Title: The contributions of decoding skill and lexical knowledge to the development of irregular word reading

Running title: Lexical knowledge and irregular word reading

Authors: Hannah M Nash [a] *, Robert Davies [b] & Jessie Ricketts[c]

[a] School of Psychology, University of Leeds, Leeds, LS2 9JT, UK

Tel: (+44) 113 3439199. E-mail: h.nash@leeds.ac.uk

http://orcid.org/0000-0002-4357-945X

* Corresponding author

[b] Department of Psychology, Lancaster University, Lancaster, UK

[c] Department of Psychology, Royal Holloway University of London, UK

The authors have no conflicts of interest to declare.

Acknowledgements: We would like to thank the research assistants who were involved in collecting the data for the studies reported here, and of course the children who took part. The first author was supported by a postdoctoral fellowship from The Division of Psychology and Language Sciences, University College London. The last author was supported by the Economic and Social Research Council (grant number ES/K008064/1). All analyses were run using the High End Computing facility at Lancaster University. We share data and code at https://osf.io/vcb83/?view_only=14609960061249b5a05ed0166feff743

The contributions of decoding skill and lexical knowledge to the development of irregular word reading

Research Highlights

- Combining results from observations of individual differences with results from training experiments, we show that reading irregular words in isolation is achieved through a combination of partial decoding and lexical knowledge.
- Familiarizing children with the phonological form of a word appears to have greater impact on learning to read it aloud than pre-training on the meaning of the word.
- Our findings expose limitations in current theoretical accounts of reading development, highlighting areas for further research.
- We recommend an instructional approach combining phonics and vocabulary training for novel irregular words.

Abstract

Two recent computational models of reading development propose that irregular words are read using a combination of decoding and lexical knowledge but differ in assumptions about how these sources of information interact, and about the relative importance of different aspects of lexical knowledge. We report developmental data that help to adjudicate these differences. Study One adopted a correlational approach to investigate the item-level relations between the ability to read a word aloud, general decoding ability, and knowledge of the word's phonological form (lexical phonology) or meaning (lexical semantics). We found that the latter three factors all influenced accuracy of oral reading. We observed trends indicating that the impact of differences in decoding skill and lexical knowledge were more prominent for irregular words. Study Two comprised two experiments in which novel irregular words were taught; in Experiment 1 we compared phonological to no pre-training, while in Experiment 2 we compared phonological to phonological plus semantics pre-training. Exposure to the phonological form of the word had a substantial impact in the early stages of learning, while the impact of adding semantics was more modest and emerged later. Our findings provide strong evidence that irregular words are read using a combination of decoding and lexical knowledge, with a greater contribution from lexical phonology than lexical semantics. Computational models of learning to read are currently unable to fully account for our data, therefore we propose some modifications. We advocate an instructional approach whereby phonics and vocabulary teaching are combined to support irregular word reading.

Keywords: reading, irregular words, decoding, vocabulary, lexical knowledge

English has an alphabetic orthography, with letters and letter patterns (graphemes) mapping to sounds (phonemes) in what can be seen to be a quasi-regular system. This means that while some grapheme-phoneme mappings follow consistent rules, some graphemes represent more than one phoneme (e.g., the letter 's' in sack and sugar), and some phonemes map to more than one grapheme (e.g., the sound /s/ in sack and cycle). Consequently, English contains regular words that conform to rules about how letters correspond to sounds and irregular words that do not. Once children have learned the rules governing spelling-sound correspondences, they can accurately decode (sound out) such regular words (e.g., bed, context) because the rule for pronouncing each grapheme yields the correct phoneme. In contrast, irregular words cannot readily be decoded because sounding out the graphemes will result in incorrect pronunciations (e.g., good, meringue). Such words present a challenge to the developing reader and to educators. An alternative way to identify the challenges presented by English orthography is in terms of spelling-sound consistency such that words with similar spellings and also sounds (in consistent neighbourhoods) will tend to be easier to read than words with similar spellings but different sounds (in inconsistent neighbourhoods) (Jared et al., 1990). At present, computational models of reading suggest that pre-existing knowledge about words is important for the development of irregular (or inconsistent) word reading. However, there is little consensus about which aspects of knowledge are important or, critically, when their influence may be observed. Here, we report data about emergent readers from two studies in which we used complementary methodologies to address the same questions: Which aspects of word knowledge contribute to irregular word reading and how does this knowledge interact with the product of decoding? Our findings have important implications for understanding how children learn to read and how best to deliver reading instruction.

We draw on computational accounts of reading development to generate expectations about the involvement of different kinds of word-level or lexical knowledge in reading aloud. We begin our discussion with two such accounts, the Self-Teaching Dual Route Cascaded model (ST-DRC; Pritchard et al., 2018) and the Self-Teaching Connectionist Dual Processing model (ST-CDP; Ziegler et al., 2014), that make an explicit link between decoding skill and learning to read words (self-teaching). We later discuss an alternative account, based on the so-called triangle model (Plaut et al., 1996; Seidenberg & McClelland, 1989), in which self-teaching is not implemented. In the self-teaching hypothesis, Share (1995) argued that children use their knowledge of letter-sound mappings to decode novel printed words and commit them to memory. In the case of irregular words, he argued that context supplements partial decoding attempts. The ST-DRC and ST-CDP accounts present important similarities, especially, in sharing the assumption that the phonological representations of words must be activated to trigger orthographic learning. However, the accounts differ in their assumptions about how candidate word forms are evoked, and about how and when lexical knowledge contributes to irregular word reading development.

The ST-DRC (Pritchard et al., 2018) incorporates self-teaching into a Dual Route Cascaded (DRC; Coltheart et al., 2001) framework. In the DRC, there are two main routes to reading aloud: going from the spellings of written words to their spoken forms. The 'sub-lexical route' comprises grapheme-phoneme conversion rules or correspondences (GPCs) that specify grapheme-phoneme mappings. The 'lexical route' comprises lexical (word-level) representations, and achieves spelling-to-sound conversion through two sub-routes: a lexical-non-semantic route that maps directly from orthographic to phonological forms (OP mappings), and a lexical-semantic route that maps from orthographic to phonological forms via semantics (OSP mappings). In the DRC, regular words can be read correctly through either route, but irregular words can only be read accurately through one of the lexical routes.

Like the DRC, the ST-DRC starts with built-in representation of GPCs and lexical semantics. However, the self-teaching part of the ST-DRC means that OP mappings are learned rather than built-in. For a regular word, grapheme decoding through the sub-lexical route activates the corresponding lexical phonological representation, resulting in the word being recognized and a new orthographic representation being created (i.e., orthographic learning). This is not possible for irregular words. The ST-DRC and DRC share the assumption of relatively high levels of phoneme-to-phonological-lexicon inhibition. This means that GPC-based activation alone cannot activate the corresponding lexical phonological representation above the threshold required for learning. The mismatch between regular versus irregular phonemes results in some inhibition. In the ST-DRC, irregular word learning is possible because support from contextual information (which, in theory, may be semantic or syntactic) evokes the activation of candidate lexical phonological representations. In the ST-DRC simulations, stronger levels of contextual support, and more specific (less ambiguous) contexts facilitated better learning.

It is possible that contextual or semantic information will also contribute to regular word reading development, especially in earlier phases of development (simulated in the ST-DRC by reduced GPC knowledge). But the importance of context in the ST-DRC implies that lexical semantic knowledge should influence reading development, principally, for irregular words. However, activation from the sub-lexical route remains important because the inclusion in simulations of even limited GPC rules lead to better learning than no rules, and more rules result in better learning. This implies that variation in decoding skill ought to have some influence in emergent reading for both regular and irregular words, with a more prominent role in reading regular words. Crucially, because the activation of lexical phonological representation is key to learning, variation in lexical phonological knowledge should influence the emergence of the capacity for, and the further development, of both regular and irregular word reading.

The ST-CDP (Ziegler et al., 2014) is a self-teaching version of the Connectionist Dual Processing model (CDP, Zorzi et al., 1998; Perry et al., 2007), a connectionist model that incorporates a non-lexical (two-layer associative network) process and a (non-semantic) lexical process. Like the ST-DRC, new orthographic representations are learned as a result of the activation of lexical phonological representations, when a child reads newly encountered word spellings. This implies, as in the ST-DRC account, that variation in lexical phonological knowledge should influence the development of regular and irregular word reading. In contrast to the ST-DRC model, the ST-CDP model assumes a relatively limited level of phoneme-to-phonological-lexicon inhibition. Granted lower levels of phoneme-phonology inhibition, encountering a novel word spelling evokes activation of a cohort of lexical phonological representations through sub-lexical spelling-sound mappings. In ST-CDP simulations (Ziegler et al., 2014), the model created orthographic representations for 80% of about 32,000 words but an explicit distinction was not made, in analyses, between Ziegler et al.'s more or less 'spelling-sound ambiguous' (or irregular) words in the training vocabulary. This means that the contributions of decoding and lexical phonological knowledge to words varying in spelling-sound ambiguity is unclear. If it is assumed that excitatory activation via sub-lexical mappings can activate phonological neighbours (including exception word pronunciations) then the ST-CDP account, like the ST-DRC account, implies that variation in decoding skill should influence development of both regular and irregular word reading.

Semantic knowledge was not built into the ST-CDP model, but it was assumed that a child could use contextual, syntactic or semantic information (including feedback from an adult during shared reading) to select the correct pronunciation from the cohort of candidates activated through decoding. Ziegler et al. (2014) do not report how the impact of semantics may differ in learning words varying in spelling-sound ambiguity; however, their results indicate that, overall, the correct phonological representations tended to be most activated most of the time in any cohort. Supposing that this is most likely to be true where, as for regular words, the correct pronunciation is entirely determined by sub-lexical mappings, then we can conjecture that semantic influence will tend to impact irregular word reading development. The ST-CDP account is silent on the time-course for the impact of semantic information on development. Key differences between the ST-DRC and ST-CDP are highlighted in Figure 1.

[Figure 1 about here]

Connectionist models based on the triangle model do not incorporate self-teaching in learning at present. Arguably, therefore, they may not apply to how children start to learn orthographic forms at the onset of reading development (Plaut et al., 1996; Seidenberg & McClelland, 1989; see also Kim et al., 2013; Armstrong et al., 2017). Connectionist triangle models vary, but all are governed by the principles of nonlinearity, adaptivity, and distributed representations (Plaut et al., 1996; Seidenberg & McClelland, 1989). They implement the assumption that the reading system operates over networks of sub-symbolic representations of orthography, phonology, and semantics, dispensing with explicit lexical representation (a key difference with the ST-DRC and ST-CDP accounts). They comprise systems that develop structure given exposure to the words in a training vocabulary and given a learning algorithm to allow adaptation in order to narrow the distance between output and target phonology. Exposure to a word will cause changes to the weights on network connections, and more frequent experience of that word will more often result in changes helpful to learning and producing it. However, the nonlinearity of the function linking input to output activation means that output activation will tend to asymptote toward the limits as input activation increases. This ensures the appearance of a gradual ceiling effect (Dilkina et al., 2008; Plaut et al., 1996; Seidenberg & McClelland, 1989).

The nonlinear input-output function means that room for improvement narrows as performance improves. Thus, the effect of regularity (or consistency) will be smaller for higher frequency words, and the effect of frequency will be smaller for regular (or consistent) words. Because the principles of nonlinearity and adaptivity apply generally, we can expect to find that as experience accumulates over a vocabulary, and performance improves overall, effects of frequency or consistency will tend to diminish, and, indeed, this is what the simulations show (Seidenberg & McClelland, 1989), consistent with studies of lifespan reading development (e.g., Davies et al., 2017). Connectionist accounts of reading development, then, imply the expectation that the difference between reading regular and irregular words (or words varying in consistency) shall tend to decrease as children develop as readers. This may be reflected in a decreasing regularity effect for older children but may also be seen in children with better decoding skills. Hence, as for the ST-CDP and ST-DRC accounts, connectionist accounts imply the expectation that variation in decoding or phonological word form knowledge should influence reading development. The gradual ceiling effect implies that such influences should be more prominent for irregular word reading. While Seidenberg and McClelland (1989) demonstrated that a connectionist system can accurately read words aloud using just an OP mapping system, Plaut et al. (1996) showed that if semantics (or, strictly, an additional source) may contribute activation to output phonology then the system will learn a division of labour in which irregular exception words will tend to be read with the help of semantics. Given the development of a division of labour in a child (a possibility that may vary, Dilkina et al., 2008), we can expect that the impact of word meaning knowledge will also be most prominent in irregular word reading. We shall return to this point in motivating Study Two.

Behavioural data supports a role for both decoding and lexical knowledge in irregular word reading but there are important limitations in previous research. While the role of decoding in regular word reading is well established (Share, 1995), measures of individual differences in decoding skill also correlate with performance in tests of irregular word reading (e.g., Baron, 1979; Wagner & Torgesen, 1987). Longitudinal studies in primary school-aged children have shown that earlier variation in decoding and lexical knowledge predict later variation in irregular word reading (Nation & Snowling, 2004; Ricketts et al., 2007). Two studies furnish observations relevant to the role of specific aspects of word knowledge in reading aloud. Nation and Cocksey (2009) examined the ability of 7-year-old children to read a word in relation to their ability to respond to the same word in auditory lexical decision and definitions tasks that tapped, respectively, knowledge of phonological forms and meanings. Word knowledge was correlated with word reading at the item level and this association was stronger for irregular words. However, once variance associated with word phonology had been accounted for, semantic knowledge did not explain additional unique variance. Nation and Cocksey (2009) concluded that only lexical phonology contributed to irregular word reading in children. However, they were not able to investigate the predictive relationship for regular words due to ceiling effects. Using a similar design, Ricketts et al. (2016) found that knowledge about a particular word's semantics, but not its phonology, predicted both regular and irregular word reading in 6-year-old children. These results suggest, potentially, that there may be separable impacts on reading aloud, in developing readers, due to individual differences in word knowledge. However, it is unclear why the findings are inconsistent and thus what role different aspects of word knowledge may play. Moreover, effects of knowledge are reported but analyses do not take into account the effects of decoding skill.

In our first study, we examined children's responses to regular and irregular words (words varying in spelling-sound consistency). We sought to observe the effects of variation in children's knowledge about the phonological form as well as the meaning of words. Addressing the limitations of previous research, our analyses took into account individual differences in decoding skills. This is an observational study and therefore dependent, in its capacity to detect effects of interest, on the nature of the sample. Thus, we follow up the first with a second study, comprising a pair of training experiments in which we manipulated the knowledge that children can use in reading irregular words. We share data and code at https://osf.io/vcb83/?view_only=14609960061249b5a05ed0166feff743.

Study One

In Study One, we asked 7-8 and 9-10 year old children to read regular and irregular words. We presented relatively difficult words to mitigate ceiling effects. Prior to our analyses, given previous observations, we expected to see the effects of: word type, with regular words read more accurately; age, with older children reading more words accurately; decoding skill differences, with better decoders reading more words; and lexical knowledge, where words associated with some phonological or semantic knowledge would be more likely to be read correctly.

We predicted that the effects of age, decoding skill and lexical knowledge would interact with the effect of word type. However, the nature of the interactions we could predict depends on which computational account is adopted. Relatively clear predictions can be derived from the ST-DRC account but predictions based on the ST-CDP account are harder to construct because the model implementation, as it is reported, is ambiguous in critical respects. Lexical-semantic 'knowledge' was not embedded in the ST-CDP, and word type was not distinguished in analyses of performance though words of varying spelling-sound ambiguity were included among stimulus words. Both self-teaching accounts predict a greater effect of variation in decoding skill in regular than in irregular word reading because of the link between regular word reading and the relative effectiveness of the sub-lexical mapping (which decoding skill measures should reflect); this was indicated in simulations of the ST-DRC but not tested in the ST-CDP. Both accounts also predict an effect of variation in lexical phonological knowledge on regular as well as irregular word reading because activation of the correct spoken word is central to learning. According to the ST-DRC, the influence of lexical semantic knowledge should be greater in irregular word reading. In that account, lexical semantic knowledge should only influence regular word reading in less skilled decoders.

In comparison, the connectionist account of reading development implies the expectation that responses to regular words should be more accurate than responses to irregular (or inconsistent) words but that this difference should tend to decrease for older children (provided age is associated with reading experience), and for children presenting stronger OP mappings (reflected in higher levels of decoding skill). In addition, while connectionist models do not incorporate lexical representations, the corresponding theoretical accounts would lead us to expect that children with stronger knowledge about word forms (cf. Harm & Seidenberg, 1999) and stronger semantic information should also perform better overall. Given the expectation of a gradual ceiling effect, connectionist accounts would also appear to predict that the regularity effect should be smaller for more experienced readers, those with greater decoding skill or word knowledge or, equally, that the effects of experience, decoding skill or knowledge would be more prominent in responses to irregular than to regular words. While our initial stimulus selection was focused on variation in spelling-sound word regularity, to check the sensitivity of our results, and to align our analyses with the connectionist framework, in alternate analyses we replaced regularity with orthographic consistency.

Method

Ethics statement

Ethical approval was gained from the Departmental Ethics Committee at University College London. Written consent was gained from parents and verbal assent from children.

Participants

The sample comprised 33 younger (7-8 years, M_{age} = 7.70 years, *SD* = .35; 16 female) and 33 older (9-10 years, M_{age} = 9.62 years, *SD* = .30; 17 female) children. They completed the Test of Word Reading Efficiency (TOWRE) (Torgesen, Wagner & Rashotte, 1999), a standardized measure of reading in which they were asked to read as many words (Sight Word Efficiency subtest) and pseudowords (Phonemic Decoding Efficiency subtest) as they could in 45 seconds. Children obtained word (younger *M* = 114.12, *SD* = 13.58, older *M* = 109.58, *SD* = 11.86) and pseudoword (younger *M* = 112.15, SD = 24.81, older *M* = 112.36, *SD* = 13.21) standard scores towards the upper end of the normal range.

Procedure

Children completed three sessions of approximately twenty minutes, each separated by one week. Three experimental tasks were administered in a fixed order, to minimize contamination between tasks (cf. Nation & Cocksey, 2009). Auditory lexical decision was assessed first so performance was not enhanced as a consequence of hearing the words in other tasks. Definition knowledge was assessed last because this task was considered relatively immune to contamination from other tasks.

Items

Regular and irregular words were drawn from the Castles and Coltheart reading test 2 (CC2, Castles, et al., 2009), which includes 40 irregular words, 40 regular words and 40 pseudowords, the words were matched for frequency, length and grammatical class. Data from only 38 items of each type were used in our analyses, as data for two of the regular items (chicken and grail) were not collected from some children due to an administrative error. The corresponding irregular words (couple and crepe) were also removed prior to analysis. The pseudoword data were not analyzed.

Experimental tasks

Auditory Lexical Decision (phonological knowledge). The regular words, irregular words and pseudowords from the CC2 were mixed with 40 pseudowords from the Children's Test of Nonword Repetition (CNRep, Gathercole & Baddeley, 1996) to ensure a balanced number of words and pseudowords. These were digitally recorded by a native female speaker and presented through headphones in a fixed random order using OpenSesame (Mathot et al., 2012). Children were instructed to say 'yes' if they thought an item was a real word or 'no' if they thought it was not, with the accuracy of responses recorded by the tester. Auditory lexical decision could be aided by knowing the meaning of the word but word meaning knowledge was not necessary to responding accurately.

Reading. The regular and irregular words were presented on computer screens (using OpenSesame), one item at a time, in the order (of difficulty) in which they appear in the CC2 test. Each word was presented in 32-point Arial font and preceded by an 800-ms blank screen. The word remained on screen until it was read by the child. The tester recorded pronunciation accuracy.

Definitions (semantic knowledge). Each word was read aloud by the experimenter and the child was asked if they had heard the word before. If they said that they had, they were asked to provide a definition in response to the questions "what is a …?" or "what does … mean?". If, given a child's initial response, the experimenter suspected additional knowledge then they used the prompt cue "can you tell me anything more about this word?" to elicit a fuller production. Accuracy of each definition was scored using pre-determined definitions collated from dictionaries. If there was any ambiguity about whether a response was correct, the scoring was agreed after discussion.

Results

Performance on the three tasks in each age group, for regular and irregular words, is summarized in Table 1. It can be seen that older children read, recognized and defined more words correctly than younger children, regular word reading was more accurate than irregular word reading in both age groups, and responses to regular items approached maximum accuracy for many children in word reading. Auditory Lexical Decision scores were higher than Definitions scores, suggesting that children were familiar with the forms of some words for which they could not produce definitions.

[Table 1 about here]

In our analysis of the Study One data, as in all analyses in this paper, we fitted a Bayesian generalized linear mixed-effects model to response accuracy -- for Study One, reading response accuracy -- using the *brms* (Bayesian regression models using 'Stan') library (Burkner, 2018; Carpenter et al., 2017). We did this because, firstly, Bayesian models virtually always converge (Kruschke & Liddell, 2018) whereas frequentist models (e.g., fitted using lme4, Bates et al., 2015) can sometimes fail to converge given more complex model structures (Bates et al., 2018; Eager & Roy, 2017; Matuschek et al., 2017) even where structures are warranted by study designs (Barr et al., 2013; in terms of random effects, see our discussion, following). Secondly, Bayesian models are scientifically advantageous because they yield a posterior probability distribution of the model parameters (including the effect coefficients). Given our data, and granted our assumptions, the posterior distribution indicates the relative probability of different values of the coefficient for each effect. We report the most probable estimate for the coefficient for the fixed effect of each experimental variable. The spread of the posterior distribution directly indicates our uncertainty about the estimate. We report 95% Credible Interval (CrI) limits to summarize that uncertainty. (We describe our analysis in more detail, and present plots illustrating the posterior distribution for each effect in each model, in OSF: analysis-details_2020-09-27.pdf.) This approach enables an inferential focus on the direction of the effects of the experimental variables (e.g., Cumming, 2014). This focus is necessitated by the theoretical and practical importance of aiming to distinguishing *how* differences in decoding skill, age or lexical knowledge may benefit reading performance.

To enable readers to compare our results with those deriving from more traditional methods, we fitted frequentist models with the same structure using the lme4 library (Bates et al., 2015; see OSF: lme4-analyses_2020-09-27.pdf.) While the effects estimates are largely comparable, we should note that some of the frequentist models are associated with convergence warnings, rendering their interpretation problematic. Encountering convergence warnings is not altogether surprising, in the context of mixed-effects analyses of response accuracy (Bates et al., 2018; Eager & Roy, 2017; Matuschek et al., 2017), and our anticipation of such difficulties was one motivation, as noted, for opting to analyze outcomes using Bayesian methods.

The model for our analysis of Study One data estimated the fixed effects on reading accuracy associated with: individual differences in age (year group, years 3 versus 5); decoding skill (standardized raw scores from the TOWRE Phonemic Decoding Efficiency subtest); whether the child's response to the same word was correct or not in the Definitions or Auditory Lexical Decision tasks; the type of word (regular vs. irregular); as well as interactions between the effects of age or decoding skill, lexical knowledge (in Definitions, Auditory Lexical Decision), and word type, including 2- and 3-way interactions. Different levels of the categorical factors (year group; Definitions or Auditory Lexical Decision accuracy) were sum coded (-1, +1). The model incorporated random effects due to unexplained deviations between sampled participants or words in intercepts, between participants in the effect of word type, and between words in the effects of year group or decoding skill. A summary of the model estimates for the fixed effects is presented in Table 2 (Random effects variance and covariance estimates are presented in OSF: analysis-details_2020-09-27.pdf).

[Table 2 about here]

We found evidence for an effect of individual differences in decoding skill (estimated effect = .95, CrI[.69, 1.23]): increasing decoding skill was associated with increasing probability that a word would be read correctly. There were positive effects of better performance in both lexical knowledge tasks, Definitions (.70, CrI[.48, .92]) and Auditory Lexical Decision (.32, CrI[.14, .50]), suggesting that children who are able to correctly define or classify a word then are also more likely to read that word correctly. Finally, we observed an effect of word type (1.92, CrI[1.28, 2.54]): regular words were more likely to be read correctly than irregular words.

For all other effects, the evidence we observed was weak. To explain what this means, we note that, whereas credible intervals may resemble confidence intervals numerically, the Credible Interval (CrI) directly indicates our uncertainty about the estimated coefficient of each effect while confidence intervals do not (Morey et al., 2015): for any CrI, we see the range of plausible coefficients for an effect, with upper and lower limits within which we may be certain with 95% probability that (given our data and model) the true parameter lies. (A 95% confidence interval for a parameter may be expected to include the true value of the parameter 95% of the time, over repeated sampling, in hypothetical studies.) The credible intervals associated with the effect of year group, and with the effects of all hypothesized interactions, each encompassed 0. This indicates that, for example, the estimate of the effect of the interaction between word type and decoding skill is compatible, with some probability, with negative, null, or positive estimates of the coefficient of the effect of the interaction. Our data were insufficient to resolve the size or the direction of the interaction effects with precision.

While our model indicates weak evidence for the hypothesized interactions, it is useful, nevertheless, to examine the marginal effects plots for these interactions (Figure 2). This is because the interactions are critical to our theoretical concerns, especially as they relate to the potential for different impacts of decoding skill, age, or lexical knowledge for responses to regular or irregular words. And it is because, as can be seen in Table 2 (and in the plots shown in OSF: analysis-details_2020-09-27.pdf), substantial proportions of the probability distribution are associated with non-zero effects of some interactions, indicating some support for the interactions, even if there is insufficient information to estimate their effects with precision. In the following, we report *how much* probability is associated with positive or negative interaction effects, i.e., for each interaction, what proportion of the posterior probability distribution is associated with an effect.

For each plot in Figure 2, it is apparent that the probability that a response to a regular word would be correct approached 1.0, irrespective of year group, differences in decoding skill, or performance in the Definitions or Auditory Lexical Decision tasks. Variation in accuracy was apparent, in contrast, in responses to irregular words. The interaction between word type and year group (Figure 2 a.; estimated effect = -.21, CrI[-.48, .05]; proportion of posterior probability distribution < 0 equals 94%) suggests that year 3 responses tended to be less accurate than year 5 responses to irregular words. The interaction between word type and decoding skill (Figure 2 b.; estimated effect = .21, CrI[-.05, .48]; proportion of probability distribution > 0 equals 94%) suggests that children who scored higher on decoding skill measures were more likely to read irregular words correctly. A small effect of individual differences in decoding skill is apparent, also, for responses to regular words but appears confined to children scoring at lower levels on measures of decoding. The interaction between word type and Definitions performance (Figure 2 c.; estimated effect = .12, CrI[-.10, .33]; proportion of posterior probability distribution > 0 equals 86%) suggests that children who could correctly define an irregular word were more likely to read it correctly. The interaction between word type and Auditory Lexical Decision performance (Figure 2 d.; estimated effect = -.10, CrI[-.28, .08]; proportion of posterior probability distribution > 0 equals 85%) suggested that children who correctly classified an irregular word were more likely to read it correctly.

[Figure 2 about here]

Consistency rather than regularity. Spelling-sound regularity is not critical to connectionist accounts of oral reading because, in the latter, the quasi-regular nature of the orthography-to-phonology mapping is captured in terms of differences in spelling-sound consistency. To test predictions deriving from connectionist accounts, therefore, we conducted additional analyses of Study One oral reading accuracy in which we fitted models that were the same in structure except that we replaced word type with the rime consistency of stimulus words. In alternate analyses, we fitted models using either word type or token (feedforward, i.e., spelling-to-sound) rime consistency (Chee et al., 2020), with consistency coded as continuously varying quantity rather than a categorical dichotomy (as in regular versus irregular word type). Words associated with higher levels of spelling-sound consistency were more likely to elicit correct reading responses. Substituting regularity with consistency did not produce different results besides, obviously, replacing a regularity with a consistency effect. Using consistency instead of regularity did not change the shape of the effects of decoding skill or performance in lexical tasks and did not change the interactions involving the impact of spelling-sound ambiguity (see OSF: sensitivity-analyses_2020-09-27.pdf).

Sensitivity checks. We examined the sensitivity of our results to variation in the way in which we coded critical variables or specified models (Depaoli & van de Schoot, 2017). We checked if our estimates depended on the specification of prior probability distributions. Models incorporating narrower priors were associated with slightly smaller effects estimates to those we report, while models incorporating more diffuse priors were associated with practically identical estimates. We examined, also, the impact of coding participant age not by year group (as a factor) but in months. This appeared, again, to have little impact on results. (These checks are outlined in OSF: sensitivity-analyses_2020-09-27.pdf.) To permit a comparison between Bayesian and frequentist estimates of effects, we fitted all models using the lme4 library (Bates et al., 2015). Estimated effects were, in some cases, slightly larger for the frequentist models but the pattern of results was otherwise the same (see OSF: lme4-analyses_2020-09-27.pdf.)

Study One Summary and Discussion

Whether we construe the predictability of orthography-to-phonology mappings in terms of regularity or consistency, children were more likely to read regular (or consistent) words more accurately than irregular (or inconsistent) words. Our observation of evidence for the impact of decoding ability is congruent with the predictions that derive from self-teaching as well as connectionist accounts of reading development. Importantly, we found evidence for distinct effects of lexical phonology and semantic knowledge while taking into account the effect, also, of decoding skill. We found limited evidence for specific effects of the hypothesized interactions. However, while our model suggested we cannot estimate the interactions with certainty, trends indicated that the impact of differences in age, decoding skill or lexical knowledge were more prominent for irregular words.

Our observation that response accuracy to regular words was at ceiling for our sample implies, interestingly, that younger children possessed knowledge sufficient to decode the stimuli. But it constrains the conclusions we can draw about the contributions of decoding skill or lexical knowledge to reading aloud different word types. It may be, as the plots of the interactions (Figure 2) clearly suggest, that the effects of age, decoding skill or lexical knowledge are stronger in determining the accuracy of responses to irregular words. These observations would be consistent with the expectations that derive from the assumption of a nonlinear activation function in connectionist accounts, and the corollary expectation of a gradual ceiling effect for influences on reading performance. Our finding that the effect of decoding skill was not stronger for regular word reading would appear to be inconsistent with the predictions deriving from self-teaching accounts of reading development. However, it may also be, as the uncertainty about the coefficients for the interactions suggests, that the population effects of individual differences in decoding skill or lexical knowledge would be similar for both regular and irregular words. Further investigation is required because the limits in our evidence mean that its theoretical significance is yet unclear regarding the influences on regular word reading. Given the young age of our youngest readers, and the relative difficulty of the chosen stimuli, we think the investigation of lexical effects on regular word reading presents an important empirical challenge. But we focus, next, on the influences of lexical knowledge in reading irregular words because, as we discuss, computational accounts of development differ most obviously in their predictions concern irregular word reading.

Behavioural studies have yielded contradictory results concerning the role of lexical knowledge. Nation and Cocksey (2009) found that lexical phonology but not semantics predicted irregular word reading. Ricketts et al. (2016) found that lexical semantics but not phonology contributed to reading both regular and irregular words. The results of Study One show that both types of lexical knowledge have an effect and, further, suggest that the effects of variation in lexical knowledge are more marked for irregular compared to regular words. The evidence from Study One was grounded in an observational sample drawing on variation in knowledge about words that were, to varying extents, already known to the children we tested. That approach had the advantage of assessing the impact of knowledge as it is found among emerging readers but, clearly, it constrained our capacity to identify the effect of knowledge differences given the lack of control over prior learning. This motivated the investigation of the effects of the knowledge that can be established through a learning intervention. This is what we report next.

Study Two: Training Experiments with Irregular Words

In Study Two, we conducted two training experiments to test the causal role of pre-training lexical knowledge in learning to read irregular words. The impact of training lexical knowledge prior to exposure to written words has been investigated in a number of studies involving adults or children. McKay et al. (2008) found that when adults received pre-training on both lexical-phonological and lexical-semantic information they were better able to read inconsistent novel words compared to when they received lexical-phonological pre-training alone. Taylor et al. (2011) replicated and extended this finding in a study using inconsistent pseudowords. In the early stages of learning, they found that both types of pre-training led to significantly better reading compared to no pre-training but, at the end of learning, the phonology plus semantic condition led to significantly better reading of pseudowords. Studies involving children have, in comparison, provided incomplete information. In some studies that involved novel words with regular or consistent pronunciations only, reading accuracy was observed to be greater following phonological pre-training while semantic information did not provide additional benefit (Duff & Hulme, 2012; McKague et al., 2001). When studies have involved irregular words (Wang et al., 2011; Wang et al., 2013), the content of oral pre-training has not been manipulated to examine the independent contributions of phonological or semantic knowledge. It is unclear, therefore, how phonological or semantic knowledge contribute to the capacity to read irregular words aloud. We conducted our study to address this gap.

Computational models of reading development differ in key respects in their predictions concerning the effects of pre-training on the development of irregular word reading. The current instantiation of the ST-DRC (Pritchard et al., 2018) assumes a high level of phoneme-to-phonological lexicon inhibition that allows the model to share the capabilities of the DRC but, critically, prevents the activation of candidate phonological word forms based on decoding alone, when a child first encounters a novel irregular word spelling. This assumption means that the model relies on contextual support to enable the activation of potential phonological word forms, at least, at the onset of word learning (the word type learning phase, in the authors' terms). It implies a sharp distinction with the ST-CDP account because the ST-CDP model, in assuming a lower level of phoneme-to-phonological lexicon inhibition, does allow the activation of phonological word form representations, given decoding, in the absence of contextual support. According to the ST-CDP account, an effect of pre-training could be seen on the first trial of written word learning, as partial decoding of an irregular word in isolation can lead to sufficient activation for recognition. In our training Experiment 1, phonological pre-training is conducted without semantic or syntactic contextual support. Thus, our observations on the impacts of phonological pre-training speak critically to a key distinction between self-teaching accounts.

The ST-DRC predicts that the inclusion of semantic information in pre-training will provide additional benefit to learning because the lexical semantic route can contribute to the activation of phonemes corresponding to lexical phonological representations. If semantic pre-training can be constructed so that, effectively, it provides support equivalent to the contextual support implemented in the ST-DRC simulations, then it may be that the impact of semantic pre-training is observed on irregular word reading from the beginning of training. The ST-CDP account assumes that a child selects the correct word form from an activated cohort of candidates using contextual, semantic or syntactic constraints. Both self-teaching accounts imply that the influence of semantics should be seen from the start of training but, given existing information, we cannot say whether that influence should be sustained over time. In our training Experiment 2, phonological pre-training is conducted with versus without semantic contextual support. Thus, our observations speak critically, also, to assumptions about the impact and timing of semantic information in reading development in self-teaching accounts.

Connectionist (triangle model) accounts, also, imply the expectation that we should see benefits of pre-training, but it is less clear which aspect of lexical knowledge will have an impact. Recent simulation results show that variation in the strength or effectiveness of phonological representations

(Harm & Seidenberg, 1999) or in semantic information (Dilkina et al., 2008; Plaut et al., 1996) should affect the development of irregular word reading. However, the supervised nature of learning in these simulations, given the use of the back-propagation learning algorithm, arguably means that connectionist accounts cannot predict the self-taught learning that may occur when an irregular word spelling is first encountered (Pritchard et al., 2018). Clearly, there is no reason in principle why connectionist models could not be extended to incorporate self-teaching, and orthographic word learning, but the difference that pre-training makes, and when that difference can be observed, has important implications for whether an extension of connectionist models would be warranted. Further, Plaut et al. (1996) argued that the influence of semantic information should appear later in development because the orthography-to-semantics (OS) mapping can influence output phonology only once the OP mapping has developed to some degree. However, the emergence of OS or OSP mappings (rather than the optimization of connection weights) over time has not yet been simulated to our knowledge (cf. Harm & Seidenberg, 2004), thus our findings potentially supply evidence about the timing of the semantic contribution that would be important, also, to the development of connectionist accounts.

We adapted the methodology of McKay et al. (2008) and Taylor et al. (2011) for use with children. Children were trained on low frequency irregular words that they were unlikely to know so that they were tested on their capacity to learn, effectively, non-words: potentially extending their vocabulary. In Experiment 1, we examined the impact of the addition of phonological pre-training to a learning program that otherwise included just the repeated opportunity to read target words aloud. Phonological pre-training involved hearing and repeating the spoken form of each word and segmenting it into phonemes. We supposed that such experience would support the creation of phonological representations for the word. In Experiment 2, we compared the impact of phonological pre-training alone to the impact of the same phonological pre-training combined with semantic pre-training. The difference between these conditions was designed to isolate the effect of exposure to semantic information. We supposed that this addition would support the creation of semantic representations.

Experiment 1: Phonology pre-training versus no pre-training

Method

Ethics statement. In both experiments, ethical approval was gained from the Departmental Ethics Committee of the Psychology Department at the University of York. Written consent was gained from parents and verbal assent from the children.

Participants. Thirty-two children aged 7-9 years participated (M_{age} = 8.36 years, *SD* = 0.63; 16 female, 16 male). The children were not reported to have any recognized special educational needs and used British English as their primary language. The children's norm-referenced scores on standardized measures of language and literacy (see following) are presented in Table 3. These scores cover a wide range but the mean standard score on the spelling test is at the upper end of the average range.

[Table 3 about here]

Standardized Measures of Language and Literacy. Standardized measures of literacy and language were included to confirm that the sample was representative and to serve as filler tasks between training and post-tests. Tasks were administered according to their manual. Children completed Card One of the Single Word Reading Test (SWRT; Foster, 2007), the Green Spelling Test of the Wide Range Achievement Test 4 (WRAT4, Wilkinson, 2006) and the vocabulary subtest of the Wechsler Intelligence Scale for Children (WISC-IV-UK; Wechsler, 2003).

Materials. Twelve low frequency irregular, inconsistently spelled, nouns were selected from the MRC Psycholinguistic Database (Coltheart, 1981; see Table A1). These were divided into two lists of six words, with no significant difference between lists in in SUBTLEX-UK log frequency Zipf values (Van Heuven et al., 2014), length in letters, phonemes or syllables ($t(10) < 1$ in all comparisons). The words varied in position of grapheme-phoneme irregularities. The assignment of list to learning condition was counterbalanced.

Procedure. The experiment used a within-participants design, with children learning 12 words: six in each condition. The conditions were administered in separate 45-minute sessions, approximately one week apart, in the child's school. The order of administration of words (lists) in conditions was counterbalanced over participants. To assess learning, we conducted pre- and post-tests of phonological knowledge and reading aloud. Post-tests were given approximately 15 minutes after training. The training, pre- and post-tests were implemented using e-Prime (v.2, Schneider et al., 2002). Figure 3 summarizes the design and procedure for both experiments. Learning without pre-training incorporated just reading opportunities (reading-only condition), while learning with phonological pre-training incorporated reading opportunities as well as phonology-focused exercises (reading-plus-phonology condition).

[Figure 3 about here]

Training. ***Phonological pre-training.*** Children were presented with each of the six words aurally during two cycles of training. The children heard a female voice say the word and then segment it into individual phonemes. The children were asked to repeat the word (e.g., "The new word is amethyst, can you say amethyst?") and then to identify its first and last sound. They were then asked if they could repeat the word once more, after which the next word was introduced. Each word was presented once per cycle, in a random order. Feedback was provided for the first and last sounds and for the final repetition, irrespective of accuracy, to equate the number of exposures (e.g., "That's right it's /a/." or "That's not quite right it's /a/").

Reading-based training. There were three cycles of reading-based training. In each cycle, the children were told they were going to practise reading some words. They saw the written form of the word and were asked to read it aloud. Their response received corrective feedback (e.g., "Yes that's right it's amethyst." or "No that's not quite right it's amethyst.") and the next word was displayed.

Words were presented in random order for each cycle in each training condition.

Pre- and Post-tests. ***Phonological choice (pre- and post-test, both conditions).*** For the pre-test, all 12 words were presented in the choice task. Two pseudoword distracters were created for each word, with either one or two differing phonemes (e.g. Tzar, Zee, Zer; see Appendix). Children heard each of the three options accompanied by a number on the screen (1, 2, 3) and were then asked to identify the real word. Children received two practice trials followed by the experimental trials. The items were presented in a random order and the position of the target word was rotated. To measure phonological learning, the task was given again at post-test for the six words taught in that session.

Reading aloud (pre- and post-test, both conditions). For the pre-test, children were asked to read aloud 36 items, consisting of the 12 experimental words, 12 visually similar pseudowords and 12 other pseudowords. The words and pseudowords were presented in a fixed random order. To measure improvement in the children's ability to read the irregular words at the end of each training session, the children were asked to read aloud 18 words: the six items they had been taught in that session; their six visually similar counterparts; and six randomly assigned pseudowords.

Test accuracy was scored by awarding one point for each correct response.

Results

We analyzed phonological choice and reading response accuracy using models fitted to estimate the effects of training condition (reading-only or reading-plus-phonology), test time, and the interaction between the effects of training condition and test time. For phonological choice, the effect of test time was estimated as a comparison between pre- and post-test times. For reading aloud, it was estimated as a comparison of responses recorded at pre-test (time 1), at each of three training sessions (test times 2, 3 and 4), and at post-test (time 5). The models included random effects due to unexplained deviations between-participants or between-words in intercepts, or in the slopes of the effects of condition, test time, and the condition x test time interaction. We present a summary of the results in Table 4, illustrating the effects of condition and time on phonological choice (plot a.) and on reading accuracy (plot b.) in Figure 4.

[Table 4, Figure 4, about here]

For the analysis of phonological choice, we observed effects of: training condition (estimated effect = .39, CrI[.08, .72]), responses were more likely to be correct under the reading-plus-phonology condition; test time (1.47, CrI[1.06, 1.93]), responses were more likely to be correct at post- than pre-test; and the interaction (.51, CrI[.18, .87]), as the increase in accuracy from pre- to post-test was greater under the reading-plus-phonology condition.

For the analysis of reading aloud, we again found effects of: training condition (estimated effect = .96, CrI[.63, 1.30]), responses were more likely to be correct under the reading-plus-phonology condition; test time (1.54, CrI[1.17, 1.91]), responses were more likely to be correct at later test times; and the interaction (.26, CrI[< -0.01, .53]), such that the increase in accuracy from pre- to post-test was greater under the reading-plus-phonology condition.

Experiment 2: Phonology pre-training versus phonology plus semantics pre-training

Method

The design and items were the same as in Experiment 1 but we recruited a new sample of children. The procedure is summarized on the right-hand side of Figure 3. Phonological pre-training incorporated reading opportunities and phonological exercises (reading-plus-phonology condition), while the semantic pre-training condition incorporated reading opportunities as well as phonological and semantic exercises (reading-phonology-semantics condition).

Participants. Thirty-four children aged 7-9 years took part (M_{age} = 8.25 years, *SD* = 0.41; 16 female, 18 male). The children's norm referenced scores on the standardized measures of language and literacy are presented in Table 3.

Training. Phonological (reading-plus-phonology) pre- training took the same form as in Experiment 1.

Semantic pre-training. Semantic training directly followed phonological training for each word. To balance the number of exposures to the word, the children did not hear the final request to repeat the word. Instead, they were told that they would learn more about the word and they then heard it once more within the definition. Children were presented with a picture and heard a short description detailing what it was, where it could be found, and another feature (e.g., "An amethyst is a precious stone, mainly found in Brazil and it is usually purple." see Appendix C for details). The picture was then removed and the children were asked if they could remember what they had learned. They received corrective feedback ("Yes that's right, we learned that it is a precious stone mainly found in Brazil and it is usually purple.").

Pre- and Post-tests. The pre- and post-tests were the same as in Experiment 1, with the addition of a semantic knowledge post-test. Following the phonological choice pre-test, the children were asked if they had heard any of the 12 words before and, if so, if they could provide a definition of their meaning. This pre-test was administered at the start of all training conditions, to check if the children were familiar with any of the words. The semantic definition test was given again at post-test, but only in the reading-phonology-semantics condition to assess semantic learning. The definitions used during training were used to score answers given at pre- and post-test (definitions are provided in Appendix B).

Results

Semantic knowledge in the phonology plus semantics condition. We tested if semantic pre-training had had an impact by conducting a post-test for each child for those six words (out of the 12) in the reading-phonology-semantics pre-training condition. A Bayesian mixed-effect model estimated the effect of test time (pre- vs. post-test) on response accuracy in the semantic test as 3.96 (CrI[2.70, 5.98]) showing that semantic pre-training improved performance on the semantic test.

Phonological choice and reading aloud. We again fitted models of phonological choice and oral reading accuracy. We present the results summary in Table 5 and illustrate the effects of condition and test time on phonological choice (plot c.) and reading accuracy (plot d.) in Figure 4.

[Table 5, about here]

For phonological choice accuracy, we observed: a very small effect of condition (estimated effect = -.02, CrI[-.35, .31]), responses were about as likely to be correct under the reading-phonology-semantics as under the reading-plus-phonology condition; an effect of test time (1.88, CrI[1.42, 2.41]), responses were more likely to be correct at post- compared to pre-test; and a small effect of the interaction between the effects of condition and test time (.09, CrI[-.26, .44]).

For the analysis of reading data, we observed: a small effect of condition (.34, CrI[-.05, .78]), responses were more likely to be correct under the reading-phonology-semantics compared to the reading-plus-phonology condition; an effect of test time (2.02, CrI[1.50, 2.60]), responses were more likely to be correct at later test time points; and an effect due to the interaction between the effects of condition and test (.34, CrI[.07, .67]), such that the increase in reading accuracy from pre- to post-test was greater under the reading-phonology-semantics than under the reading-plus-phonology condition. Inspection of Figure 4 d. shows how the rate of increase in reading accuracy was steeper under the combined pre-training condition.

Sensitivity checks for Experiments 1 and 2. We examined the sensitivity of our results to variation in the way in which we coded critical variables or specified models. The effects estimates were similar over variation in the specification of priors and the number of MCMC samples (see OSF: sensitivity-analyses_2020-09-27.pdf). The same model specifications when fitted using lme4 yielded slightly larger effects estimates (see OSF: lme4-analyses_2020-09-27.pdf). Where credible intervals excluded 0 in our results, there effects were significant in the frequentist analyses, though sometimes lme4 model diagnostics indicated boundary singular fits.

In the analysis of reading aloud, we specified time as a continuous numeric variable (times 1-5), while in the analysis of phonological choice, in contrast, we specified time as a categorical effect (pre- vs. post-test). The difference in approach was not arbitrary: to us, it made sense to treat five different test times as variation akin to numeric variation whereas two test-times could only be pre-test or only post-test. However, we checked if coding time as a number could have influenced our results so we fitted models in which, for reading, we sum-coded time as a factor with five levels (1-5). Note that fitting a model of time coded as a factor will yield multiple estimates -- of the differences in outcomes observed at different levels of time -- whereas a model of the impact of continuously varying time yields one estimate of the rate at which outcomes vary over different times. Reassuringly, for Experiment 1 data, we found that there were effects of condition, test time and the interaction between condition and test time (see OSF: sensitivity-analyses_2020-09-27.pdf). For Experiment 2 data, we replicate the effect of time on reading performance, irrespective of coding, and while the interaction between condition and time effects is less precisely estimated if we code time as a factor rather than as a numeric quantity, we do see, interestingly, that the effect of pre-training condition is more pronounced at later compared to earlier time points, as Figure 4 makes clear.

Study Two exploratory analysis. We further examined the effects of pre-training across Experiments 1 and 2 in exploratory analyses. As described, we had taken measures of spelling ability (WRAT), vocabulary (WISC-IV), and word reading ability (SWRT) for all children recruited to Study Two. Thus, we had the opportunity to investigate whether individual differences modulated training response. It was possible that children who were more skilled would be better able to take advantage of the training or, in contrast, that children who were less skilled would tend to show the impact of training more strongly.

We examined, firstly, whether children in the two experiments responded in similar ways to phonological pre-training, analyzing the phonological choice and reading data collected under the phonological pre-training condition in both experiments. A model with experiment as a fixed effect (Experiment 1 vs. 2), incorporating random effects of participants and words on intercepts, and of words on the slope of the experiment effect, yielded little or no evidence for the effect of experiment on either phonological choice (estimated effect = -.03, CrI[-.21, .17]) or reading aloud (estimated effect = .01, CrI[-.22, .24]). This suggests consistency between experiments in the response to pre-training.

To investigate how the response to learning conditions varied, given differences in reading-related skills, we analyzed outcome data across Study Two Experiments to estimate the effects of condition, test time, and variation in vocabulary knowledge, reading and spelling ability, as well as the effects of the two- or three-way interactions between the effects of individual differences and the effects of condition and time. In this analysis, learning condition had three levels: reading-only; reading-plus-phonology or reading-phonology-semantics. As before, we sum coded the factors and, in the analysis of phonological choice, we treated time as a categorical variable (pre- vs. post-test) while, in the analysis of reading, we treated time as a standardized numeric variable (1-5 test times). We standardized the individual differences measures to z-scores before use as predictors. Models incorporated random effects taking into account variance due to differences between-participants or between-words in intercepts or in the effects of condition and time, as well as differences between words in the effects of individual differences. We present a summary of the results in Table 6.

[Table 6, about here]

The phonological choice model indicates little evidence for effects of individual differences or of interactions involving individual differences in vocabulary, reading or spelling ability. However, we observed effects of time (estimated effect = 1.84, CrI[1.46, 2.25]), and of the interaction between time and at least one condition contrast, that between reading-only and reading-phonology-semantics conditions (estimated effect = .47, CrI[.04, .95]).

The reading model indicates, similarly, weak evidence for effects of individual differences or of interactions involving individual differences in vocabulary, reading or spelling ability. For reading, however, we observed effects of learning condition, such that the contrast between reading-only and reading-plus-phonology (estimated effect = .49, CrI[.21, .77]), and the contrast between reading-only and reading-phonology-semantics learning conditions (estimated effect = 1.00, CrI[.56, 1.46]) suggested the benefits of different kinds of pre-training. We again observed an effect of test time (estimated effect = 1.75, CrI[1.42, 2.09]). And we again found an interaction between time and the contrast between reading-only and reading-phonology-semantics conditions (estimated effect = .17, CrI[.13, .79]), suggesting that the combined pre-training furnished the strongest contrast with reading-only learning and, for reading, that it grew in prominence over later test times.

The observation of these effects indicates that the benefits of pre-training conditions are robust to individual differences in vocabulary, reading or spelling abilities. However, the reading results present intriguing evidence for an interaction between the effects of differences in vocabulary and the effect of the contrast between reading-only and reading-phonology-semantics conditions (estimated effect = -.37, CrI[-0.78, .05]). The data do not allow us to estimate the interaction with precision, but a marginal effects plot (Figure 5) suggests that the administration of phonology or phonology-plus-semantics pre-training is most helpful for children scoring at the lowest vocabulary levels.

[Figure 5, about here]

Study Two Summary and Discussion

Study Two is the first developmental study to show that learning about the phonological form and meaning of an irregular word helps in learning to read it aloud. Improved performance in the phonological choice and definitions tasks at post-test indicates that pre-training supports the emergence of phonological and semantic representations. For subsequent reading aloud, our evidence indicates distinct benefits of phonological and of combined phonological and semantic pre-training. It further indicates differences in the time course of learning effects with important theoretical implications. The exploratory analysis hints that pre-training may compensate for weaker vocabulary in some children.

Phonological pre-training appeared to have a substantive initial impact on reading aloud (seen from reading test time 2, the first after pre-training, Figure 4 b). This is observed, in the absence of context, in the comparison between reading-plus-phonology and reading-only training conditions in Experiment 1. It is consistent with the results from adult learning reported by Taylor et al. (2011). And it is consistent with the assumption in the ST-CDP account that print presentation evokes the activation of phonological word representations through decoding. But it appears to be inconsistent with the assumption in the ST-DRC that, for irregular words, in the absence of contextual support, decoding cannot lead to sufficient activation of phonological representations for word learning to occur. An immediate benefit for reading due to phonological pre-training cannot yet be accommodated by connectionist models that do not incorporate self-teaching mechanisms. Relatively poor reading performance on the first exposure to the written word in the reading-only learning condition is consistent with both the ST-CDP and the ST-DRC, where an irregular word cannot be decoded without activation of a phonological representation.

We found that performance in the reading task improved continuously over the course of the learning trials across learning conditions (Figure 4 b). This suggests that opportunities to read aloud the words and receive feedback may have sustained the development of representations helpful to reading. The finding that phonological pre-training provides a distinct additional benefit which tends to grow over time (at a faster rate) implies that the further development of OP mappings leads to better performance in reading aloud. This observation is consistent with results from connectionist simulations demonstrating the benefit of phonological pre-training as well as the benefits of more effective phonological representations (Harm et al., 1994; Harm & Seidenberg, 1999).

We found that the combination of semantic with phonological pre-training had additional benefits for reading irregular words, most markedly at later test times. This observation is consistent with the results reported by Taylor et al. (2011) for adults, extending our understanding of the impacts of semantic pre-training to children. Finding an effect of semantic pre-training on irregular word reading development appears to be broadly consistent with the predictions that can be derived from both self-teaching and connectionist accounts. However, the sustained nature of the semantic benefit (Figure 4 d) would seem to be more consistent with the assumption in the ST-DRC (Pritchard et al., 2018) and connectionist accounts (Plaut et al., 1996) that activation from semantics contributes to the activation of output phonology for irregular words, and with the idea that the semantic contribution will tend to grow as OS or OSP mappings develop (Plaut et al., 1996).

General Discussion

We examined the contributions of phonological and semantic knowledge to regular and irregular word reading, and to the development of the capacity to read irregular words. Models of reading development that incorporate self-teaching mechanisms (ST-DRC, Pritchard et al., 2018; ST-CDP, Ziegler et al., 2014) assume, in common, that the activation of phonological word form representations is essential to learning lexical orthography-to-phonology (OP) mappings. Both ST-DRC and ST-CDP accounts assume that decoding is required to activate phonological representations, and both incorporate the contribution of context to activation of the phonology for irregular words. But the accounts differ in whether phonological representations for irregular words can be evoked to a level sufficient for learning in the absence of context: it is possible in current instantiations of the ST-CDP but not in the ST-DRC. In comparison, connectionist triangle model accounts do not, at present, incorporate self-teaching mechanisms nor assume lexical representations, but associated simulations (Harm et al., 1994; Harm & Seidenberg, 1999) demonstrate the impact on reading development of phonological pre-training and of variation in the effectiveness of phonological representations. In addition, a series of simulations have demonstrated the importance of the contribution from semantic knowledge to reading, especially, irregular words (Dilkina et al., 2008; Plaut et al., 1996) while verbal accounts have conjectured that the semantic influence should be late emerging. Connectionist accounts, in general, predict the emergence of a gradual ceiling effect (Plaut et al., 1996; Seidenberg & McClelland, 1989) which would tend to predict larger effects of decoding skill or word knowledge on irregular word reading. Our observations reveal aspects of the performance of children, in cross-sectional comparisons, and in response to training, congruent with the predictions that can be derived from these accounts. They locate limitations, however, in all three accounts which identify targets for future extensions.

Study One results showed that children's reading accuracy was greater for regular than for irregular words, and that variation in decoding skill, as well as in lexical phonological and semantic knowledge, contributed to reading accuracy overall. Children who could demonstrate accurate lexical phonological and semantic knowledge were more likely to read regular or irregular words accurately. Our analyses suggested trends (not resolved with certainty) where differences in age, decoding skill or lexical knowledge more strongly affected irregular word than regular word reading. The difference in trends between word types may well be because regular word reading accuracy was near to ceiling. Even so, the effect of variation in decoding skill evident in irregular word reading indicates that the relative efficiency of sub-lexical OP mappings tends to influence the accuracy of exception word coding. Likewise, while the effect of age may reflect the impact of exposure, the effects of variation in lexical knowledge show that differences in the strength or quality of word form or meaning knowledge are important to irregular word reading. Our investigation of the impact of phonological training, in Study Two Experiment 1, revealed the near-term impact of pre-training in the absence of context. Phonological pre-training continued to exert a sustained influence on irregular word reading though the effect tended to diminish as performance approached ceiling. A comparison between Study Two Experiments 1 and 2 showed that the phonological training intervention could be replicated. The results from our comparison of the impacts of phonological versus phonology-plus-semantics training, in Study Two Experiment 2, showed that the addition of semantic pre-training resulted in an extra benefit for reading performance, increasing through later test times. The Experiment 2 analyses indicated that interventions designed to support the establishment of lexical knowledge had separable effects in distinct phonological and semantic influences on irregular word reading. Exploratory analyses suggested that semantic training for irregular words may benefit more children with weaker vocabularies.

Our observation in Study One that accuracy in auditory lexical decision predicted accuracy in reading aloud is consistent with the assumption that self-teaching depends on the activation of lexical phonology, and with the assumption in connectionist triangle model accounts that reading development benefits from more effective phonological representations. Study One, in examining reading responses to words that have been learned (more or less) offers a retrospective picture of what factors may have promoted prior learning. If phonological representations are better specified (even if only to the extent they support accurate auditory lexical decision) then they ought to have been activated more readily in the learning of lexical OP mappings. The self-teaching accounts can be distinguished from connectionist triangle model accounts because the latter do not, yet, incorporate a self-teaching mechanism. The ST-DRC and ST-CDP can be distinguished from each other because the ST-CDP allows decoding to activate candidate lexical phonological representations in response to irregular word spellings, in the absence of context, while the ST-DRC does not. In Study Two, we tested these predictions in an examination of what phonological pre-training can achieve as reading opportunities accumulate. Our observation, in Experiment 1, that phonological pre-training benefits performance in irregular word reading from the first test time after training is consistent with the ST-CDP but not the ST-DRC account.

Whether or not phonological representations can be activated in the absence of context depends, in the self-teaching models, on the level of inhibition permitted between phoneme and lexical phonological representations. Pritchard et al. (2018) argued that because the ST-DRC implemented the same parameterization as the DRC model, including a high level of phoneme-phonological lexicon inhibition, the ST-DRC shared the capacity of the DRC to account for benchmark effects in skilled adult readers. Our results suggest that future simulation work is now required to show whether a relaxation in the phoneme-phonology parameter would enable the ST-DRC to simulate, without a contribution from context, the evocation of cohorts of phonological representations in response to novel irregular words. Further, given the aims set out by Pritchard et al. (2018), where the capabilities of prior models should be retained in descendant implementations, future simulation work is required to show whether an ST-DRC instance with reduced phoneme-phonology inhibition could both simulate the effect of phonological training without context and the benchmark capabilities of the skilled reader. Pritchard et al. (2018) note that it has not yet been reported whether the ST-CDP model is capable of simulating both the acquisition of OP mappings and the critical effects observed in skilled reading.

It is worth reflecting, at this point, on the broader theoretical implications of our findings. A more satisfactory account of the behavioural data may be furnished by computational models in which sub-lexical decoding is permitted to evoke a cohort of lexical phonological representations in the absence of context. But what does this imply about what children do? One possibility is that if beginning readers are less sure of their decoding ability compared to more skilled readers then they are more likely to consider multiple phonological word forms as candidates for the correct pronunciation of a letter string. We do not have to assume that children explicitly consider candidates for the pronunciation of a lexical string. It may be that, in beginning readers, encounters with new irregular words result, at first, in the activation of a cohort of candidate words and that selection of a pronunciation for a letter string derives, for example, from a probability distribution such that different candidates are associated with differing plausibility. It may be that in beginning readers, the plausibility distribution is broader, encompassing more candidates, or allocating more equal levels of plausibility to alternate candidates. More skilled readers may be less likely to activate a broader cohort of candidates, consistent with increased levels of phoneme-phonological inhibition. Whether decoding yields activation of broader cohorts of phonological word forms in earlier, less skilled readers, whether such activation narrows over time, and whether, indeed, this is a developmental process associated with children's evaluation of alternate decoding candidates, seem to us to be fruitful possibilities for future research.

Future simulation work is required to show, also, whether connectionist triangle models can both simulate the initial learning of OP mappings and the emergence of skilled reading behavior. However, the finding that phonological pre-training provides an additional benefit that initially tends to grow over time is consistent in detail with the results from previous simulations. Harm and Seidenberg (1999) found that phonological pre-training benefited reading development compared to the absence of pre-training but that the difference was small and reduced rapidly. In contrast, strikingly, variant simulations reported by Harm, Altmann and Seidenberg (1994) showed that the impact of phonological pre-training could be found to be substantial and sustained, with reading performance rising faster, and remaining better at asymptote, exactly as we observed (Figure 4 b). The difference between the Harm and Seidenberg (1999) and the Harm et al. (1994) simulations is that, in the first, phonological training and reading training events are interleaved but, in the second, a block of phonological pre-training preceded reading training. Of course, just as in the Harm et al. (1994) simulations, in our Study Two, phonological pre-training precedes reading training events. What we find, then, is that a match in the organization of phonological pre-training and then reading training is associated with a remarkable match in the time course and relative extent of the benefit of phonological pre-training.

In both studies, decoding skill was found to be essential in emergent irregular word reading. The results of Study One show that higher levels of decoding skill predict greater probability of accurate reading aloud most prominently for irregular words. The results of the analysis of reading trial data in Study Two show that the output phonology of novel irregular words can be decoded, in the absence of context, though more readily following phonological or phonological plus semantic pre-training. This shows the importance of partial decoding in irregular word reading: applying decoding skills will contribute to the activation of the correct pronunciation. The capacity of sub-lexical mappings to contribute activation to the phoneme level is integral to self-teaching for both ST-DRC and ST-CDP models. This implies that an overall decoding effect is consistent with such accounts though, as discussed, the growth in irregular word reading in Study Two trials, in the absence of context, would seem to be inconsistent with the ST-DRC. However, the assumptions of self-teaching accounts appear to predict that the effect of individual differences in decoding skill (as seen in Study One) should be more important for regular than for irregular word reading. This is because regular word phonology is, by definition, capable of being wholly determined by the outputs of sub-lexical OP mappings or GPCs. Hence, it follows that children who present higher levels of decoding skill ought to demonstrate a higher level of accuracy on regular word reading. We observed, instead, that the decoding effect tended be apparent more strongly in irregular than in regular word reading. This finding appears to be inconsistent in detail with the expectations deriving from self-teaching accounts. In contrast, it is consistent with the expectations that follow from the assumption, in connectionist accounts (Plaut et al., 1996; Seidenberg & McClelland, 1989), that input-output activation functions take a non-linear form and, consequently, that a gradual ceiling effect will tend to emerge given higher levels of factors (like decoding skill) that may drive improvements in reading skill. This is because the impact of higher levels of decoding skill will tend to diminish as reading approaches maximum accuracy and, in our sample, that ceiling is approached for regular words before irregular words.

Our observation that decoding contributes to irregular word reading extends our understanding of reading development because previous behavioural studies did not take into account individual differences in decoding ability (Nation & Cocksey, 2009; Ricketts et al. 2016). Future research is required to identify if the use of more challenging words would reveal stronger evidence for an interaction between the effects of decoding skill and word spelling-sound regularity or consistency. However, our results show that the range of decoding skills that can be found in children as young as seven years is likely sufficient to decode even regular words that are chosen to be relatively challenging. The simulations reported by Pritchard et al. (2018) indicate that regular word reading accuracy can be at or near ceiling (Simulation 2), granted full knowledge of GPCs, given weak levels of contextual support. This implies that to observe an effect of decoding skill on regular word reading researchers will need to test children with incomplete GPC knowledge, sampling from ages younger than seven years.

In both studies, we detected an independent effect of lexical semantic knowledge on emergent reading. In Study One, accuracy in the semantic definition task predicted accuracy in reading aloud, more in irregular than in regular word reading. In the ST-DRC and in connectionist accounts there is an explicit assumption that semantic activation can be transmitted to the phoneme level to enable phonological word form output in reading aloud. Consistent with our findings, connectionist triangle model simulation results (Harm & Seidenberg, 2004; Plaut et al., 1996) indicate that the influence of semantic information will tend to be stronger for irregular (or inconsistent) word reading. In Study Two, we observed a later emerging influence of semantic pre-training, similar to that reported by Taylor et al. (2011) in adult learning. Intriguingly, this result matches the expectation that arises, in the connectionist account, from the conjecture that mappings need to become established in order to allow the evolution of a semantic influence (Plaut et al., 1996). Our Study One finding is consistent, in principle, also, with the ST-DRC assumptions that activation from the semantic level may support the activation of lexical phonological representations, and that such support will be more necessary for irregular words that cannot be decoded accurately given GPC knowledge alone. The ST-CDP is somewhat less explicit on how semantics may contribute to phonological activation. However, contextual information is assumed in the ST-CDP to enable the selection of the appropriate phonological representation, given the decoding-based activation of a cohort of candidate word forms, when a novel irregular word is encountered. We think that future simulation work is required to examine in detail how semantic information may contribute to reading development in self-teaching models.

In the Experiments in Study Two, corrective feedback was provided during the reading-based training. It could be argued that this is a form of supervised feedback. In the ST-CDP simulations, the role of context was not modelled, and the correct phonological representation was selected simply if it was in the cohort. Ziegler et al. (2014) go on to say that an external teacher is required for the selection of word candidates with ambiguous spellings. Proving feedback in this way is akin to what happens during supervised reading when a child encounters an irregular word, However, we recognize that reading is not always supervised and it would be theoretically and practically interesting to see how learning progresses for irregular words with just pre-training but no feedback during reading-based learning.

In our Experiments, written word training followed pre-training on the same day. It is possible that recent exposure to the spoken words may have exaggerated the role of lexical knowledge. Alternatively, not having had time to consolidate oral learning may have reduced its impact. Some studies of word learning suggest that the impact of semantics may be greater after a period of consolidation (e.g., Clay et al., 2007). Manipulating the delay between the two aspects of training would also provide information about when is best to teach children about the spoken and written forms of words: should they be taught at the same time, on the same day but with spoken first, or with spoken information taught first with time for consolidation? The addition of semantics to pre-training led to a modest boost at post-test. One possibility is that semantic training leads to better quality lexical representations that are maintained for longer. Including a delayed post-test (e.g., 24 hours) would test this hypothesis. Previous studies have shown that context is more beneficial for learning irregular words than regular words, when the phonology and meaning of both have been pre-trained (e.g., Wang et al., 2011). If context works to pre-activate lexical-semantic representations (as in the ST-DRC) then presenting words in context should only have a beneficial effect if pre-training has included semantics. Finally, the more modest boost associated with semantic pre-training may have reflected the fact that semantics were added to pre-training in phonology and this pre-training substantially improved reading. It is possible that reducing the amount of phonological pre-training would reveal a greater role for semantic knowledge.

Our results have practical implications for classroom instruction. The role of decoding in irregular word reading indicates that the teaching of systematic phonics will support the development of irregular, as well as regular, word reading even though irregular words can only be partially decoded. Our Study Two finding that phonological pre-training led to a substantial benefit in the early stages of learning to read irregular words implies that exposing children to the spoken form of an irregular word, even if there is no time for detailed semantic instruction, will benefit reading development for the most difficult, irregular, words. Our exploratory analysis of the role of individual differences in learning to read irregular words suggests that phonological pre-training may be more beneficial for children with weaker vocabulary knowledge. However, future research is needed in which a sample of emergent readers is recruited, stratified by vocabulary knowledge, to examine whether pre-training effects are, indeed, stronger for lower vocabulary ranges. Our results show, also, that even if the contribution of semantic knowledge is, by comparison, modest it is likely important for irregular or inconsistent word reading. In addition, the role of semantic knowledge is likely to be greater for irregular words encountered in text, where context can activate semantic knowledge to support reading.

In conclusion, our findings provide strong evidence that irregular or inconsistent words presented in isolation are read using a combination of decoding and lexical knowledge, and that word-level phonological and semantic knowledge have important, separable, influences on performance in emergent reading performance. Results from a study of training indicate that the impact of phonological pre-training is apparent early in the accumulation of opportunities to practice reading aloud, and continues to be sustained over time even as reading performance approaches maximum accuracy. The additional benefit of incorporating semantic information in pre-training is identified but it appears to grow stronger over time. Our results motivate simulation work to examine potential extensions to computational models of reading development, based on self-teaching accounts, that could accommodate the effect of phonological pre-training on irregular word reading, when irregular words are presented in isolation, the influence of semantic knowledge, and the greater effect of decoding for irregular than for regular word reading. While connectionist accounts appear to be consistent with our results in a number of respects, future simulation work is required, also, to examine whether a connectionist model of reading development can incorporate self-teaching at the onset of orthographic learning. Our findings lead us to advocate an instructional approach whereby phonics instruction is combined with vocabulary instruction to support children’s reading of irregular words.

References

Armstrong, B.C., Dumay, N., Kim, W., & Pitt, M.A. (2017). Generalization from newly learned words reveals structural properties of the human reading system. Journal of Experimental Psychology: General, 146, 227-249.

Baron, J. (1979). Orthographic and word-specific mechanisms in children's reading of words. *Child Development*, 60-72. https://doi.org/10.2307/1129042

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.2012.11.001Bates, D., Maechler, M., & Bolker, B. Walker., S. (2015). Fitting linear mixed-effects models using lme4. *J Stat Softw*, *67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... & Green, P. (2018). Package 'lme4'. *Version*, *1*, 17.

Bürkner, P. C. (2018). *Advanced bayesian multilevel modeling with the R package brms*. R J. 10, 395–411. doi: 10.32614. RJ-2018-017.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of statistical software*, 76(1). https://doi.org/10.18637/jss.v076.i01

Castles, A., Coltheart, M., Larsen, L., Jones, P., Saunders, S., & McArthur, G. (2009). Assessing the basic components of reading: A revision of the Castles and Coltheart test with new norms. *Australian Journal of Learning Difficulties*, 14(1), 67-88. https://doi.org/10.1080/19404150902783435

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

Clay, F., Bowers, J. S., Davis, C. J., & Hanley, D. A. (2007). Teaching adults new words: The role of practice and consolidation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(5), 970–976. https://doi.org/10.1037/0278-7393.33.5.970

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology*, 33A (4), 497-505. https://doi.org/10.1080/14640748108400805

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

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, 25(1), 7-29.

Davies, R. A., Arnell, R., Birchenough, J. M., 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. https://doi.org/10.1037/xlm0000366

Depaoli, S., & Van de Schoot, R. (2017). Improving transparency and replication in Bayesian statistics: The WAMBS-Checklist. Psychological methods, 22(2), 240.

https://doi.org/10.1037/met0000065

Dilkina, K., McClelland, J. L., & Plaut, D. C. (2008) A single system account of semantic and lexical deficits in five semantic dementia patients, Cognitive Neuropsychology, 25:2, 136-164, DOI: 10.1080/02643290701723948

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

Eager, C., & Roy, J. (2017). Mixed effects models are sometimes terrible. *arXiv preprint arXiv:1701.04858*. https://doi.org/10.1063/pt.5.028530

Foster, H. (2007). Single word reading test 6-16. Windsor, England: Nfer-Nelson Publishing Company Ltd.

Gathercole, S., & Baddeley, A. (1996). *The children's test of nonword repetition (CNRep)*. London: The Psychology corporation.

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. https://doi.org/10.1037/0096-1523.5.4.674

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

Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. Journal of memory and language, 29(6), 687-715. https://doi.org/10.1016/0749-596x(90)90044-z

Kim, W., Pitt, M. A., & Myung, I. J. (2013). How do PDP models learn quasiregularity? Psychological Review, 120, 903-916

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, *25*(1), 178-206. https://doi.org/10.3758/s13423-016-1221-4

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, 94, 305-315. https://doi.org/10.1016/j.jml.2017.01.001

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior research methods*, 44(2), 314-324. https://doi.org/10.3758/s13428-011-0168-7

McKague, M., Pratt, C. & Johnston, M. B. (2001). The effect of oral vocabulary on reading visually novel words: a comparison of the dual-route cascaded and triangle frameworks. *Cognition*, 80, 231-262. https://doi.org/10.1016/S0010-0277(00)00150-5

McKay, A., Davis, C., Savage, G. & Castles, A. (2008). Semantic involvement in reading aloud: evidence from a nonword training study. *Learning, Memory, and Cognition*, 34 (6), 1495-1517. https://doi.org/10.1037/a0013357

Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E. J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review*, 23(1), 103-123.

Nation, K. & Cocksey, J. (2009). The relationship between knowing a word and reading it aloud in children's word reading development. *Journal of Experimental Child Psychology*, 103, 296–308. https://doi.org/10.1016/j.jecp.2009.03.004

Ouellette, G., & Beers, A. (2010). A not-so-simple view of reading: How oral vocabulary and visual-word recognition complicate the story. *Reading and Writing, 23*(2), 189-208. doi:10.1007/s11145-008-9159-1

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

Plaut, D. C., McClelland, J. L., Seidenberg, M. S. & Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological Review*, 103 (1), 5-115. https://doi.org/10.1037/0033-295X.103.1.56

Pratt, C., Johnston, M. B., & Morton, A. (1999). The effect of oral instantiation on children's nonword reading. *Australian Journal of Psychology*, 51, 39.

Pritchard, S., Coltheart, M., Marinus, E., & Castles, A. (2018). A computational model of the self-teaching hypothesis based on the dual-route cascaded model of reading. *Cognitive science*, *42*(3), 722-770. https://doi.org/10.1111/cogs.12571

Ricketts, J., Nation, K., & Bishop, D. V. M. (2007). Vocabulary is important for some, but not all reading skills. *Scientific Studies of Reading*, 11, 235–257. https://doi.org/10.1080/10888430701344306

Ricketts, J., Davies, R., Masterson, J., Stuart, M., & Duff, F. J. (2016). Evidence for semantic involvement in regular and exception word reading in emergent readers of English. *Journal of Experimental Child Psychology*, 150, 330-345. http://dx.doi.org/10.1016/j.jecp.2016.05.013

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

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

Share, D. L. (1995). Phonological recoding and self-teaching: sine qua non of reading acquisition. *Cognition*, 55, 151-218. https://doi.org/10.1016/0010-0277(94)00645-2

Taylor, J. S. H., Plunkett, K., & Nation, K. (2011). The influence of consistency, frequency, and semantics on learning to read: An artificial orthography paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(1), 60. https://doi.org/10.1037/a0020126

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of word reading efficiency*. Austin, TX: Pro-Ed.

van Heuven, W.J.B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, 67, 1176-1190. https://doi.org/10.1080/17470218.2013.850521

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological bulletin*, *101*(2), 192. https://doi.org/10.1037/0033-2909.101.2.192

Wang, H. C., Castles, A., Nickels, L., & Nation, K. (2011). Context effects on orthographic learning of regular and irregular words. *Journal of experimental child psychology*, 109(1), 39-57. https://doi.org/10.1016/j.jecp.2010.11.005

Wang, H. C., Nickels, L., Nation, K., & Castles, A. (2013). Predictors of orthographic learning of regular and irregular words. *Scientific Studies of Reading*, 17(5), 369-384. https://doi.org/10.1080/10888438.2012.749879

Wechsler, D. (2003). *Wechsler intelligence scale for children – fourth edition: Administration manual*. Lutz, FL: Psychological Assessment Resources, Inc.

Wilkinson, G. S. & Robertson, G. J. (2006*). Wide range achievement test 4 (WRAT4): Professional Manual*. Lutz, FL: Psychological Assessment Resources, Inc.

Ziegler, J. C., Perry, C., & Zorzi, M. (2014). Modelling reading development through phonological decoding and self-teaching: implications for dyslexia*. Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1634), 20120397. https://doi.org/10.1098/rstb.2012.0397

Zorzi, M., Houghton, G., & Butterworth, B. (1998). Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance*, 24(4), 1131. https://doi.org/10.1037/0096-1523.24.4.1131

Table 1

Performance on the reading aloud, auditory lexical decision and definitions tasks in Study 1 as a function of orthographic regularity and age group averaged across participants (top) and items (below)

	Younger		Older	
By-participants (max 38)	*M*	*SD*	*M*	*SD*
Reading Regular	33.45	5.24	35.33	2.79
Reading Irregular	19.21	5.47	24.21	4.03
Auditory lexical decision Regular	27.06	3.54	29.94	3.11
Auditory lexical decision Irregular	26.42	2.69	29.42	3.61
Definitions Regular	20.06	3.90	23.48	4.15
Definitions Irregular	20.97	3.66	24.97	3.52
By-items (max 33)	*M*	*SD*	*M*	*SD*
Reading Regular (max	29.61	5.27	30.71	3.63
Reading Irregular	19.03	12.34	18.53	12.08
Auditory Lexical decision Regular	25.03	9.27	24.61	9.92
Auditory Lexical Irregular	23.92	10.50	24.39	11.02
Definitions Regular	18.92	12.32	18.87	12.29
Definitions Irregular	20.08	12.63	20.39	12.41

Table 2

Study one – model summary for reading aloud: regularity

	Estimate	Error	l-95%	u-95%
Intercept	2.45	0.32	1.83	3.10
Year group	0.19	0.14	-0.08	0.46
Decoding	0.95	0.14	0.69	1.23
Definitions	0.70	0.11	0.48	0.92
Auditory lexical decision	0.32	0.09	0.14	0.50
Word type	1.92	0.32	1.28	2.54
Year group x Definitions	0.13	0.09	-0.04	0.31
Year group x Auditory lexical decision	0.08	0.08	-0.09	0.24
Decoding x Definitions	0.05	0.08	-0.11	0.21
Decoding x Auditory lexical decision	0.07	0.08	-0.09	0.23
Year group x Word type	-0.21	0.14	-0.48	0.05
Decoding x Word type	0.21	0.13	-0.05	0.48
Definitions x Word type	0.12	0.11	-0.10	0.33
Auditory lexical decision x Word type	-0.10	0.09	-0.28	0.08
Year group x Definitions x Word type	-0.01	0.09	-0.18	0.17
Year group x Auditory lexical decision x Word type	0.10	0.08	-0.06	0.26
Decoding x Definitions x Word type	-0.11	0.08	-0.26	0.04
Decoding x Auditory lexical decision x Word type	-0.10	0.08	-0.25	0.06

Table 3

Mean (standard deviation) scores on the tests of language and literacy for the children in Study 2, Experiment 1 and Experiment 2

	Experiment 1	Experiment 2
Test	*M* norm-referenced score (*SD*)	*M* norm-referenced score (*SD*)
Single word reading	107.47 (14.15)[a]	114.24 (12.67)[a]
Spelling	113.69 (11.11)[a]	117.24 (13.83)[a]
Expressive vocabulary	9.28 (2.19)[b]	9.91 (2.21)[b]

Notes: [a]Standard score (*M* = 100, *SD* = 15); [b]Scaled Score (*M* = 10, *SD* = 3)

Table 4

Study 2 – Experiment 1 model summaries for phonological choice and reading aloud.

	Phonological choice				Reading aloud			
	Estimate	Error	l-95%	u-95%	Estimate	Error	l-95%	u-95%
Intercept	1.50	0.34	0.85	2.20	0.75	0.43	-0.10	1.64
Condition	0.39	0.17	0.08	0.72	0.96	0.17	0.63	1.30
Test time	1.47	0.22	1.06	1.93	1.54	0.18	1.17	1.91
Condition x Test time	0.51	0.17	0.18	0.87	0.26	0.13	< -0.01	0.53

Table 5

Study 2 – Experiment 2 model summaries for phonological choice and reading aloud.

	Phonological choice				Reading aloud			
	Estimate	Error	l-95%	u-95%	Estimate	Error	l-95%	u-95%
Intercept	1.68	0.37	0.99	2.45	2.07	0.51	1.06	3.09
Condition	-0.02	0.17	-0.35	0.31	0.34	0.21	-0.05	0.78
Test time	1.88	0.25	1.42	2.41	2.02	0.27	1.50	2.60
Condition x Test time	0.09	0.18	-0.26	0.44	0.34	0.15	0.07	0.67

Table 6

Study 2 –model summaries for phonological choice and reading aloud; collated data

	Phonological choice				Reading aloud			
	Estimate	Error	l-95%	u-95%	Estimate	Error	l-95%	u-95%
Intercept	1.75	0.36	1.08	2.51	1.23	0.46	0.33	2.16
zSWRT	< 0.01	0.24	-0.48	0.45	0.31	0.24	-0.15	0.79
zVOCAB	0.03	0.20	-0.35	0.42	0.36	0.21	-0.05	0.76
zWRAT	0.28	0.24	-0.17	0.75	0.39	0.25	-0.08	0.88
Condition2	0.13	0.19	-0.24	0.49	0.49	0.14	0.21	0.77
Condition3	0.35	0.26	-0.16	0.87	1.00	0.23	0.56	1.46
Test time	1.84	0.20	1.46	2.25	1.75	0.17	1.42	2.09
Experiment	-0.13	0.14	-0.40	0.14	-0.16	0.15	-0.47	0.14
zSWRT x Condition2	0.34	0.26	-0.16	0.87	0.24	0.18	-0.11	0.59
zSWRT x Condition3	-0.20	0.39	-0.97	0.54	-0.18	0.28	-0.73	0.35
zVOCAB x Condition2	-0.06	0.22	-0.49	0.37	-0.01	0.16	-0.32	0.29
zVOCAB x Condition3	-0.18	0.30	-0.77	0.41	-0.37	0.21	-0.78	0.05
zWRAT x Condition2	-0.19	0.26	-0.70	0.31	-0.11	0.21	-0.51	0.29
zWRAT x Condition3	0.35	0.37	-0.35	1.09	0.38	0.27	-0.15	0.91
zSWRT x Test time	0.21	0.22	-0.22	0.64	0.21	0.16	-0.10	0.52
zVOCAB x Test time	-0.14	0.18	-0.49	0.21	<-0.01	0.13	-0.27	0.26
zWRAT x Test time	0.13	0.23	-0.32	0.59	0.01	0.16	-0.30	0.32
Condition2 x Test time	0.15	0.18	-0.20	0.51	-0.02	0.12	-0.25	0.20
Condition 3 x Test time	0.47	0.23	0.04	0.95	0.45	0.17	0.13	0.79
zSWRT x Condition2 x Test time	0.15	0.26	-0.34	0.67	<-0.01	0.15	-0.29	0.28
zSWRT x Condition3 x Test time	-0.08	0.36	-0.81	0.62	0.08	0.22	-0.36	0.52
zVOCAB x Condition2 x Test time	-0.29	0.24	-0.76	0.17	0.14	0.12	-0.10	0.38

zVOCAB x Condition3 x Test time	-0.01	0.30	-0.60	0.59	-0.11	0.17	-0.45	0.23
zWRAT x Condition2 x Test time	-0.09	0.26	-0.60	0.41	0.08	0.16	-0.23	0.38
zWRAT x Condition3 x Test time	0.25	0.35	-0.42	0.94	0.07	0.23	-0.37	0.53

Figure 1

Key differences between the ST-DRC and ST-CDP

Figure 2

Study one – reading aloud: marginal effect plots for key interactions

Figure 3

A summary of the procedure for Study 2. Conditions: reading-only (R); reading-plus-phonology (P+R) phonological pre-training; reading-phonology-semantics (P+S+R) combined phonological and semantic pre-training.

Figure 4

Study two: experiments 1 and 2 – phonological choice and reading aloud: plots for condition x test time interactions

Figure 5

Study two – reading aloud: marginal effect plot for condition x vocabulary interaction

Table A1

The two lists of irregular words used in Study 2, with the semantic information provided for each word and the phonological distractors used in the phonological choice test

List	Word	Freq [a]	Length	Phonemes	Syllables	Semantic information	One feature distractor	Two feature distractor
List 1	chamois	2.35	7	5	2	A chamois is a small, wild antelope. It lives in the mountains and has two horns on its head.	shammay	sammay
	amethyst	3.01	8	7	3	An amethyst is a precious stone. It's mainly found in Brazil and is usually purple.	ammethost	annuthost
	plaid	4.16	5	4	1	Plaid is a fabric that people wear. It has a pattern where different colours are crossed over each other.	plag	prag
	algae	3.44	5	4	2	Algae is a plant that's found in water. It doesn't have a stem or leaves.	alga	alda
	croquet	3.23	7	5	2	Croquet is a game that people play outside using wooden balls and mallets.	crokar	crotar
	niche	3.45	5	3	1	A niche is a small hole or gap, usually in a wall.	naysh	naych
List 2	chassis	3.14	7	4	2	A chassis is the frame of a car. Other parts of the car are mounted onto it.	shossee	shossoo
	crypt	2.87	5	5	1	A crypt is a room under a church. It's sometimes used as a burial place.	crupt	clupt
	pancreas	2.83	8	8	3	The pancreas is a gland near the stomach. It makes juices to help break down food.	pandreas	pandreeaff
	okapi	1.77	5	5	3	An okapi is a small deer-like animal. It lives in Africa and has a long tongue.	odahpi	odahper

debris	3.71	6	5	2	Debris is the remains of something that's been broken or destroyed. Sometimes you find it on a beach after a shipwreck.	depree	deproo
tsar	3.17	4	3	1	A tsar is an emperor or a king who rules Russia.	zee	zer

Note: [a] Frequency SUBTLEX_UK LogFreq Zipf value