

How granularity of orthography-phonology mappings affect reading development: Evidence from a computational model of English word reading and spelling

Alfred Lim and Beth A. O'Brien
Nanyang Technological University, Singapore
alfred.lim@nie.edu.sg
beth.obrien@nie.edu.sg

Luca Onnis
University of Genoa, Italy
luca.onnis@unige.it

Abstract

It is widely held that children implicitly learn the structure of their writing system through statistical learning of spelling-to-sound mappings. Yet an unresolved question is how to sequence reading experience so that children can ‘pick up’ the structure optimally. We tackle this question here using a computational model of encoding and decoding. The order of presentation of words was manipulated so that they exhibited two distinct progressions of granularity of spelling-to-sound mappings. We found that under a training regime that introduced written words progressively from small-to-large granularity, the network exhibited an early advantage in reading acquisition as compared to a regime introducing written words from large-to-small granularity. Our results thus provide support for the grain size theory (Ziegler and Goswami, 2005) and demonstrate that the order of learning can influence learning trajectories of literacy skills.

1 Introduction

Reading science provides evidence of the developmental path to acquiring reading for alphabetic languages (Ehri, 2005; Rayner et al., 2001). From parsing the speech stream into words in infancy (Christiansen et al., 2006; Saffran et al., 1997), to familiarizing with print in the preschool years (Thompson, 2009) — these activities lead to the accrual of key knowledge for learning to read. Knowledge about the language’s phonotactic and

graphotactic properties and symbolic representations with abstract letter units is necessary for the forthcoming insight that print represents spoken language (the alphabetic principle). Subsequent to this insight, children are ready to take on the process of learning the precise mapping of print-to-speech.

At its basis, learning to read involves learning to *decode* a script into oral language representations. The question arises as to the optimal input for learning this orthography-to-phonology mapping in an alphabetic system, especially for languages that have deep orthographies, such as English. Shallow orthographies (e.g., Finnish, Spanish) have a more precise match between letters and sounds; whereas deep orthographies match phonemes to graphemes (one or more letters) in an inconsistent way — with multiple spellings per phoneme, or multiple pronunciations per grapheme — and, thus, have a greater number of GPCs (grapheme-phoneme correspondences). Therefore, reading acquisition is found to occur at a comparatively slower rate for readers in deep as compared to shallow orthographies (Ellis et al., 2004; Georgiou et al., 2008; Florit and Cain, 2011).

The deep orthographic complexity of English also partly results from variation in the functional units of the writing system — graphemes — which may consist of a single letter (e.g., *a*), or multiple letters (e.g., *ay*, *aye*). While skilled adult readers have unitized these subword patterns (Rey et al., 2000), beginning readers need to acquire these patterns of graphemes and their mappings to phonemes. Here we consider this mapping problem along two dimensions: **(1)** the *granularity* of the units of analysis to be picked up at any given time; and **(2)** the ordering of learning such units and types.

A fruitful approach to examining the GPC learning process is through computational mod-

Copyright ©2020 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

elling (Monaghan and Ellis, 2010; Perry et al., 2019; Pritchard et al., 2016). Specifically, connectionist models are sensitive to the timing and ordering of learning events, in that they learn incrementally. This feature is particularly apt for modeling reading development, as it affords simulating the incremental nature of a child learning to read new words daily, as schooling progresses. Order effects as well as frequency trajectory effects have been documented in previous connectionist models (Mermillod et al., 2012), and here we are interested in comparing learning trajectories for particular training orderings for reading development.

To this end, we present connectionist networks with small batches of words, which we test regularly for accuracy until a given criterion across the batch is achieved — in essence, an adaptive training regime. Using this approach, we can address long-standing issues in the area of reading education with a more systematic approach to understanding how print-to-speech mappings are learned (Rueckl, 2016).

Below we briefly review why print-to-speech decoding can be a hard problem, both for learners and for researchers trying to understand its mechanisms. Then, we discuss dimensions of granularity derived from the literature, and offer a first set of connectionist simulations of the order of reading acquisition of American English.

1.1 Is there an optimal reading experience?

The psycholinguistic grain size theory (Ziegler and Goswami, 2005) has generated much research on reading acquisition, including across different alphabetic languages. It espouses that granularity for oral and written language development proceed in different directions — from larger to smaller, vs. from smaller to larger units. Thus, the mismatch in unit or “grain” sizes available over development introduces a disparity in learning the mapping between orthography and phonology.

This learning challenge has led to investigations of behavioral interventions for teaching reading at either whole word or subword levels (National Reading Panel, 2001; McArthur et al., 2015), showing an advantage for subword approaches emphasizing letter, grapheme or larger (subsyllable onset-rime) units (Rayner et al., 2001; Ehri et al., 2001; Torgerson et al., 2006; Olson and Wise, 1992; Ecalle et al., 2009). At the same time, the optimal subword grain size has been debated. De-

velopmentally, Treiman et al. (2006) reported that children appear to initially attend to small units (graphemes), before gradually showing an influence of surrounding graphemes when confronted with inconsistencies in pronunciation.

Thus, in the current study we focus on single grapheme to phonemes, or single phoneme to grapheme mappings in our inquiry of granularity and learning to read. In this way, we make no assumptions about a beginning reader’s knowledge of subword units or syllable structure, instead assuming all letters are created equal (whether vowels or consonants) and that the reading system must initially acquire knowledge of print patterns for GPC on its own, through experience with the print input. **Granularity** was, therefore, operationalised for each word as the difference between the number of letters (N_{letter}) and phonemes (N_{phon} ; i.e., $N_{letter} - N_{phon}$). For example, the granularity of the word *mince* ($N_{letter} = 5$, $N_{phon} = 4$) is 1, and the granularity of *thought* ($N_{letter} = 7$, $N_{phon} = 3$) is 4. A granularity of 0, hence, indicates that the word comprises of no multi-letter grapheme (e.g., *held*, *storm*).

The aim of this study was to systematically examine granularity related to the learning of GPC and word decoding. Theoretical accounts of the best representational units for learning to read have not been explicitly tested in the modelling literature to our knowledge. This, in turn, may inform instructional practices as to the best approaches for optimizing the learning curve, and results can be interpreted in terms of optimal child developmental trajectories and reading curricula (McKeown et al., 2017).

2 Method

2.1 Model Architecture

The model had four types of layers: orthographic, phonological, hidden, and clean-up (see Figure 1). The orthographic and phonological layers were each connected to a clean-up layer that mediated connections within the respective units, creating an attractor network that settles into a stable pattern over time (Harm and Seidenberg, 1999).

The orthographic layer was composed of 260 units, corresponding to 10 positions \times 26 possible letters. Words were coded as vowel-centred, such that the fourth slot was filled with the left-most vowel of a word (e.g., *mince* \rightarrow _ _ m i n c e _ _ _).

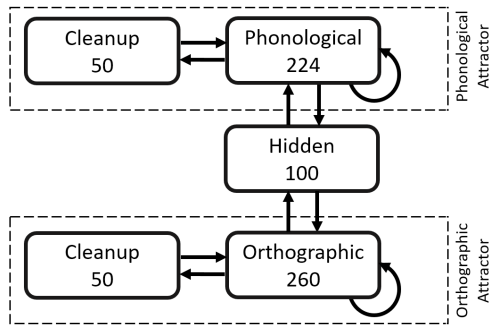


Figure 1: Architecture of the implemented reading model.

A word’s phonology was represented with nodes coding features of phonemes (8 positions \times 28 possible phonological features = 224 units). Pronunciation of each word was positioned with the vowel at the fourth slot (e.g., *mince* \rightarrow /_ _ m i n s _ _/). Each phoneme was encoded by a binary vector of 28 phonological features taken from PHOIBLE (Moran and McCloy, 2019), an online repository of cross-lingual phonological data. A list of phonemes and their respective phonological features used in the present work can be found in the Open Science Framework (OSF) repository for this project (<https://osf.io/hj96x/>).

While traditionally the problem of learning to read is conceptualized in terms of decoding unidirectionally from orthography to phonology, research suggests that children engage in spelling words simultaneously as they learn to decode. In addition, *feedback* sound-to-spelling relations are also informative in establishing mappings for reading. Thus, we implemented a new model with a bidirectional network architecture that connects orthographic-to-phonological and phonological-to-orthographic layers via the hidden units.

2.2 Training Procedure

The model was trained with a learning rate of 0.05 using a back-propagation through time (BPTT) algorithm with input integration and a time constant of 0.5 (Harm and Seidenberg, 1999; Plaut et al., 1996). Each word item was clamped and presented for six time ticks, and then in an additional six time ticks, the model was required to reproduce the target pattern of the word by the final 12th tick. The weight connections are updated based on cross-entropy error computed between the target and the actual activation of the output units.

Training proceeded in two distinct stages re-

flecting naturalistic child language development: (1) a pre-literacy training stage, in which the model was trained to learn the phonology-to-phonology mappings with an accuracy of 99%; and (2) a literacy training stage, in which the model was trained on both **decoding** (orthography-to-phonology) and **encoding** (phonology-to-orthography) tasks in a sequential manner. The pre-literacy stage of training was intended to mimic the fact that children develop oral skills through hearing and speaking long before learning to read.

Models were trained with a cumulative process of learning to encode and decode, whereby words with different granularity were introduced to the model in either an ascending or descending sequence. These two models were referred to as **small-to-large (SL)** and **large-to-small (LS)** from here onwards.

Words were first sorted with regard to their granularity, followed by a second-level sorting criterion to arrange words with the same granularity in order of decreasing frequency. The first batch of words in each training regime, therefore, comprises of high frequency words that are of either the smallest (e.g., *fix*, *lynx*) or largest (e.g., *bought*, *should*) granularity in the corpus. During training, words were sampled according to their frequency from the Word Frequency Guide (WFG) corpus (Zeno et al., 1995), and the resulting probability values were normalized over all words in the training set. Correspondingly, low frequency words had a lower probability of being presented to the model during training as compared to high frequency words [e.g., $P(yules) = 0.05$ vs. $P(of) = 0.97$].

2.2.1 Adaptive training

Teachers introduce written words progressively to their pupils, and regularly assess progress before introducing new words. Likewise, our model training introduced batches of 45 new words at a time. Importantly, a new batch of words was introduced only after model performance exceeded a criterion threshold of 70% combined accuracy for the decoding and encoding tasks on trained words — which included only words that the model had been trained on cumulatively up to the last training epoch. This tested the network success at reproducing the training set to which it had been progressively exposed, and allowed us to compare the rates of learning under different training regimes.

2.3 Testing Procedure

Two complementary tests are carried out every 100 training epochs: (1) a *total vocabulary test* which uses words from the entire corpus, regardless of whether they have been presented to the model in previous training phases; and (2) an *untrained pseudo-words test* which uses a fixed set of pronounceable and spellable monosyllabic non-words. This pseudo-word set is derived from previous empirical studies on developmental reading skills (Torgesen et al., 1999). Thus, with these tests we assess the network’s (1) transfer and (2) decoding abilities.

Because no learning occurs during testing, the same set of test words and non-words can be used routinely as novel testing items after 100 training epochs. This represents a considerable advantage with respect to behavioural longitudinal experiments, where successive test sessions can suffer from previous exposure effects.

Each test was administered twice, once in a decoding task and again in an encoding task. The **decoding task** activated the orthographic pattern for a given test word on the orthographic layer, say, *eye*, and measured the accuracy of the network to reproduced the corresponding target phonological word (/aɪ/) on the phonological layer. Conversely, the **encoding task** activated the phonological pattern for a given word on the phonology layer, say, /aɪ/, and measured the accuracy of the network to reproduced the corresponding target orthographic word (*eye*) on the orthographic layer.

Similar to the training procedure, each test word item was clamped and presented for six time samples, and then in an additional six time samples, the model was required to produce the target phonological/orthographic pattern of the word. An output was scored as correct when the target nodes were active with a value ≥ 0.75 , and concurrently the other nodes were inactive (≤ 0.25). Intermediate values were considered incorrect.

2.4 Corpus

All stimuli were monosyllabic American English words. The CHILDES database (MacWhinney, 2000), WFG corpus (Zeno et al., 1995), and the Phonemic Decoding Efficiency sub-test of the TOWRE (Torgesen et al., 1999) were used for pre-literacy training ($N = 5032$), literacy training ($N = 4394$), and pseudo-words testing ($N = 163$), respectively. The full list of words and their respec-

tive granularity and batch number can be found on OSF (<https://osf.io/hj96x/>).

To check whether frequency covaries with grain size, and may therefore confound the order effect, we conducted Spearman’s correlations across the training regime between batch number and mean log frequency per batch. This was done for each training order: **small-to-large (SL)** and **large-to-small (LS)**. Importantly, while batch number was significantly correlated with frequency for both training orders [SL: $r_s(96) = -0.43$, $p < .001$; LS: $r_s(96) = -0.30$, $p = .003$], the relation was in the same, negative direction in both cases — ensuring that frequency was not systematically tied to grain size. Rather, the result was from the second-level sorting by frequency in descending order.

To identify the possible relationship between the granularity and consistency of the mapping for the units to be learned, we calculated the decoding and encoding consistency measures to reflect how often the orthographic/phonological unit was spelled/pronounced in the same way as it was across all words (Berndt et al., 1987). The procedure required the conditional probabilities of GPCs and PGCs to be computed as they occur in the corpus [e.g., the probability of the grapheme *ew* being pronounced as /o/ is, $P(/o/|ew) = 0.057$].

We then derived a *composite consistency score* to account for the two measures (decoding and encoding), with a higher score representing higher overall bi-directional word consistency. Consistency was found to correlate negatively with granularity increases [SL: $r_s(96) = -0.69$, $p < .001$; LS: $r_s(96) = 0.71$, $p < .001$], indicating that words with smaller granularity were more consistent.

3 Results

At the time of writing, each model had been trained on 67 out of 98 batches of words (or 3015 out of 4394 unique words). While incomplete, our preliminary observations suggest a clear difference in rate of learning across the two training regimes.

Results are summarised in Figure 2, and show that under a training regime that introduces written words in batches progressively from small-to-large granularity, the network exhibited an early advantage in reading acquisition as compared to a regime introducing written words from large-to-

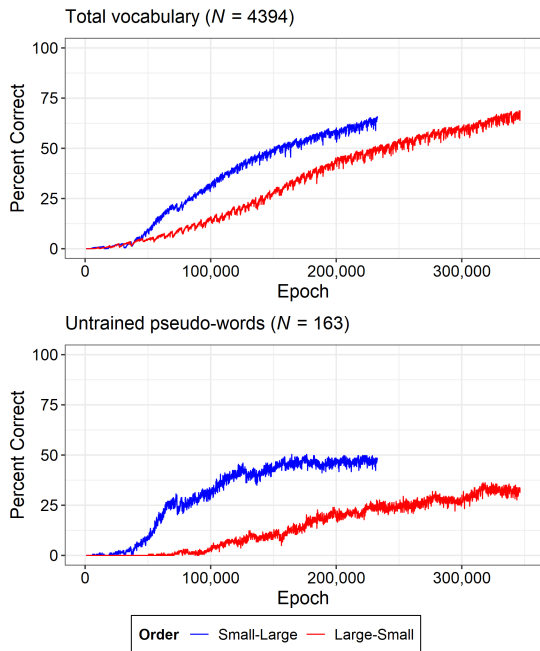


Figure 2: Models’ accuracy in the encoding task when tested against the full vocabulary and a set of pseudo-words. Test results compare models trained on two ordering regimes based on granularity of the orthography-phonology mappings.

small granularity.

The two types of repeated tests served to evaluate the accuracy of phonological output for: (a) total vocabulary (including trained and untrained words) and (b) pseudo-words. Both tests measured the ability of the networks to generalize to unseen but orthographically legal strings (see Figure 2). Specifically, the SL and LS models took 232800 and 346400 epochs, respectively, to reach the criterion threshold of 70% accuracy for all 67 batches of words that were introduced cumulatively over time. Apart from reaching the criterion threshold earlier, the SL model also performed better than the LS model in pseudo-words test (47.85% vs. 33.13%) at the end of preliminary training.

4 Discussion

As the process of learning to read requires picking up and internalizing representational units of print associated with sound, the ordering of training input to the reading system becomes paramount. How best to order input and maximize learning efficiency has been debated in the literacy education field. This study capitalizes on a computational modelling approach to this issue, using a highly

controlled context without the ethical concerns of human learning studies. Directly contrasting the effects of two literacy training regimes differing in granularity order, the simulation results support better learning with smaller, less complex orthographic units, as predicted from corpus-based research (Vousden, 2008). At training stages comprising of 3015 words, we found that the model initially trained with words of smaller granularity performed and generalized to pseudo-words better than the model trained with larger granularity. The LS model did require significantly more training epochs to reach the same performance as the SL model.

Essentially, when children learn to read, they must navigate the structure of their language and its writing system. Granularity and consistency are important aspects of this structure, and both impact reading performance. Adult readers are slower to identify letters within a multi-letter grapheme (Smith and Monaghan, 2011; Rey et al., 2000), suggesting that graphemes are functional reading units. Furthermore, Rastle and Coltheart (1998) found that naming latencies were slower for pseudo-words with, as compared to without, multi-letter graphemes. Adult word naming and lexical decision are also faster for consistent words (Andrews, 1982; Jared, 1997; Jared, 2002), and consistent words are more accurately read and spelled by children (Alegria and Mousty, 1996; Lété et al., 2008; Weekes et al., 2006).

Granularity and consistency have been regarded to be associated (Treiman et al., 1995), and our corpus analysis revealed this as well — monosyllabic English words of smaller granularity tend to be more consistent than words with larger granularity. This relationship indicates that granularity and consistency may not be entirely disentangled, at least for English. With this in mind, the SL model was first exposed to words of smaller granularity that were also more consistent in their GPC and PGC (phoneme-grapheme correspondence) mappings. Thus consistency and granularity may be two sides of the same coin, and when manipulated they could lead to faster or slower rates of convergence. Importantly, the current model included bidirectional links between orthographic and phonological units, simulating the real-world scenario that children acquire decoding and encoding skills simultaneously.

These findings have implications for educa-

tional planning for early literacy. In particular, our pilot simulation provides preliminary evidence on the potential utility of manipulating the order of training in terms of word granularity to unveil facilitative effects on literacy acquisition. Reading instruction can consider the early acquisition of words with smaller granularity, or more consistency. However, we note that the present findings are based on the analysis of monosyllabic words only and should not be generalized to multisyllabic words directly. Future work can consider using models that are capable of reading multisyllabic words (Perry et al., 2010), or explore the link between granularity and consistency across languages that are either less or more orthographically transparent.

Acknowledgments

Support comes from the Education Research Funding Programme of the National Institute of Education (NIE), Nanyang Technological University, Singapore, grant #OER0417OBA.

References

- Jesus Alegria and Philippe Mousty. 1996. The development of spelling procedures in french-speaking, normal and reading-disabled children: Effects of frequency and lexicality. *Journal of experimental child psychology*, 63(2):312–338.
- Sally Andrews. 1982. Phonological recoding: Is the regularity effect consistent? *Memory & Cognition*, 10(6):565–575.
- Rita S. Berndt, James A. Reggia, and Charlotte C. Mitchum. 1987. Empirically derived probabilities for grapheme-to-phoneme correspondences in english. *Behavior Research Methods, Instruments, & Computers*, 19(1):1–9.
- Morten H. Christiansen, Stephen A. Hockema, and Luca Onnis. 2006. Using phoneme distributions to discover words and lexical categories in unsegmented speech. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 28.
- Jean Ecalte, Annie Magnan, and Caroline Calmus. 2009. Lasting effects on literacy skills with a computer-assisted learning using syllabic units in low-progress readers. *Computers & Education*, 52(3):554–561.
- Linnea C. Ehri, Simone R. Nunes, Steven A. Stahl, and Dale M. Willows. 2001. Systematic phonics instruction helps students learn to read: Evidence from the national reading panel’s meta-analysis. *Review of Educational Research*, 71(3):393–447.
- Linnea C. Ehri. 2005. Learning to read words: Theory, findings, and issues. *Scientific Studies of reading*, 9(2):167–188.
- Nick C. Ellis, Miwa Natsume, Katerina Stavropoulou, Lorenc Hoxhallari, Victor H. P. Van Daal, Nicoletta Polyzoe, Maria-Louisa Tsipa, and Michalis Petalas. 2004. The effects of orthographic depth on learning to read alphabetic, syllabic, and logographic scripts. *Reading research quarterly*, 39(4):438–468.
- Elena Florit and Kate Cain. 2011. The simple view of reading: Is it valid for different types of alphabetic orthographies? *Educational Psychology Review*, 23(4):553–576.
- George K. Georgiou, Rauno Parrila, and Timothy C. Papadopoulos. 2008. Predictors of word decoding and reading fluency across languages varying in orthographic consistency. *Journal of Educational Psychology*, 100(3):566.
- Michael W. Harm and Mark S. Seidenberg. 1999. Phonology, reading acquisition, and dyslexia: insights from connectionist models. *Psychological review*, 106(3):491.
- Debra Jared. 1997. Spelling–sound consistency affects the naming of high-frequency words. *Journal of Memory and Language*, 36(4):505–529.
- Debra Jared. 2002. Spelling-sound consistency and regularity effects in word naming. *Journal of Memory and Language*, 46(4):723–750.
- Bernard Lété, Ronald Peereman, and Michel Fayol. 2008. Consistency and word-frequency effects on spelling among first-to fifth-grade french children: A regression-based study. *Journal of Memory and Language*, 58(4):952–977.
- Brian MacWhinney. 2000. *The CHILDES project: The database*, volume 2. Psychology Press.
- Genevieve McArthur, Saskia Kohnen, Kristy Jones, Philippa Eve, Erin Banales, Linda Larsen, and Anne Castles. 2015. Replicability of sight word training and phonics training in poor readers: a randomised controlled trial. *PeerJ*, 3:e922.
- Margaret G. McKeown, Paul D. Deane, and Ren R. Lawless. 2017. *Vocabulary assessment to support instruction: Building rich word-learning experiences*. Guilford Publications.
- Martial Mermillod, Patrick Bonin, Alain Méot, Ludovic Ferrand, and Michel Paindavoine. 2012. Computational evidence that frequency trajectory theory does not oppose but emerges from age-of-acquisition theory. *Cognitive Science*, 36(8):1499–1531.
- Padraic Monaghan and Andrew W. Ellis. 2010. Modeling reading development: Cumulative, incremental learning in a computational model of word naming. *Journal of Memory and Language*, 63(4):506–525.

- Steven Moran and Daniel McCloy, editors. 2019. *PHOIBLE 2.0*. Jena: Max Planck Institute for the Science of Human History.
- National Reading Panel. 2001. *Report of the National Reading Panel: Teaching Children to Read (Report NIH Pub. No. 00-4769)*. National Institutes of Health.
- Richard K. Olson and Barbara W. Wise. 1992. Reading on the computer with orthographic and speech feedback. *Reading and Writing*, 4(2):107–144.
- Conrad Perry, Johannes C. Ziegler, and Marco Zorzi. 2010. Beyond single syllables: Large-scale modeling of reading aloud with the connectionist dual process (cdp++) model. *Cognitive psychology*, 61(2):106–151.
- Conrad Perry, Marco Zorzi, and Johannes C. Ziegler. 2019. Understanding dyslexia through personalized large-scale computational models. *Psychological science*, 30(3):386–395.
- David C. Plaut, James L. McClelland, Mark S. Seidenberg, and Karalyn Patterson. 1996. Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological review*, 103(1):56.
- Stephen C. Pritchard, Max Coltheart, Eva Marinus, and Anne Castles. 2016. Modelling the implicit learning of phonological decoding from training on whole-word spellings and pronunciations. *Scientific studies of reading*, 20(1):49–63.
- Kathleen Rastle and Max Coltheart. 1998. Whammy and double whammy: Length effects in nonword naming. *Psychonomic Bulletin and Review*, 5:277–282.
- Keith Rayner, Barbara R. Foorman, Charles A. Perfetti, David Pesetsky, and Mark S Seidenberg. 2001. How psychological science informs the teaching of reading. *Psychological science in the public interest*, 2(2):31–74.
- Arnaud Rey, Johannes C. Ziegler, and Arthur M. Jacobs. 2000. Graphemes are perceptual reading units. *Cognition*, 75(1):B1–B12.
- Jay G. Rueckl. 2016. Toward a theory of variation in the organization of the word reading system. *Scientific Studies of Reading*, 20(1):86–97.
- Jenny R. Saffran, Elissa L. Newport, Richard N. Aslin, Rachel A. Tunick, and Sandra Barrueco. 1997. Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological science*, 8(2):101–105.
- Alastair C. Smith and Padraic Monaghan, 2011. *What are the functional units in reading? Evidence for statistical variation influencing word processing*, volume 20, pages 159–172. World Scientific. Proceedings of the 12th Neural Computation and Psychology Workshop.
- Ross A. Thompson. 2009. Doing what doesn't come naturally. *Zero to Three Journal*, 30(2):33–39.
- Carole Torgerson, Greg Brooks, and Jill Hall. 2006. *A systematic review of the research literature on the use of phonics in the teaching of reading and spelling*. DfES Publications Nottingham.
- Joseph K. Torgesen, Richard K. Wagner, Carol A. Rashotte, Elaine Rose, Patricia Lindamood, Tim Conway, and Cyndi Garvan. 1999. Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology*, 91(4):579.
- Rebecca Treiman, John Mullennix, Ranka Bijeljac-Babic, and E. Daylene Richmond-Welty. 1995. The special role of rimes in the description, use, and acquisition of english orthography. *Journal of Experimental Psychology: General*, 124(2):107.
- Rebecca Treiman, Brett Kessler, Jason D. Zevin, Suzanne Bick, and Melissa Davis. 2006. Influence of consonantal context on the reading of vowels: Evidence from children. *Journal of Experimental Child Psychology*, 93(1):1–24.
- Janet I. Vousden. 2008. Units of english spelling-to-sound mapping: a rational approach to reading instruction. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 22(2):247–272.
- Brendan S. Weekes, Anne E. Castles, and Robert A. Davies. 2006. Effects of consistency and age of acquisition on reading and spelling among developing readers. *Reading and Writing*, 19(2):133–169.
- Susan Zeno, Stephen H. Ivens, Robert T. Millard, and Raj Duvvuri. 1995. *The educator's word frequency guide*. Touchstone Applied Science Associates.
- Johannes C. Ziegler and Usha Goswami. 2005. Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychological bulletin*, 131(1):3.