. O-P regularity	−0.32	−0.30	0.65	0.67	0.62	0.53
2. Imageability	—	0.26	−0.39	−0.51	−0.44	−0.38
3. Frequency		—	−0.16	−0.33	−0.10	−0.04
4. Letter-Word Identification			—	0.78	0.83	0.85
5. Word Attack				—	0.74	0.65
6. Passage Comprehension					—	0.80
7. Reading Fluency						—

$-0.31 \geq r\text{'s} \geq -0.58$), their correlations with sensitivity to frequency were generally lower ($-0.04 \geq r\text{'s} \geq -0.33$). Nonetheless, and despite the fact that reliance on frequency was generally not significantly related to reading performance above and beyond reliance on O-P regularity and imageability, it is worth mentioning that sensitivity to frequency was generally negatively associated with reading skills. This aligns with findings with adult readers showing that higher reading skills are correlated with lesser impact by word frequency in different tasks (e.g. Kuperman & Van Dyke, 2013; Schilling, Rayner, & Chumbley, 1998). Nonetheless, our data suggest that O-P regularity and imageability are more strongly correlated with early reading skills compared to sensitivity to frequency.

Lastly, we examined whether sensitivity to O-S and O-P processes is associated not only with continuous measures of reading, but also with categorical TD/RD diagnosis – that is, whether the two slope scores can distinguish between children with and without RD diagnosis. A logistic model with TD/RD diagnosis of each child and slope scores of O-P regularities and imageability revealed that indeed, higher sensitivity to O-P regularities and lower reliance on imageability were associated with a TD (rather than a RD) diagnosis (sensitivity to O-P regularities: B = 2.21, SE = 0.52, Z = 4.21, $p < 0.001$; imageability B = −0.78, SE = 0.35, Z = −2.24, $p = 0.02$; McFadden's multiple-R^2 = 45%). Fig. 3 visually depicts these results, showing that participants without reading disabilities (TD readers) are those who show greater sensitivity to O-P regularities and lower sensitivity to imageability compared to their peers (i.e. TD readers are concentrated in a limited portion of the regularity by imageability space, generally exhibiting larger than average sensitivity to regularity, and smaller than average sensitivity to imageability). Among RD readers, however, there is substantial variability in the extent of sensitivity to the two sources of information.
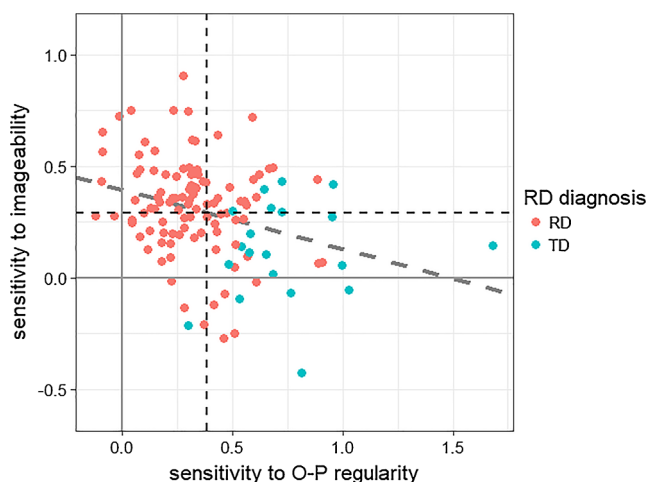

Fig. 3. RD diagnosis (in color) as a function of sensitivity to imageability (y-axis) and O-P regularity (x-axis). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Study 2

The primary goal of Study 2 is to replicate the findings of Study 1. We utilize a similar task to estimate individuals' reliance on O-S and O-P associations during word naming, and examine the relation between these processes and individual differences in reading skills. Study 2 includes an even larger sample of children compared to Study 1. Moreover, this sample consists of a cohort of children which are more representative of typical variability in reading skills, as they were sampled from the general population of early readers without over-sampling of children with RD. This enables us to examine whether reliance on the two processes is associated with variability in reading skill also among children without a formal diagnosis of RD.

Methods

Participants

A total of 282 children (143 boys) participated in Study 2. All children were in the 2nd to 5th grade, and their mean age was 9.43 (SD = 1.17, range = 7.3–13.08). Participants came from both public and private schools, and the sample was made up of 51% African American, 36% Caucasian, 11.5% Hispanic, and 1.5% Mixed Race students. As expected given the differences in the samples, participants had higher average standardized reading assessment scores compared to those of Study 1 (Letter-Word Identification: M = 97.59, SD = 11.47; Word Attack: M = 98.12, SD = 9.56; Passage Comprehension: 90.40, SD = 11.41; all $p\text{'s} < 0.001$).

Materials, design and procedure

The word naming task contained the same items as Study 1. Students were asked to read the words in a list format as quickly and accurately as they could and a research assistant coded the child's responses. Audio of the experimental session was recorded and experimenters could go back and listen to the recordings in cases where they could not decide whether a correct response was made during the experimental session. Children also participated in three sub-tests of reading skills from the Woodcock-Johnson III: Letter-Word Identification, Word Attack, and Passage Comprehension.

Results and discussion

Group-level analysis

As in Study 1, we first ran a logistic mixed-model to estimate the group-level effects in the word naming task. The model again included frequency, imageability, and O-P surprisal (all standardized) as well as all interactions as fixed effects, with accuracy as the dependent variable. The model included by-subject and by-item random intercepts, and by-subject random slopes for O-P surprisal and frequency (the maximal random effect structure that converged; Barr et al., 2013). Collinearity in the model was again small (all $|r| < 0.25$). As can be seen in Table 4, higher frequency and lower O-P surprisal (i.e. higher O-P regularity) were both associated with higher naming accuracy. In contrast, the effect of imageability on group level accuracy did not reach significance. Thus, on average, naming performance for this group of readers was not significantly impacted by imageability, as opposed to performance in the sample of Study 1. One possible reason for this arises from the fact that on average Study 2's participants are better readers than the mostly RD sample of Study 1 (and see aggregated analysis section, below). The effects of the interactions mirrored those in Study 1: There was a sub-additive frequency by O-P surprisal interaction (i.e. a stronger O-P regularity effect in infrequent words), and a super-additive interaction between frequency and imageability (a stronger frequency effect in highly imageable words).

Individual-differences analysis

Four children had mean word naming accuracy of more than 98%

Table 4
Group-level effects on accuracy in the word naming task in Study 2.

Predictor	β (coefficient)	SE	z value	p value
IMG	0.128	0.087	1.47	.14
O-P REG	−0.584	0.091	−6.40	**< .001**
FREQ	1.028	0.089	11.46	**< .001**
IMG*O-P REG	−0.067	0.088	−0.76	.44
IMG*FREQ	0.269	0.099	2.70	**.007**
O-P REG*FREQ	0.218	0.090	2.41	**.016**
IMG*O-P REG*FREQ	0.172	0.095	1.82	.07

Notes: IMG: imageability; O-P REG: O-P regularity (surprisal); FREQ: Frequency.

and their data were therefore removed from further analysis. For each of the remaining 278 children (accuracy range: 9–97%, mean = 78.9%, SD = 19.8%), we again estimated sensitivity to O-P regularity, imageability, and frequency, using logistic models predicting accuracy of each subject at each trial from each predictor. Note that we again 'flipped' the slope scores of O-P surprisal such that higher scores reflect more sensitivity to O-P regularities.

Fig. 4 shows the spread of individuals along the two critical dimensions: sensitivity to O-P regularity and imageability. Despite the fact that there was no group-level effect of imageability in the mixed-model above, mean slope scores of sensitivity to O-P regularity and imageability both significantly differed from 0 (vertical and horizontal dashed lines: O-P regularities: one-sample $t(277) = 26.41, p < 0.001$; imageability: $t(277) = 13.72, p < 0.001$). This suggests that the lack of observed sensitivity to imageability at the group level is related to the inclusion in the model of the higher order interactions, substantial shrinkage, and/or substantial variance across items (which mixed-models take into account; see Jaeger, 2008 for discussion). Note also that although sensitivity to O-P regularity and imageability were again numerically negatively correlated, this correlation was very weak and failed to reach significance ($r = -0.06, p = 0.29$; and see aggregated analysis section, below).

Most importantly, however, and in line with the results of Study 1, sensitivity to imageability and O-P regularity were again associated with reading skills. This is reflected in Fig. 4 which depicts letter-word identification scores (in color scale) as a function of the two slope scores. The full results are shown in Table 5, which shows the outputs of three multiple regression models with the three Woodcock-Johnson
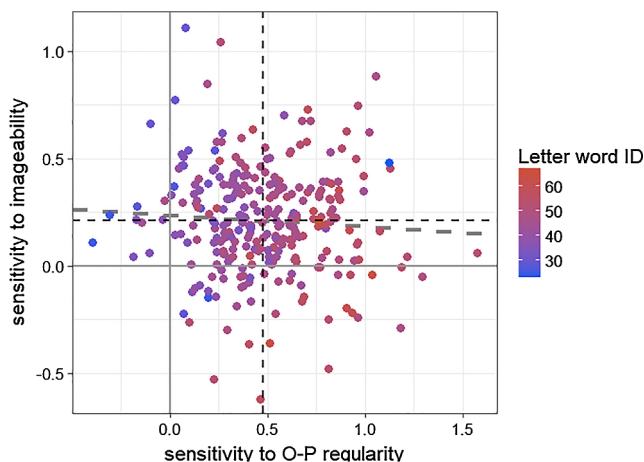

Fig. 4. Variability among participants in Study 2 in sensitivity to imageability (y-axis) and O-P regularities (x-axis). Vertical and horizontal dashed lines show mean sensitivity to O-P regularity and imageability, respectively. Dashed trend line show correlation between the two slope measures. Color scale presents raw score on Letter-Word Identification task. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 5
Regression models predicting reading skill from sensitivity to imageability and O-P regularities in Study 2.

Predictor	β (coefficient)	SE	z value	p value
Dependent variable: Letter-Word Identification (adj-R² = 38%)				
O-P REG	4.59	0.36	12.73	**< .001**
IMG	−1.03	0.36	−2.85	**.005**
Dependent variable: Word Attack (adj-R² = 39%)				
O-P REG	3.51	0.28	12.60	**< .001**
IMG	−1.08	0.28	−3.89	**< .001**
Dependent variable: Passage Comprehension (adj-R² = 24%)				
O-P REG	2.35	0.27	8.81	**< .001**
IMG	−0.74	0.27	−2.77	**.006**

Notes: IMG: imageability; O-P REG: O-P regularity (surprisal); FREQ: Frequency.

sub-tests as dependent variables. As in Study 1, in all three models we found that higher sensitivity to O-P regularity and less sensitivity to imageability were associated with better reading skills. Also note that again effect sizes were large (adj-R² values: Word Attack: 39%; Letter-Word Identification: 38%; Passage Comprehension: 24%), albeit somewhat lower than in Study 1 (and see aggregated analysis section for further discussion). As for sensitivity to frequency, we again found that it had only limited relations to reading skills beyond sensitivity to imageability and O-P regularity: Adding slopes of sensitivity to frequency did not improve model fit for Letter-Word Identification and Passage Comprehension ($p > 0.05$), although it did have a significant yet small added effect on Word Attack scores (β = −0.84, SE = 0.28, $p = 0.003$, Δadj-R² = 1.7%). Table 6 includes the full correlation matrix with the three slope scores and the Woodcock-Johnson sub-tests. In line with the results of Study 1, the reading skills scores were significantly positively correlated with sensitivity to O-P regularity, significantly negatively correlated with imageability, and negatively (yet generally insignificantly) correlated with sensitivity to frequency.

Aggregated analysis

As noted above, participants in the two studies were sampled from two different populations: Although the sample of Study 1 consisted mostly of RD readers, participants in Study 2 were sampled from the general population of early readers. Indeed, this different sampling was reflected in significant differences in the three sub-components of the Woodcock-Johnson sub-tests that were administered in both samples (see above). Importantly, the difference between the samples was also reflected in the slope scores extracted from the word naming task. Thus, in line with the results above, the better readers of Study 2 showed greater mean slopes of O-P regularity and smaller slopes of imageability than the mostly-RD participants of Study 1 (O-P regularity: $t(397) = 3.02; p = 0.003$; imageability: $t(397) = 3.01; p = 0.002$).

Despite the different samples, the results of Studies 1 and 2 point to a similar general conclusion: In both studies greater sensitivity to O-P regularity and lower sensitivity to imageability were associated with higher reading skills. Unsurprisingly, aggregating the data from the two

Table 6
Bi-variate Pearson correlations of sensitivity to O-P regularities, imageability, and frequency with standardized reading measures in Study 2. Significant correlations (p < 0.05) are shown in bold.

Measure	2	3	4	5	6
1. O-P regularity	−0.06	−0.17	**0.61**	**0.60**	**0.47**
2. Imageability	—	0.13	**−0.17**	**−0.22**	**−0.18**
3. Frequency		—	−0.10	**−0.26**	−0.04
4. Letter-Word Identification			—	**0.76**	**0.74**
5. Word Attack				—	**0.62**
6. Passage Comprehension					—

Table 7
Regression models predicting reading skill from sensitivity to imageability and O-P regularities in the aggregated analyses (n = 399).

Predictor	β (coefficient)	SE	z value	p value
Dependent variable: Letter-Word Identification (adj-R^2 = 43%)				
O-P REG	4.83	0.31	15.76	< .001
IMG	−1.39	0.31	−4.53	< .001
Dependent variable: Word Attack (adj-R^2 = 45%)				
O-P REG	3.84	0.25	15.67	< .001
IMG	−1.61	0.25	−6.57	< .001
Dependent variable: Passage Comprehension (adj-R^2 = 32%)				
O-P REG	2.69	0.23	11.91	< .001
IMG	−1.12	0.23	−4.95	< .001

Notes: IMG: imageability; O-P REG: O-P regularity (surprisal); FREQ: Frequency.
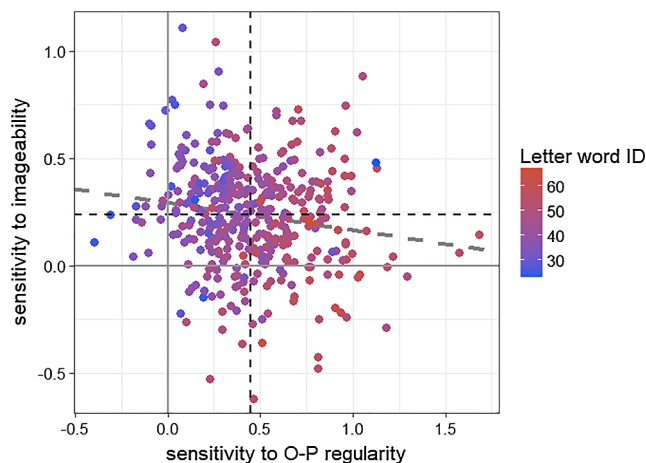

Fig. 5. Variability among participants across the two studies in sensitivity to imageability (y-axis) and O-P regularities (x-axis). Vertical and horizontal dashed lines show mean sensitivity to O-P regularity and imageability, respectively. Dashed trend line show correlation between the two slope measures. Color scale presents raw score on Letter-Word Identification task. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

samples (*n* = 399) showed that sensitivity to O-P regularity and imageability were both strongly associated with reading skills, as expected given the separate results of Study 1 and 2 (see Table 7 for output of regression models predicting performance on the Woodcock-Johnson sub-tests, and Fig. 5 for an aggregated scatter plot).

Despite the similarities in the effects of the slope scores on reading across the two studies, there were two differences between the samples that warrant additional investigation. First, whereas in Study 1 we observed a significant negative correlation (of $r = -0.32$) between O-P regularity and imageability, no significant correlation between the slope scores was observed in Study 2 ($r = -0.06$), and these two correlation coefficients significantly differed from one another ($Z = 2.47$, $p = 0.01$). Second, effects sizes (i.e., percent of variance explained) were generally higher among the lower-skilled readers of Study 1 compared to those of Study 2 (see R^2 values in Table 2 vs. Table 5, above).

These two discrepant findings raise the possibility that the development of the two processes and their relation to reading behavior is non-linear. For example, it is possible that knowledge of the correspondences between graphemes and phonemes follows a non-linear trajectory, developing rapidly (and showing substantial individual differences) among early/poor readers, while its growth saturates (or even decreases) among readers that are more further along the literacy acquisition trajectory. This prediction conforms with developmental studies documenting a non-linear trajectory of O-P consistency effects (e.g.

Weekes et al., 2006), and with non-linear effects in language and skill acquisition more broadly (e.g. Farnia & Geva, 2011; Newell, 1991). Importantly, aggregating the data from the two studies allowed us to achieve sufficient power to examine the possibility of such non-linear effects using more sensitive statistical analysis.

To do so, we ran Generalized Additive Models (GAM) on the aggregated data from the two studies (*n* = 399 children). In contrast to linear models, in GAMs the relation between a dependent variable and a predictor or a set of predictors does not have to be linear. Instead, GAMs use smoothed terms of the predictor(s) to find the function that best fits the outcome. Then, these models can be compared to models without smoothed terms for one or more of the predictors, to examine whether the non-linear smoothing indeed accounts for the data better than a simple linear function (i.e. whether the non-linear transformation is warranted by the data). Here we report two GAMs using the *mgcv* package in R (Wood, 2011), both with O-P regularity and imageability as predictors. The dependent variable in the first model was Letter-Word Identification scores, and Word Attack scores in the second.

The results of the GAMs are presented in Fig. 6. As can be seen in the top panel, the estimated effect of imageability on the two dependent variables was negative as expected, and importantly - almost entirely linear. Indeed, a Chi-square test comparing the GAMs with smoothed terms for both O-P regularity and imageability to models without a smoothed term for imageability revealed no significant difference in models' fit (*p* > 0.05 for both dependent variables). In contrast, the positive effect of O-P regularities on reading skills was non-linear: Model comparison revealed a significant increase in model fit for models with a non-linear term for O-P regularity (for both models: *p* < 0.001). More specifically, the impact of sensitivity to O-P regularity on reading skills had a seemingly logarithmic trajectory, with a strong positive impact on reading at lower skills, which saturates at higher skill level (see Fig. 5, top panel). The joint impact of sensitivity to imageability and O-P regularities on the two dependent variables are depicted in the bottom panel of Fig. 5, which presents predicted reading scores given a reader's sensitivity to O-P regularities and imageability. This again shows the clear relation between the two dimensions and reading skills, while also taking into account the non-linear impact of sensitivity to O-P regularity on reading.

Lastly, we wished to examine the role of vocabulary knowledge in the development of the reliance on O-P and O-S processes and their associations with reading skills. This analysis is motivated by behavioral studies showing that reading development is related not only to decoding skills but to oral learning skills more broadly (e.g. Nation & Snowling, 2004), and by recent computational work suggesting that vocabulary skill is associated with a more efficient division of labor (as better oral language skills reflect better knowledge of S-P and P-S mappings, which is then beneficial for the development of an efficient division of labor between O-P and O-S; Chang & Monaghan, 2019; Chang, Monaghan, & Welbourne, 2019). Indeed, we found significant correlations between a measure of vocabulary knowledge (raw scores in the vocabulary sub-test of the WASI-II, Wechsler, 2011) and sensitivity to O-P and O-S: a positive correlation with reliance on O-P ($r = 0.2$), and a negative association with O-S ($r = -0.27$ both p's < 0.001). Importantly, however, even after controlling for vocabulary skills we found significant associations (with large effect sizes) between sensitivity to O-P and O-S and reading outcomes (Table 8). Thus, although better knowledge of S-P/P-S does support the development of a more efficient division of labor as reflected in word naming (i.e. more reliance on O-P and less reliance on O-S), the associations between sensitivity to O-P and O-S and reading skill cannot be explained solely by variance in vocabulary knowledge.

General discussion

The current work revisits a basic question: What are the processes that lead to successful early reading acquisition? From an SL
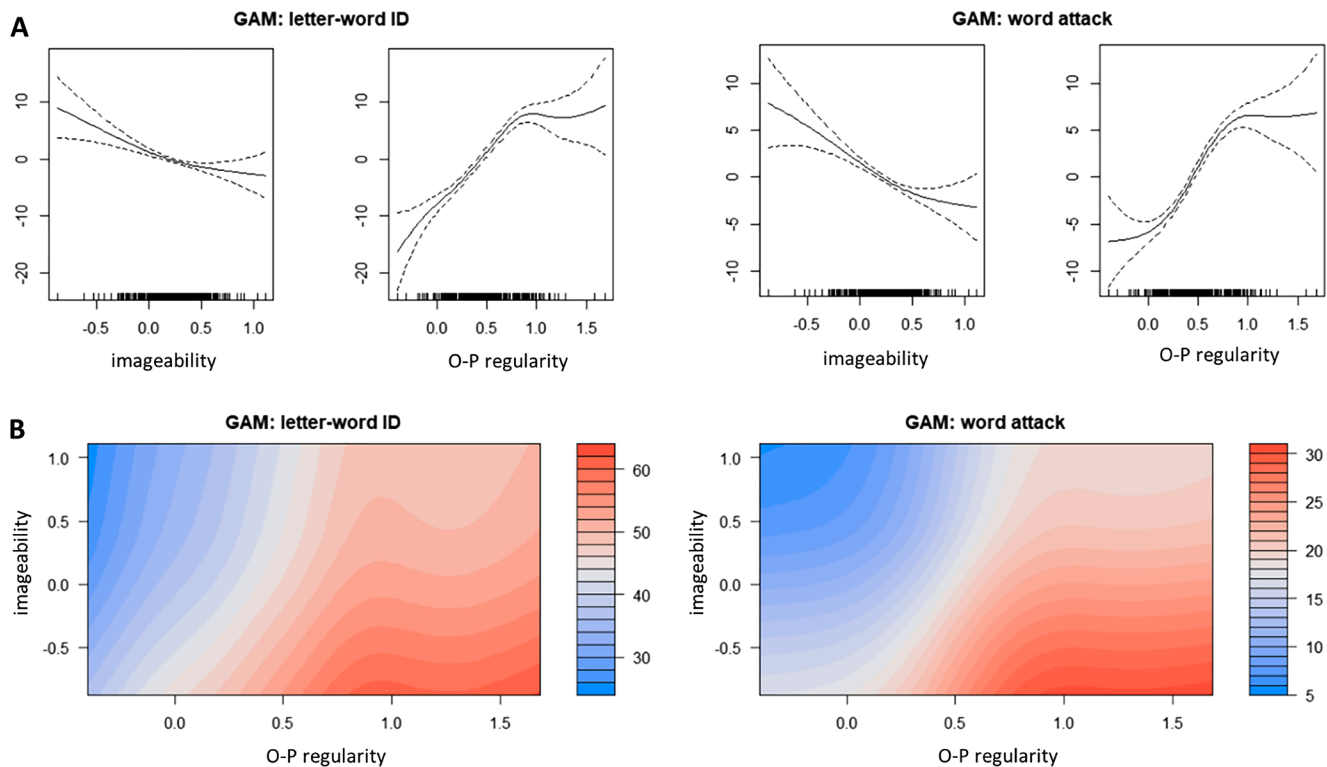
Fig. 6. GAM results based on the aggregated data (*n* = 399). Panel A: Estimated effects of the smoothed terms of sensitivity to imageability and O-P regularity on (residualized) Letter-Word Identification (left) and Word Attack (right) scores. Panel B: Predicted scores in Letter-Word Identification (left) and Word Attack (right) as a function of sensitivity to O-P regularity (x-axis) and imageability (y-axis).

Table 8

Regression models predicting reading skill from sensitivity to imageability and O-P regularities in the aggregated analyses (n = 399), while controlling for vocabulary knowledge.

Predictor	β (coefficient)	SE	z value	p value	ΔR^2
Dependent variable: Letter-Word Identification (adj-R^2 = 47%)					
O-P REG	4.55	0.30	15.24	< .001	34%
IMG	−0.97	0.31	−3.17	.002	
Vocabulary	1.79	0.31	5.81	< .001	5%
Dependent variable: Word Attack (adj-R^2 = 48%)					
O-P REG	3.63	0.24	15.05	< .001	36%
IMG	−1.29	0.25	−5.24	< .001	
Vocabulary	1.30	0.25	5.22	< .001	4%
Dependent variable: Passage Comprehension (adj-R^2 = 49%)					
O-P REG	2.32	0.20	11.65	< .001	20%
IMG	−0.53	0.20	−2.60	.009	
Vocabulary	2.33	0.21	11.31	< .001	17%

Notes: IMG: imageability; O-P REG: O-P regularity (surprisal); FREQ: Frequency. ΔR^2 values are the difference in R^2 between the full model and a partial model that includes only the relevant predictor(s).

perspective, achieving proficient reading requires the assimilation of statistical regularities present across varied linguistic dimensions in the writing system. Thus, early readers who are better in assimilating these regularities are expected to be more proficient readers than those with poorer SL abilities. The current results provide strong evidence for this theoretical claim. Namely, data from almost 400 early readers of English show that children who show greater sensitivity to the correspondences between print and speech in word naming have better reading skills. In parallel, individuals who display greater reliance on associations between print and meaning, and particularly those who are also not sensitive to O-P correspondences, are characterized by lower reading abilities.

A small number of previous studies have investigated differences in

the imageability effect between sub-groups of children who vary in their reading skills. For example, Coltheart et al. (1988) found that imageability facilitated word naming for the poorer readers among a sample of nine-year-old children but had no effect on the performance of the better readers. Relatedly, although Schwanenflugel and Noyes (1996) found limited evidence that imageability affects word naming by 3rd- and 5th-grade children, their results also suggested that imageability effects are larger for poorer readers. Joining these studies is the work by Davies et al. (2017) showing that the magnitude of the imageability effect is modulated by both age and reading skill in a sample of 8–83 year-old readers, with larger imageability effects for younger, and poorer, readers. Our results extend these findings in an important way: Because we manipulated both imageability *and* O-P regularity, we documented the joint contribution of the reliance on regularities in O-P as well as O-S over a large sample of children with diverse reading skills. Similarly, a handful of previous studies have investigated individual differences in the division of labor between O-P and O-S contributions to word naming in adults (e.g. Strain & Herdman, 1999; Woollams et al., 2016). Our study extends this work by investigating these processes during a period of the developmental trajectory for which we have theoretical and computational insights (e.g. Harm & Seidenberg, 2004) but a dearth of empirical data. In line with this computational work, our results suggest that developing an efficient division of labor between O-P and O-S is crucial for early reading success.

Although the main focus of the present investigation is individual-level variation, the current results also present intriguing group-level effects that shed additional light on the development of the division of labor between O-P and O-S processes. Specifically, in both studies we observed a similar pattern of interactions between frequency, imageability, and O-P regularity on group-level performance: We found that whereas the O-P regularity effect was stronger with infrequent compared to frequent words, the imageability effect was *stronger* with frequent compared to infrequent words. Further, neither the interaction

between imageability and O-P regularity nor the three-way interaction of frequency, regularity, and frequency were statistically reliable. Thus, whereas the O-P regularity by frequency interaction resembles the effect typically observed in adults, the lack of interactions between imageability and regularity, and the opposite interaction between imageability and frequency, differed from previous studies with adult samples (e.g. Strain et al., 1995). This pattern of interactions may be due to differences in the rate of learning of O-P and O-S associations: O-S associations are learned more slowly and thus at the age range of the children in our sample the O-S associations for lower-frequency and/or O-P irregular words may be too weak to contribute significantly to word naming (see Duff & Hulme, 2012, and Laing & Hulme, 1999, for similar results in a word learning paradigm). From this perspective, the group-level results compliment the individual-level findings: O-P associations are learned earlier and are strongly related to reading skills among early readers; full knowledge of the more arbitrary O-S associations is expected to emerge only later and at this developmental stage general reliance on O-S (when other sources of information are present) is associated with lower reading skills (see also Harm & Seidenberg, 2004).

Although the results of our study shed light on an important aspect of literacy acquisition, they also raise several issues that should be addressed by further research. One issue concerns the nature of the mechanism that gives rise to individual differences in sensitivity to O-P regularities. As mentioned above, some portion of the variance in reliance on O-P (as well as on O-S) was related to vocabulary skills: A measure of vocabulary skill was positively associated with the magnitude of the O-P effect, and negatively associated with the imageability effect. This suggests that one factor that contributes to the development of O-P knowledge (and a more efficient division of labor more broadly) is better knowledge of the associations between phonology and semantics, in line with recent computational work (Chang & Monaghan, 2019; Chang et al., 2019). At the same time, oral language skills did not account for variability in O-P fully, and an open question therefore is what additional mechanism(s) give rise to these individual differences. One possibility is that these differences are related to the effectiveness of the mechanism that tracks the statistical relationships between orthographic and phonological units. That is, perhaps readers who are less impacted by O-P regularities are less adept at forming associations between well-established orthographic and phonological representations based on the reliability of these relationships. Alternatively, it is possible that these readers' failure in forming these associations is the result of a deficiency in the orthographic or phonological representations themselves – for example, unstable phonological representations, resulting in poor associations with orthographic representations. These two possibilities reflect the principles of computational models of reading (Harm & Seidenberg, 1999; Harm, McCandliss, & Seidenberg, 2003; Rueckl et al., 2019), which suggest that differences in the quality of O-P relations are related to both learning-related parameters (e.g. learning rate, number of hidden units) and parameters related to the quality of the phonological representations (e.g. phonological noise). These two possibilities also relate to the model of SL by Frost and colleagues, which posits that individual differences in a given SL task are the result of the interplay between individuals' ability to compute regularities between established representations, and their encoding abilities that enable the efficient formation of these representations to begin with (Frost, Armstrong, Siegelman, & Christiansen, 2015). Also of note is that given the presumed involvement of SL in word segmentation (Saffran, Aslin, & Newport, 1996), impaired phonological representations may also be related to SL deficits. To dissociate between these two possible explanations (and their relative contribution), future studies can consider measures that directly assess the quality of orthographic and phonological representations (e.g. phonological awareness), and further explore the relations between sensitivity to O-P regularities, quality of representations, oral language skills, and reading.

A second issue concerns the relationship between the attunement of

the reading system to statistical regularities and the manifestation of those regularities in behavior. From our perspective, reading behavior is jointly determined by both O-P and O-S regularities; the influence of either depends on how well both have been assimilated as well as how reliable they are in a particular task. In the context of word naming, the behavioral manifestation of O-P regularities (as assessed by the O-P regularity effect) can be taken as a reasonably direct index of the degree to which those regularities have been assimilated into the organization of the reading system. Thus, with regard to our experiments, the magnitude of the O-P regularity effect can be taken as an indicator of how well an individual had assimilated the statistical regularities of the O-P mapping. In contrast, as illustrated by the modulation of the imageability effect on word naming by factors such as word frequency and O-P consistency, the manifestation of semantic regularities in this task is dependent on whether other regularities (in particular O-P regularities) serve as an efficient source of information to guide word reading. Thus, although differences in the magnitude of the imageability effect clearly reveal differences in the overall organization of the reading system, they should not be taken as a (context-free) measure of the strength of an individual's O-S (and S-P) knowledge. Rather, an individual's imageability effect is a proxy of the use of unreliable information despite the availability of more reliable information. Indeed, although we observed that poorer readers exhibited *larger* imageability effects, there is reason to believe that poorer readers are *less* adept at assimilating O-S regularities in general (see Keenan & Betjemann, 2008). Going forward, it will be vital to develop methods that allow us to assess how well O-S regularities have been assimilated across different reading tasks, to identify the factors related to individual differences in the ability to assimilate them, and to determine whether these are the same factors underlying variability in the assimilation of O-P regularities as well (see Sawi & Rueckl, 2019, for discussion).

Beyond these specific issues, we believe that our study has broader implications for the study of reading, as it provides a glimpse of an important but understudied aspect of literacy acquisition. More research is needed not only to gain a more complete picture of the division of labor between O-P regularities and imageability across the course of literacy acquisition, but also to reveal the role of other regularities (e.g. morphological regularities, O-P regularities at different grain sizes) over this same time course. In addition, whereas the experiments reported here focused on the results of long-term SL—that is, the behavioral consequence of learning that took place prior to the experiment—it is critical that future research target the learning mechanisms themselves. In this regard, it is interesting to note that recent findings suggest that O-P regularities and imageability play a facilitatory role in word learning, a role that further varies as a function of individual-level reading skill (e.g., Duff & Hulme, 2012; Steacy & Compton, 2019).

Finally, our study has implications for the study of SL more broadly. It is useful to consider the differences between the results of the current study and those of previous studies examining the correlations between SL tasks and reading skills. Of particular note is the strong predictive power of the measures examining reliance on O-P and O-S compared to the much weaker correlations observed in correlational studies of 'typical' SL tasks and reading outcomes. Some of the reasons for this difference are methodological. Specifically, SL tasks are oftentimes characterized by compromised reliability, especially with children (Arnon, 2019a; West, Vadillo, Shanks, & Hulme, 2018), which may limits the tasks' predictive power. Importantly, however, we believe that this discrepancy also originates from theoretical reasons. First, there is growing evidence that SL is not (or at least not exclusively) a domain-general construct, but rather shows patterns of modality and information-specificity (e.g. Arciuli, 2017; Siegelman et al., 2017). Per this domain-specific view of SL, learning one type of regularity (e.g. learning the transitional probability between two shapes) may be different than learning another (e.g. an association between a letter and a sound). Second, and relatedly, our results highlight the different nature

of learning of regularities in 'typical' SL tasks and those that are actually assimilated by learners in real-world tasks (e.g. reading). Namely, typical SL studies focus on constrained, highly (if not fully) regular, and uni-modal statistical contingencies that are learned over a short period of time. In contrast, the current approach examines reliance on more complex types of regularities, which vary in their reliability, that are concurrently presented to learners, and that are assimilated over long time-scales. Our results suggest that these more complex regularities are those that play a role in reading acquisition, more so than the simplified regularities typically studied in SL paradigms. We therefore believe that in order for SL theory to scale-up to the actual challenges presented by real-world learning tasks, it must go beyond simplified learning scenarios, and focus on these types of more complex and subtle regularities.

Acknowledgements

Supplementary Material

The full list of items (with their order of presentation, frequency, imageability, O-P surprisal values, and mean accuracy levels in Studies 1 and 2) is available at: https://osf.io/hkpy2/.

References

Arciuli, J. (2017). The multi-component nature of statistical learning. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 372*(1711), 20160058. https://doi.org/10.1098/rstb.2016.0058.

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.

Arciuli, J., & Simpson, I. C. (2012). Statistical learning is related to reading ability in children and adults. *Cognitive Science, 36*(2), 286–304. https://doi.org/10.1111/j.1551-6709.2011.01200.x.

Arnon, I. (2019a). Do current statistical learning tasks capture stable individual differences in children? An investigation of task reliability across modality. *Behavior Research Methods*. https://doi.org/10.3758/s13428-019-01205-5.

Arnon, I. (2019b). Statistical learning, implicit learning, and first language acquisition: A critical evaluation of two developmental predictions. *Topics in Cognitive Science, 11*(3), 12428. https://doi.org/10.1111/tops.12428.

Arrington, C. N., Malins, J. G., Winter, R., Mencl, W. E., Pugh, K. R., & Morris, R. (2019). Examining individual differences in reading and attentional control networks utilizing an oddball fMRI task. *Developmental Cognitive Neuroscience, 38*, 100674. https://doi.org/10.1016/j.dcn.2019.100674.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*(4), 321–324. https://doi.org/10.1111/1467-9280.00063.

Barakat, B. K., Seitz, A. R., & Shams, L. (2013). The effect of statistical learning on internal stimulus representations: Predictable items are enhanced even when not predicted. *Cognition, 129*(2), 205–211. https://doi.org/10.1016/j.cognition.2013.07.003.

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. Retrieved from http://cran.r-project.org/package=lme4.

Borleffs, E., Maassen, B. A. M., Lyytinen, H., & Zwarts, F. (2017). Measuring orthographic transparency and morphological-syllabic complexity in alphabetic orthographies: A narrative review. *Reading and Writing, 30*(8), 1617–1638. https://doi.org/10.1007/s11145-017-9741-5.

Borowsky, R., & Masson, M. E. J. (1996). Semantic ambiguity effects in word identification. *Journal of Experimental Psychology: Learning Memory and Cognition, 22*(1), 63–85. https://doi.org/10.1037/0278-7393.22.1.63.

Brown, P., Lupker, S. J., & Colombo, L. (1994). Interacting sources of information in word naming: A study of individual differences. *Journal of Experimental Psychology: Human Perception and Performance, 20*(3), 537–554. https://doi.org/10.1037/0096-1523.20.3.537.

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.

Chang, Y. N., Monaghan, P., & Welbourne, S. (2019). A computational model of reading across development: Effects of literacy onset on language processing. *Journal of Memory and Language, 108*, 104025. https://doi.org/10.1016/j.jml.2019.05.003.

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.

Coltheart, V., Laxon, V. J., & Keating, C. (1988). Effects of word imageability and age of acquisition on children's reading. *British Journal of Psychology, 79*(1), 1–12. https://doi.org/10.1111/j.2044-8295.1988.tb02270.x.

Cortese, M. J., & Simpson, G. (2000). Regularity effects in word naming: What are they? *Memory & Cognition, 28*(8), 1269–1276. https://doi.org/10.3758/bf03211827.

Dale, R., Duran, N. D., & Morehead, J. R. (2012). Prediction during statistical learning, and implications for the implicit/explicit divide. *Advances in Cognitive Psychology, 8*(2), 196–209. https://doi.org/10.2478/v10053-008-0115-z.

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.

Dotsch, R., Hassin, R. R., & Todorov, A. (2017). Statistical learning shapes face evaluation. *Nature Human Behaviour, 1*, 1. https://doi.org/10.1038/s41562-016-0001.

Duff, F. J., & Hulme, C. (2012). The role of children's phonological and semantic knowledge in learning to read words. *Scientific Studies of Reading, 16*(6), 504–525. https://doi.org/10.1080/10888438.2011.598199.

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.

Elleman, A. M., Steacy, L. M., & Compton, D. L. (2019). The role of statistical learning in word reading and spelling development: More questions than answers. *Scientific Studies of Reading, 23*, 1–7. https://doi.org/10.1080/10888438.2018.1549045.

Erickson, L. C., & Thiessen, E. D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review, 37*, 66–108.

Farnia, F., & Geva, E. (2011). Cognitive correlates of vocabulary growth in English language learners. *Applied Psycholinguistics, 32*, 711–738. https://doi.org/10.1017/S0142716411000038.

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.

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

Frost, R., Armstrong, B. C., & Christiansen, M. H. (2019). Statistical learning research: A critical review and possible new directions. *Psychological Bulletin, 145*(12), 1128–1153. https://doi.org/10.1037/bul0000210.

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences, 19*(3), 117–125. https://doi.org/10.1016/j.tics.2014.12.010.

Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What predicts successful literacy acquisition in a second language? *Psychological Science, 24*(7), 1243–1252. https://doi.org/10.1177/0956797612472207.

Gabay, Y., Thiessen, E. D., & Holt, L. L. (2015). Impaired statistical learning in developmental dyslexia. *Journal of Speech, Language, and Hearing Research, 58*, 934–945. https://doi.org/10.1044/2015.

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.

Grainger, J., Colé, P., & Segui, J. (1991). Masked morphological priming in visual word recognition. *Journal of Memory and Language, 30*, 370–384. https://doi.org/10.1016/0749-596X(91)90042-I.

Graves, W. W., Binder, J. R., Desai, R. H., Humphries, C., Stengel, B. C., & Seidenberg, M. S. (2014). Anatomy is strategy: Skilled reading differences associated with structural connectivity differences in the reading network. *Brain and Language, 133*, 1–13. https://doi.org/10.1016/j.bandl.2014.03.005.

Harm, M. W., McCandliss, B. D., & Seidenberg, M. S. (2003). Modeling the successes and failures of interventions for disabled readers. *Scientific Studies of Reading, 7*(2), 155–182. https://doi.org/10.1207/s1532799xssr0702_3.

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review, 106*, 491–528. https://doi.org/10.1037/0033-295X.106.3.491.

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.

Hoffman, P., Lambon Ralph, M. A., & Woollams, A. M. (2015). Triangulation of the neurocomputational architecture underpinning reading aloud. *Proceedings of the National Academy of Sciences, 112*(28), E3719–E3728. https://doi.org/10.1073/pnas.1502032112.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007.

Jared, D. (2002). Spelling-sound consistency and regularity effects in word naming.

Journal of Memory and Language, 46(4), 723–750. https://doi.org/10.1006/jmla. 2001.2827.

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.

Kahta, S., & Schiff, R. (2019). Deficits in statistical leaning of auditory sequences among adults with dyslexia. *Dyslexia, 25*(2), 142–157. https://doi.org/10.1002/dys.1618.

Kearns, D. M., Rogers, H. J., Koriakin, T., & Al Ghanem, R. (2016). Semantic and phonological ability to adjust recoding: A unique correlate of word reading skill? *Scientific Studies of Reading, 20*(6), 455–470. https://doi.org/10.1080/10888438. 2016.1217865.

Keenan, J. M., & Betjemann, R. S. (2008). Comprehension of single words: The role of semantics in word identification and reading disability. In E. L. Grigorenko, & A. J. Naples (Eds.). *Single-word reading: Behavioral and biological perspectives* (pp. 191–209). Denver, CO: Lawrence Erlbaum Associates Publishers.

Kuperman, V., & Van Dyke, J. A. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology: Human Perception and Performance, 39*(3), 802–823. https://doi.org/10.1037/a0030859.

Laing, E., & Hulme, C. (1999). Phonological and semantic processes influence beginning readers' ability to learn to read words. *Journal of Experimental Child Psychology, 73*, 183–207. https://doi.org/10.1006/jecp.1999.2500.

Long, J. A. (2019). Interactions: Comprehensive, user-friendly toolkit for probing interactions (R package version 1.1.0). Retrieved from 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.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*(3), 101–111. https://doi.org/10.1016/S0010-0277(01)00157-3.

McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews. Neuroscience, 4*(4), 310–322. https://doi.org/10.1038/nrn1076.

Monaghan, P., Chater, N., & Christiansen, M. H. (2005). The differential role of phonological and distributional cues in grammatical categorisation. *Cognition, 96*(2), 143–182. https://doi.org/10.1016/j.cognition.2004.09.001.

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.

Nation, K., & Snowling, M. J. (2004). Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of Research in Reading, 27*(4), 342–356. https://doi.org/10.1111/j.1467-9817.2004.00238.x.

Newell, K. M. (1991). Motor skill acquisition. *Annual Review of Psychology, 42*, 213–237. https://doi.org/10.1146/annurev.ps.42.020191.001241.

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.

Perfetti, C. A., Beck, I., Bell, L. C., & Hughes, C. (1987). Phonemic knowledge and learning to read are reciprocal: A longitudinal study of first grade children. *Merrill-Palmer Quarterly, 33*, 283–319.

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.

Qi, Z., Sanchez Araujo, Y., Georgan, W. C., Gabrieli, J. D. E., & Arciuli, J. (2019). Hearing matters more than seeing: A cross-modality study of statistical learning and reading ability. *Scientific Studies of Reading, 23*(1), 101–115. https://doi.org/10.1080/10888438.2018.1485680.

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.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews. Cognitive Science, 1*, 906–914. https://doi.org/10.1002/wcs.78.

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., Zevin, J. D., & Wolf, H., VII (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*. Baltimore, MD: Brookes Publishing Co.

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. In J. L. McClelland, & D. E. Rumelhart (Vol. Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2*, (pp. 216–271). Cambridge, MA: MIT Press.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294. 1926.

Salimpoor, V. N., Zald, D. H., Zatorre, R. J., Dagher, A., & McIntosh, A. R. (2015). Predictions and the brain: How musical sounds become rewarding. *Trends in Cognitive Sciences, 19*(2), 86–91. https://doi.org/10.1016/j.tics.2014.12.001.

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.

Scarborough, H. S. (1998). Predicting the future achievement of second graders with reading disabilities: Contributions of phonemic awareness, verbal memory, rapid naming, and IQ. *Annals of Dyslexia, 48*(1), 115–136. https://doi.org/10.1007/s11881-998-0006-5.

Schilling, H. E. H., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory and Cognition, 26*(6), 1270–1281. https://doi.org/10.3758/BF03201199.

Schmalz, X., Altoè, G., & Mulatti, C. (2017). Statistical learning and dyslexia: A systematic review. *Annals of Dyslexia, 67*(2), 147–162. https://doi.org/10.1007/s11881-016-0136-0.

Schmalz, X., Moll, K., Mulatti, C., & Schulte-Körne, G. (2019). Is statistical learning ability related to reading ability, and if so, why? *Scientific Studies of Reading, 23*(1), 64–76. https://doi.org/10.1080/10888438.2018.1482304.

Schwanenflugel, P. J., & Noyes, C. R. (1996). Context availability and the development of word reading skill. *Journal of Literacy Research, 28*(1), 35–54. https://doi.org/10.1080/10862969609547909.

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.). *yslexia across languages: Orthography and the brain–gene–behavior link* (pp. 146–168). Baltimore, MD: 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.

Shankweiler, D., Lundquist, E., Katz, L., Stuebing, K. K., Fletcher, J. M., Brady, S., ... Shaywitz, B. A. (1999). Comprehension and decoding: Patterns of association in children with reading difficulties. *Scientific Studies of Reading, 3*(1), 69–94. https://doi.org/10.1207/s1532799xssr0301_4.

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., 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), https://doi.org/10.1098/rstb.2016.0059.

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language, 81*, 105–120.

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, 1–21*. https://doi.org/10.3758/s13428-019-01317-y.

Sigurdardottir, H. M., Danielsdottir, H. B., Gudmundsdottir, M., Hjartarson, K. H., Thorarinsdottir, E. A., & Kristjánsson, Á. (2017). Problems with visual statistical learning in developmental dyslexia. *Scientific Reports, 7*(1), 606. https://doi.org/10.1038/s41598-017-00554-5.

Spencer, M., Kaschak, M. P., Jones, J. L., & Lonigan, C. J. (2014). Statistical learning is related to early literacy-related skills. *Reading and Writing, 28*(4), 467–490. https://doi.org/10.1007/s11145-014-9533-0.

Steacy, L. M., & Compton, D. L. (2019). Examining the role of imageability and regularity in word reading accuracy and learning efficiency among first and second graders at risk for reading disabilities. *Journal of Experimental Child Psychology, 178*, 226–250. https://doi.org/10.1016/j.jecp.2018.09.007.

Steacy, L. M., Compton, D. L., Petscher, Y., Elliott, J. D., Smith, K., Rueckl, J. G., ... 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.

Torkildsen, J.v. K., Arciuli, J., & Wie, O. B. (2019). Individual differences in statistical learning predict children's reading ability in a semi-transparent orthography. *Learning and Individual Differences, 69*, 60–68. https://doi.org/10.1016/J.LINDIF. 2018.11.003.

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., 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., Harvey, H., Aronoff, M., & Rastle, K. (2018). Skilled readers' sensitivity to meaningful regularities in English writing. *Cognition*. https://doi.org/10.1016/j.cognition.2018.09.013.

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry and Allied Disciplines, 45*(1), 2–40. https://doi.org/10.1046/j.0021-9630.2003.00305.x.

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence, Second Edition (WASI-II).*

San Antonio, TX: NCS Pearson.

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.

West, G., Vadillo, M. A., Shanks, D. R., & Hulme, C. (2018). The procedural learning deficit hypothesis of language learning disorders: We see some problems. *Developmental Science, 21*(2), https://doi.org/10.1111/desc.12552.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society. Series B: Statistical Methodology, 73*(1), 3–36. https://doi.org/10.1111/j.1467-9868.2010.00749.x.

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

Yap, M. J., Balota, D. A., Sibley, D. E., & Ratcliff, R. (2012). Individual differences in visual word recognition: Insights from the English Lexicon Project. *Journal of Experimental Psychology: Human Perception and Performance, 38*(1), 53–79. https://doi.org/10.1037/a0024177.

Zeno, S., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide.* Brewster, NY: 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.