# FRONTIERS IN THE ACQUISITION OF LITERACY

EDITED BY : Claire M. Fletcher-Flinn
PUBLISHED IN : Frontiers in Psychology

**frontiers** Research Topics

## About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: **researchtopics@frontiersin.org**

# FRONTIERS IN THE ACQUISITION OF LITERACY

Topic Editor:
**Claire M. Fletcher-Flinn,** University of Auckland, New Zealand

Learning to read, and to spell are two of the most important cultural skills that must be acquired by children, and for that matter, anyone learning a second language. We are not born with an innate ability to read. A reading system of mental representations that enables us to read must be formed in the brain. Learning to read in alphabetic orthographies is the acquisition of such a system, which links mental representations of visual symbols (letters) in print words, with pre-existing phonological (sound) and semantic (comprehension) cognitive systems for language.

Although spelling draws on the same representational knowledge base and is usually correlated with reading, the acquisition processes involved are not quite the same. Spelling requires the sequential production of letters in words, and at beginning levels there may not be a full degree of integration of phonology with its representation by the orthography. Reading, on the other hand, requires only the recognition of a word for pronunciation. Hence, spelling is more difficult than reading, and learning to spell may necessitate more complete representations, or more conscious access to them.

The learning processes that children use to acquire such cognitive systems in the brain, and whether these same processes are universal across different languages and orthographies are central theoretical questions. Most children learn to read and spell their language at the same time, thus the co-ordination of these two facets of literacy acquisition needs explication, as well as the effect of different teaching approaches on acquisition. Lack of progress in either reading and/or spelling is also a major issue of concern for parents and teachers necessitating a cross-disciplinary approach to the problem, encompassing major efforts from researchers in neuroscience, cognitive science, experimental psychology, and education.

The purpose of this Research Topic is to summarize and review what has been accomplished so far, and to further explore these general issues. Contributions from different perspectives are welcomed and could include theoretical, computational, and empirical works that focus on the acquisition of literacy, including cross-orthographic research.

# Table of Contents

# Editorial: Frontiers in the acquisition of literacy

Claire M. Fletcher-Flinn*

*School of Psychology, The University of Auckland, Auckland, New Zealand*

Reading and writing are fundamental to full participation in our societies, yet how children acquire such a large system of interconnected representations of print words, their meanings, and phonology in the brain remains unclear. As the teaching of literacy takes up a large proportion of classroom time in the early years, increasing knowledge about children's learning processes should result in better approaches to the teaching of reading and spelling. These insights would be particularly useful from a clinical perspective for the treatment of developmental disabilities, such as dyslexia and dysgraphia.

Important questions need to be addressed, and given the many different influences and overlapping processes on literacy learning, the answers are not straightforward. What are the learning processes? Are they the same across different orthographies, or do different orthographies require different skills and learning processes? What is the relationship between reading and spelling? How do they interact and augment each other? What is the effect of different teaching approaches on children's emerging reading system? Do early reading comprehension problems disappear over time? What are the predictors of children's reading attainment?

This E-Book responds to some of these general questions in the form of original research and opinion articles. Taken together the articles contribute new perspectives and challenges to current reading acquisition theories, and present new research on early reading skills, reading instruction, and spelling.

Thompson (2014a) claims that most reading acquisition theories are limited by their specification of letter-sound requirements to a particular class of teaching approaches. Acknowledging such limitations is an important step in the development of reading acquisition theories that are potentially more useful. Fletcher-Flinn (2014) merges ideas from dynamic systems theory with developmental data from a precocious reader. Reference is made to Knowledge Systems theory, which offers a more varied range of theoretical applicability.

Share (2014) asserts that there is a general belief in the superiority of alphabetic writing systems that has hindered progress in the development of a universal model of learning to read. Some counterevidence is presented, and in the context of a more general question on "optimality," Share proposes a universal model of reading based on a broader novice-to-expert dualism. Nag (2014) maintains that specification of the learning mechanisms involved in reading akshara units, the symbols used in many writing systems (alphasyllabaries) of Southern Asia, present a challenge to alphabet-based theories of reading acquisition. The akshara units, unlike alphabet letters, map onto multiple sublexical levels of phonology determined by context. In order to ensure the development of an inclusive reading science, and a more comprehensive and universal theory of literacy learning, Nag argues that consideration of these orthographic-specific features of reading are needed. At the same time, it is possible that some general cognitive features of information processing, as they relate to reading acquisition, may be orthography-independent. Colé et al. (2014) provides evidence that early progress in young French children in both word reading and reading comprehension was related to cognitive flexibility in the coordination of orthographic, phonological, and semantic information.

How early, and in what form does SES as a distal predictor of reading achievement manifest itself? Robins et al. (2014) found that lower SES parents of preschoolers asked fewer questions about letters, and focused more on memorizing sequences of the alphabet than higher SES parents. The persistence of a conversational focus on letters within the child's name also differentiated the groups. These differences could put lower SES children at a disadvantage when entering school, resulting in poor rates of literacy. Tse and Nicholson (2014) addressed the performance gap of low SES children in New Zealand schools with an intervention comparing three teaching approaches: Big Book (shared reading), explicit instruction in phoneme awareness and phonics, and a combined approach. The latter produced better results on a range of literacy measures compared with the combined averaged scores of the other two groups. Thompson (2014b) took the opportunity to comment on ambiguities that are often unrecognized but affect the validity of such intervention research, and Nicholson and Tse (2015) provide a rebuttal. These discussions are thoughtful contributions to methodological issues in intervention research.

How do fluent readers distinguish between words that look similar but whose meaning differ? Using masked form priming, Bhide et al. (2014) found no evidence that increases in print vocabulary size predicted precise orthographic representations, and suggested spelling skill might be more important. Ouellette and Tims (2014) examined whether this "spelling advantage" might be due to the motoric component of writing. There was no effect of modality (printing or typing) for Grade two children,

suggesting that stored orthographic detail is independent of input. Of interest, pre-existing keyboard skills affected learning.

With regard to literacy impairments, Critten et al. (2014) found no difference for the spelling of words with inflectional morphemes by children with specific language impairment (SLI) and spelling-matched controls. However, the SLI group was less accurate when spelling words with derivational morphemes. The authors conclude that this indicates a specific impairment when making orthographic and phonological shifts from base words. This outcome has useful teaching applications for SLI children. Ricketts et al. (2014) showed that children identified at 9 years with poor reading comprehension had lower educational achievement at 11 and 16 years than a reading (decoding and comprehension), and non-verbal reasoning matched group, and were below national performance norms. They point out that these children are at risk from an early age of a compromised future with regard to further training and employment.

This E-Book contains an excellent collection of cutting edge scientific research and opinions at the frontiers of literacy acquisition. The reader will find new perspectives and questions derived from the reported findings, and these can serve as a springboard for new research in this field.

## Acknowledgments

## References

Bhide, A., Schlaggar, B. L., and Barnes, K. A. (2014). Developmental differences in masked form priming are not driven by vocabulary growth. *Front. Psychol.* 5:667. doi: 10.3389/fpsyg.2014.00667

Colé, P., Duncan, L. G., and Blaye, A. (2014). Cognitive flexibility predicts early reading skills. *Front. Psychol.* 5:565. doi: 10.3389/fpsyg.2014.00565

Critten, S., Connelly, V., Dockrell, J. E., and Walter, K. (2014). Inflectional and derivational morphological spelling abilities of children with specific language impairment. *Front. Psychol.* 5:948. doi: 10.3389/fpsyg.2014.00948

Fletcher-Flinn, C. M. (2014). Learning to read as the formation of a dynamic system: evidence for dynamic stability in phonological recoding. *Front. Psychol.* 5:660. doi: 10.3389/fpsyg.2014.00660

Nag, S. (2014). Alphabetism and the science of reading: from the perspective of the akshara languages. *Front. Psychol.* 5:866. doi: 10.3389/fpsyg.2014.00866

Nicholson, T., and Tse, L. (2015). What are the criteria for a good intervention study? Response: "Unrecognized ambiguities in validity of intervention research: an example on explicit phonics and text-centered teaching." *Front. Psychol.* 6:508. doi: 10.3389/fpsyg.2015.00508

Ouellette, G., and Tims, T. (2014). The write way to spell: printing vs. typing effects on orthographic learning. *Front. Psychol.* 5:117. doi: 10.3389/fpsyg.2014.00117

Ricketts, J., Sperring, R., and Nation, K. (2014). Educational attainment in poor comprehenders. *Front. Psychol.* 5:445. doi: 10.3389/fpsyg.2014.00445

Robins, S., Ghosh, D., Rosales, N., and Treiman, R. (2014). Letter knowledge in parent–child conversations: differences between families differing in socio-economic status. *Front. Psychol.* 5:632. doi: 10.3389/fpsyg.2014.00632

Share, D. (2014). Alphabetism in reading science. *Front. Psychol.* 5:752. doi: 10.3389/fpsyg.2014.00752

Thompson, G. B. (2014a). Discovering and accounting for limitations in applications of theories of word reading acquisition. *Front. Psychol.* 5:579. doi: 10.3389/fpsyg.2014.00579

Thompson, G. B. (2014b). Unrecognized ambiguities in validity of intervention research: an example on explicit phonics and text-centered teaching. *Front. Psychol.* 5:1535. doi: 10.3389/fpsyg.2014.01535

Tse, L., and Nicholson, T. (2014). The effect of phonics-enhanced Big Book reading on the language and literacy skills of 6-year-old pupils of different reading ability attending lower SES schools. *Front. Psychol.* 5:1222. doi: 10.3389/fpsyg.2014.01222

# Discovering and accounting for limitations in applications of theories of word reading acquisition

## G. Brian Thompson *

*School of Educational Psychology and Pedagogy, Victoria University of Wellington, Wellington, New Zealand*
*Correspondence: brian.thompson@vuw.ac.nz*

More attention to the discovery of the limitations of current theories of word reading acquisition would enable progress in development of theories with a wider and more varied range of valid and useful applications. This general opinion is illustrated here with work that makes such an attempt.

There have been recent occasional attempts to apply computational connectionist models of adult word reading in simulations of children's normal progress in word and pseudoword reading (Hutzler et al., 2004; Powell et al., 2006), but they failed unless modifications were made that included adding "context-free" letter-sound correspondences to the initial training of the model. Taught phonics sounds for letters are such, as they are not bound to features within a word, such as position and/or the context of other (adjacent or otherwise) letter-sound correspondences of the word. Adding the phonic sounds was justified as a representation of the way children learnt because it was how they were taught reading. This introduces a major potential limitation in application of the model, in so far as it was improved for only one type of teaching, that with phonics. A new multiple-route theory (Grainger et al., 2012) of learning to read words has been proposed which may appear to avoid that problem but much of the learning of the beginner reader is modeled as in the theory of Share (1995). This, however, also requires full knowledge of "context-free" letter sounds for the initial development of word reading (Share, p. 164), as does the widely recognized theory of Ehri (1999, 2005, 2012). The illustration for my opinion focuses on the discovery of

the limitations of this feature common to these theories, and on development of theory that accounts for evidence beyond the limitations.

## FAILURE IN APPLICATION OF LETTER-SOUND REQUIREMENTS FOR DEVELOPMENT

There has been a claim (Thompson and Johnston, 1993; Ramus, 2004) of potential limitations of testing theories of reading acquisition on mainly those children receiving teaching in just the tradition in which the theory was developed. We cite data of participants from a teaching tradition very different from that in which the Share and Ehri theories were developed. A tradition has been common across New Zealand (since the late 1960's) in which neither context-free letter sound knowledge or explicit phonics were taught, and the emphasis was on text-centered teaching (Thompson, 1993) with individualized provision of multiple brief story texts at finely adjusted difficulty levels. In that country a sample with normal word reading progress, and 9 months of reading instruction, obtained a mean accuracy of 83% for names of the lower-case alphabet letters, 76% accuracy for the context-free phonic letter sounds of the 9 letters (*b, d, j, k, o, p, t, v, z*) with a name having the initial pronunciation element compatible with that phonic letter sound, but 51% accuracy for the sounds of the other 16 letters without such compatibility with the letter name. (Calculated from Thompson et al., 1999, that specifies the range of pronunciations obtained, and those acceptable as letter sounds, among these children who were not taught them. The letter *q* was

not included due to the high rate of visual confusion with *p*.) A sample of 11-year-olds with normal progress (relative to both local and U.S. norms) in the same school system and tradition of teaching obtained a mean accuracy of 99% for the letter names, 90% for the sounds of the 9 letters compatible with the letter name, but 62% for the sounds of the 16 not compatible (Fletcher-Flinn and Thompson, 2004, p. 315). Moreover, a sample of adult university students with above average reading skill (relative to U.S. norms), who as children had been taught in that tradition in this school system, showed a similar result (Thompson et al., 2009). The conclusion is that successful readers in this teaching tradition did not meet the requirement of the theories (cited above) for full knowledge of context-free letter sounds for success in acquiring word reading.

## WHAT ACCOUNTS FOR THIS FAILURE IN APPLICATION?

Aside from letter-sound knowledge, it may be that knowledge of letter identities *per se* can be acquired within the context of words, as children begin learning to read. In a series of studies relating to letter identities, 5-year-old children, after 9 months of reading instruction with normal progress, could cope fully with the substitution of upper case for lower case in their knowledge of identities of letters out of the context of words. The children responded to upper-case letters comprising those eight that were visually dissimilar to the corresponding lower-case form (Aa, Bb, Dd, Ee Gg, Hh, Nn, Rr) with accuracy as high as the lower case and very close to ceiling (Thompson and Johnston, 2007).

But, in reading identification of familiar print words comprising these upper-case letters, the children were much less accurate than their high accuracy for the same words in lower case. Moreover, these results were replicated across children in the New Zealand teaching tradition and that in Scotland with explicit phonics. Other evidence in the study showed that these children were using some form of letter identities to read the words, rather than global visual features of the words. In another study, an experiment with training initially unknown and similarly constructed lower case words had similar effects in which gains in lower case accuracy were large but transfer gains for upper case were much smaller. This was despite equal proficiency in knowledge of identities of the letters in the two case forms when out of the context of words (Thompson et al., 2008). Hence, the processes for letter identities that are bound to word context for identification of words can function differently from those for "context-free" letter identities (Thompson, 2009).

Such a difference in processes may also be expected to occur for the beginner reader's use of letter-sound knowledge. In a sample of normal-progress New Zealand 5- and 6-year-olds there was evidence they had some knowledge of letter-sound relations that was bound to a sublexical function of their emerging *reading vocabulary* (which is the stored knowledge of the letters of the word, i.e., the lexical orthographic representation, along with the associated phonological and lexical-semantic representations). The children's relative accuracy of letter-sound relations in their reading of simple pseudowords (e.g., *ob, bu, et,* . . . that simulate new print words) was predicted from the distribution of occurrence of within-word positions of these sublexical relations among the vocabularies of the children's reading books. For example, these small vocabularies rarely included words with a final *b* letter, although an initial *b* was common, whereas *t* in both final and initial positions was common. The children's pseudoword reading accuracy reflected these (and similar) distributions of the positional sublexical letter-sound relations in their print word experience. They gave no segmented pronunciation of component

letters, contrary to what would be expected in an explicit phonics response. Moreover, a replication of the task was conducted, and also confirmation by a successful prediction of positive effects (relative to controls) on pseudoword reading accuracy from experimental training that introduced words with final *b* into the children's reading vocabularies (Thompson et al., 1996). For children receiving explicit phonics instruction there has not been a complete replication involving the training experiment. The pseudoword reading task, however, was presented to such a sample of children in the U.S. who were of a comparable reading level, with the result showing no significant within-word positional effects (Fletcher-Flinn et al., 2004). Context-free letter sounds have no coding of position or other contextual feature of words. Hence, this result was expected for their phonological decoding of the pseudowords, if they were often using taught context-free letter sounds, rather than knowledge of letter-sound relations bound to a sublexical function of their emerging reading vocabularies.

These results are consistent with the Knowledge Sources theory that was developed to include different sources of knowledge to account for the acquisition of word reading in a tradition of text-centered teaching as well as a tradition with explicit phonics (Thompson et al., 1996; Fletcher-Flinn and Thompson, 2004; Thompson and Fletcher-Flinn, 2006, 2012). In this theory, as soon as the child, with support from parent or teacher, has acquired reliable reading of a few words, and has attended to "the relationship in which letters of words often match sound units of the spoken word" (Thompson and Fletcher-Flinn, 2012, p. 254), they can independently extract from their emerging reading vocabulary some letter-sound information coded with sublexical features. This coding can commence for position within the word and then expand to include the contexts of other letter-sound correspondences within the word (Thompson et al., 1996; Thompson and Fletcher-Flinn, 2006). Such sublexical information is available for a frequently implicit form of "phonological recoding" (involving generation of responses to new print words) that does not require full knowledge of context-free letter sounds, as

in the theories of Share or Ehri. This form of phonological recoding assists the child in acquiring representations of new or unfamiliar print words, thus extending the child's reading vocabulary, which in turn is a basis for extracting more advanced sublexical letter-sound knowledge. It implies a recursive process that can start very early in the child's development of reading. The theory, however, also accounts for children's successes in using, as in explicit phonics, the other form of phonological recoding that is initially dependent on full knowledge of context-free letter sounds.

There are other studies, which examine samples of beginner readers who have reached the same developmental level of word reading but have differences in processes of word reading acquisition according to whether they are receiving teaching with explicit phonics or text-centered teaching (Connelly et al., 2009). In a study involving three countries (Thompson et al., 2008) the teaching with explicit phonics produced beginner readers who had much higher accuracy in pseudoword reading than those receiving text-centered teaching, although both had reached the same level of word reading accuracy. Nevertheless, for reading text, the context in which most useful word reading occurs, teaching with explicit phonics produced a much slower speed of text reading (for equal word reading accuracy). This was apparently not due to their slower responses to the unfamiliar words but mainly to their lower level of practice with words in text, and hence with the lexical-semantic and syntactic relations among the words. It is consistent with the Knowledge Sources theory to infer that the corresponding greater number of exposures to print words from text, along with the associated orthographic-phonological, orthographic-lexical, and lexical-semantic/syntactic relations, reduces the need for developing a high level of expertise in a form of phonological recoding initially based on knowledge of context-free letter sounds.

The specific focus here has been on success and failure of several theories in accounting for differences in beginner learning processes arising from varied traditions for initial teaching of reading. Knowledge Sources theory, however, has been discovered to have valid applications

beyond that. The theory has been applied to a form of alphabetic orthography very different from English. Within Japanese hiragana there is a secondary phonemic function for which there are 36 *yoo-on* symbols, which are formed from some of the basic symbols that otherwise represent syllables (Coulmas, 2003). Beyond the application limits of other theories, Knowledge Sources theory accounted for results from training experiments on the initial learning of both Japanese beginners and second-language learners, as well as evidence from skilled hiragana readers on the generalization limits of their implicit knowledge of the formation principle for this phoneme representation within hiragana (Fletcher-Flinn et al., 2014). The theory has also been applied to the case of a 3-year-old precocious reader. Beyond the limitations of the other theories, the recursive learning processes that use sublexical information from the child's emerging reading vocabulary accounted for this child's underdeveloped context-free letter sounds (Fletcher-Flinn and Thompson, 2000, 2004).

By discovery of one limitation of some current theories as applied to children in a teaching tradition outside that in which those theories were developed, an alternative theory was formed that offers a tested account of reading acquisition with a wider and more varied range of applications. This is just one illustration. Other limitations await discovery in these and other theories.

## REFERENCES

Connelly, V., Thompson, G. B., Fletcher-Flinn, C. M., and McKay, M. F. (2009). "Does the type of reading instruction have an influence on how readers process print?" in *Contemporary Perspectives on Reading and Spelling*, eds C. Wood and V. Connelly (Abingdon, Oxon: Routledge), 239–253.

Coulmas, F. (2003). *Writing Systems: An Introduction to Their Linguistic Analysis*. Cambridge: Cambridge University Press.

Ehri, L. C. (1999). "Phases of development in learning to read words," in *Reading Development and the Teaching of Reading*, eds J. Oakhill and R. Beard (Oxford: Blackwell), 79–108.

Ehri, L. C. (2005). "Development of sight word reading: phases and findings," in *The Science of Reading: A Handbook*, eds M. J. Snowling and C. Hulme (Oxford: Blackwell), 135–154.

Ehri, L. C. (2012). "Why is it important for children to begin learning to read in kindergarten?" in *Contemporary Debates in Childhood Education and Development*, eds S. Suggate and E. Reese (Abington, Oxon: Routledge), 171–180.

Fletcher-Flinn, C. M., Shankweiler, D., and Frost, S. J. (2004). Coordination of reading and spelling in early literacy development: an examination of the discrepancy hypothesis. *Read. Writ.* 17, 617–644. doi: 10.1023/B:READ.0000044297.85675.f5

Fletcher-Flinn, C. M., and Thompson, G. B. (2000). Learning to read with underdeveloped phonemic awareness but lexicalized phonological recoding: a case study. *Cognition* 74, 177–208. doi: 10.1016/S0010-0277(99)00072-4

Fletcher-Flinn, C. M., and Thompson, G. B. (2004). A mechanism of implicit lexicalized phonological recoding used concurrently with underdeveloped explicit letter-sound skills in both precocious and normal reading development. *Cognition* 90, 303–335. doi: 10.1016/S0010-0277(03)00162-8

Fletcher-Flinn, C. M., Thompson, G. B., Yamada, M., and Meissel, K. (2014). Learning of a formation principle for the secondary phonemic function of a syllabic orthography. *Read. Writ.* 27, 875–904. doi: 10.1007/s11145-013-9473-0

Grainger, J., Lété, B., Bertand, D., Dufau, S., and Ziegler, J. G. (2012). Evidence for multiple routes in learning to read. *Cognition* 123, 280–292. doi: 10.1016/j.cognition.2012.01.003

Hutzler, F., Ziegler, J. C., Perry, C., Wimmer, H., and Zorzi, M. (2004). Do current connectionist learning models account for reading development in different languages? *Cognition* 91, 273–296. doi: 10.1016/j.cognition.2003.09.006

Powell, D., Plaut, D., and Funnell, E. (2006). Does the PMSP connectionist model of single word reading learn to read in the same way as a child? *J. Res. Read.* 29, 299–250. doi: 10.1111/j.1467-9817.2006.00300.x

Ramus, F. (2004). "The neural basis of reading acquisition," in *The Cognitive Neurosciences, 3rd Edn.*, ed M. S. Gazzaniga (Cambridge, MA: MIT Press), 815–824.

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

Thompson, G. B. (1993). "Appendix: reading instruction for the initial years in New Zealand Schools," in *Reading Acquisition Processes*, eds G. B. Thompson, W. E. Tunmer, and T. Nicholson (Clevedon, Avon: Multilingual Matters), 148–154.

Thompson, G. B. (2009). The long learning route to abstract letter units. *Cogn. Neuropsychol.* 26, 50–69. doi: 10.1080/02643290802200838

Thompson, G. B., Connelly, V., Fletcher-Flinn, C. M., and Hodson, S. J. (2009). The nature of skilled adult reading varies with type of instruction in childhood. *Mem. Cogn.* 37, 223–234. doi: 10.3758/MC.37.2.223

Thompson, G. B., Cottrell, D. S., and Fletcher-Flinn, C. M. (1996). Sublexical orthographic-phonological relations early in the acquisition of reading: the knowledge sources account. *J. Exp. Child Psychol.* 62, 190–222. doi: 10.1006/jecp.1996.0028

Thompson, G. B., and Fletcher-Flinn, C. M. (2006). "Lexicalised implicit learning in reading acquisition: the knowledge sources theory," in *Cognition and Language: Perspectives from New Zealand*, eds C. M. Fletcher-Flinn and G. M. Haberman (Brisbane,QLD: Australian Academic Press), 141–156.

Thompson, G. B., and Fletcher-Flinn, C. M. (2012). "Toward better teaching: revising the fundamentals of learning to read," in *Contemporary Debates in Childhood Education and Development*, eds S. Suggate and E. Reese (Abington, Oxon,: Routledge), 250–260.

Thompson, G. B., Fletcher-Flinn, C. M., and Cottrell, D. S. (1999). Learning correspondences between letters and phonemes without explicit instruction. *Appl. Psycholinguistics* 20, 21–50. doi: 10.1017/S0142716499001022

Thompson, G. B., and Johnston, R. S. (1993). "The effects of type of instruction on processes of reading acquisition," in *Reading Acquisition Processes,* eds G. B. Thompson, W. E. Tunmer, and T. Nicholson (Clevedon, Avon: Multilingual Matters), 74–90.

Thompson, G. B., and Johnston, R. S. (2007). Visual and orthographic information in learning to read and the influence of phonics instruction. *Read. Writ.* 20, 859–884. doi: 10.1007/s11145-007-9050-5

Thompson, G. B., McKay, M. F., Fletcher-Flinn, C. M., Connelly, V., Kaa, R. T., and Ewing, J. (2008). Do children who acquire word reading without explicit phonics employ compensatory learning? Issues of phonological recoding, lexical orthography, and fluency. *Read. Writ.* 21, 505–537. doi: 10.1007/s11145-007-9075-9

# Learning to read as the formation of a dynamic system: evidence for dynamic stability in phonological recoding

*Claire M. Fletcher-Flinn**

*College of Education, University of Otago, Dunedin, New Zealand*

Two aspects of dynamic systems approaches that are pertinent to developmental models of reading are the emergence of a system with self-organizing characteristics, and its evolution over time to a stable state that is not easily modified or perturbed. The effects of dynamic stability may be seen in the differences obtained in the processing of print by beginner readers taught by different approaches to reading (phonics and text-centered), and more long-term effects on adults, consistent with these differences. However, there is little direct evidence collected over time for the same participants. In this study, lexicalized (implicit) phonological processing, and explicit phonological and letter-sound skills are further examined in a precocious reader whose early development at 3 and 5 years has been extensively described (Cognition, 2000, 2004). At ages 10 and 14 years, comparisons were made with these earlier reports and skilled adult readers, using the same tasks for evidence of changes in reading processes. The results showed that along with an increase of reading accuracy and speed, her pattern of lexicalized phonological responses for reading did not change over time. Neither did her pattern of explicit phonological and letter-sound skills, aspects of which were inferior to her lexicalized phonological processing, and word reading. These results suggest dynamic stability of the word reading system. The early emergence of this system with minimal explicit skill development calls into question developmental reading theories that require such skills for learning to read. Currently, only the Knowledge Sources theory of reading acquisition can account for such findings. Consideration of these aspects of dynamic systems raise theoretical issues that could result in a paradigm shift with regard to best practice and intervention.

Keywords: dynamic systems, dynamic stability, theories of reading, reading acquisition, precocious reading, phonological recoding

## INTRODUCTION

Children learn to read by forming links between mental representations of visual symbols (letters) in print words, and their pre-existing phonological (sound) and semantic (comprehension) representations for spoken language. The challenge for developmental theories of reading is to propose how such a reading system of connected representations might be formed. The purpose of this study was to consider learning to read as the formation of a dynamic system, and to test the concept of "dynamic stability" by examining behavioral data from a precocious reader for changes in her processing of print over time. Although such cases in the population are somewhat rare (1-3.5%, Jackson, 1992), it does not necessarily follow that the cognitive processes of learning to read in precocious readers, or their reading system components are dissimilar to other normal-progress readers. Most theories of reading make this "similarity" assumption and apply it to another small population, those having reading acquisition difficulties (3 to 10%, Snowling, 2013). According to Jackson and Coltheart (2001, p. 156), referring to precocious readers, even a single case provides unique "…opportunities to test hypotheses about conditions that are necessary for successful reading acquisition." At the very least, such cases can contribute to the demarcation of limits on the range of application of current theories.

### CONNECTIONIST VIEWS OF READING

Current views of word reading, such as those from connectionist frameworks (e.g., Hutzler et al., 2004; Powell et al., 2006) assume neural system dynamics (analogous to computation in the brain), and address the general issue of how orthographic (print word) inputs are mapped onto spoken language (phonological) processes. These computer models (neural nets) have been applied to simulate the formation of a reading system in the brain through an initial architecture, and an extremely large corpus of words, input trials (exposures), learning rules, and error feedback. The initial architecture is changed as a result of these experiences, and the implemented reading model is compared with predictions from theories that are based on behavioral studies, and neuropsychological evidence on brain function. Although these connectionist models are purported to represent children's capacities and knowledge of reading at specific points in their learning, they are at best only approximations (Seidenberg, 2007, p. 3), and perhaps, not surprisingly, they have been criticized for lacking in developmental plausibility (Cassidy,

1990; but see Seidenberg, 2007), and ecological validity (Hutzler et al., 2004).

Notwithstanding skepticism about their utility, several theoretical aspects of these connectionist models may be relevant for gaining a deeper understanding and contribute to an explanation of how children learn to read. The basic notion of an *emergent* (or self-organizing) process based on the interaction of simple components that gradually results in relatively robust complex structure is central to connectionist theory. The initial probabilistic outputs of connectionist reading models increasingly reflect the statistical structure of the word-training corpus as the system becomes fully trained and implemented. Such learning occurs by modifying the connectivity between processing units according to a weight adjustment algorithm as a function of the word-training experience. However, due to the provision of *supervised* training (of the target responses), as well as the initial pre-training on grapheme-phoneme correspondences of some models (Perry et al., 2007), most connectionist models are only "superficially" self-organizing, and are embedded with theoretical assumptions about how children learn to read. Two models (Dufau et al., 2010; Glotin et al., 2010) more closely resemble emergent principles and use developmentally appropriate lexical databases and unsupervised training. The principle of self-organization without top-down (deterministic) direction has been claimed as ubiquitous in biological development (van Geert, 2008) and in our natural (McClelland et al., 2010) and social environments (Bronfenbrenner, 1979).

## THEORIES OF LEARNING TO READ

In contrast to these connectionist theories, the focus of standard developmental theories of learning to read (e.g., Ehri, 1999, 2005) have been on the explicit skills claimed necessary for beginning reading, in particular, phonemic awareness and knowledge of letter-sound correspondences. These skills are used when a child attempts to read an unfamiliar word through a taught heuristic known as non-lexical (explicit) phonological recoding, in which each letter-sound is pronounced in sequence and with the deletion of unnecessary vowel sounds, they are recombined into a word (e.g., "ba – aa – ga," for *bag*). The degree to which these explicit skills have been learnt forms the basis for summary (descriptive) performance accounts of the initial stages, or phases of the developmental theories. Share (1995) further emphasized the importance of non-lexical phonological recoding by claiming that it was a "self-teaching" device with successful attempts at recoding enabling the acquisition of word-specific orthographic knowledge required for skilled reading. Although concerned with learning, standard theories of reading have little to say about the dynamics of an emergent reading system, as children need to be first taught (explicitly) the basic skills, which are related but beyond the system (Jackson and Coltheart, 2001). These theories are also limited in explanatory power as they have been developed and tested with children who are taught to read with a phonics approach (Ramus, 2004; Fletcher-Flinn and Thompson, 2010).

An alternative reading acquisition theory, Knowledge Sources theory, also based on behavioral evidence, does share some principles of self-organization and implicit learning (Thompson et al., 1996; Thompson and Fletcher-Flinn, 2006, 2012). In this theory,

it is claimed that sublexical patterns of print word input and corresponding information from the child's phonological lexicon are induced implicitly as soon as the child attends to the relationship between letters and sounds within words, and a few words are stored in an orthographic lexicon (reading vocabulary with associated lexical meanings). These sublexical relations (ISRs) are induced from information across the child's emerging reading vocabulary. They are used to generate responses to new words through lexicalized (implicit) phonological recoding, and the ISRs are continually updated as new words enter the child's orthographic lexicon. Explicit skill learning is not necessary beyond a very rudimentary level. However, if such explicit skills are taught to children, they are considered another source of knowledge for the generation of responses to unfamiliar words.

Despite some shared concepts, Knowledge Sources theory differs from connectionist accounts in the specification of an orthographic lexicon. In connectionist accounts word knowledge is distributed and stored in the connections between the units rather than in an orthographic lexicon, resulting in the phenomenon of "catastrophic forgetting" (when new information interferes with old). Share, in his developmental theory, posits an orthographic lexicon from "self-learning," contingent on a later major shift from reliance on non-lexical phonological recoding to lexical processes. This contrasts with Knowledge Sources theory in which learning to read is viewed as an emergent, dynamic and *continuous* process based on learning across phonological and orthographic lexicons. The orthographic lexicon for direct access to word representations (of both phonology and lexical meanings), and lexicalized phonological recoding processes based on ISRs are available soon after reading commences.

Evidence for the induction of ISRs has been accumulating and was shown experimentally for beginner readers of English (Thompson et al., 1996; Fletcher-Flinn and Thompson, 2000, 2004; Fletcher-Flinn et al., 2004), and more recently for the acquisition of a phonemic function of hiragana, a syllabic orthography, in beginner readers of Japanese (Fletcher-Flinn et al., 2014). These results indicate the possibility of a universal process of acquisition, which seems plausible if learning to read is the formation of a dynamic system (Seidenberg, 2011).

## THE DYNAMIC STABILITY OF A READING SYSTEM

Another general property of dynamic systems, important for theories of the acquisition of reading, is *dynamic stability*, in which patterns once acquired, are not easily modified or perturbed. According to Rolls (2012), recurrent (input) patterns promote stability, and *attractor* networks in the brain (neurons that collectively settle into stable patterns of firing) enable memories to be stored and recalled. These "integrate-and-fire" neural nets, when modeled in real continuous time, have the advantage of very fast recall. The gradual incremental changes (short-term dynamics) of the connection weights of an attractor network determine the final steady state of the connectionist system (long-term dynamics) (Munakata and McClelland, 2003; van Geert, 2008).

Dynamic stability is not an aspect that is within the scope of standard developmental theories of learning to read, insofar as their focus is on broad phases (e.g., Ehri, 1999, 2005) or changes in processing (Share, 1995). However, it has been considered by Knowledge Sources theory with regard to the developmental continuity of lexicalized phonological recoding for Maxine, a precocious reader who has been extensively studied from prior to the age of 2 years (Fletcher-Flinn and Thompson, 2000) and continued until the age of 7 years (Fletcher-Flinn and Thompson, 2004). It was shown that her processing of lexicalised phonological recoding was developmentally stable from 3 to 5 years of age, encompassing word-reading levels from 8 to 14 years (Fletcher-Flinn and Thompson, 2004).

With regard to normal-progress children learning to read, the long-term effects of dynamic stability may be seen from differences obtained when comparing the processing of print by 6-year-old beginner readers making normal-progress (Connelly et al., 2001) with phonics and text-centered approaches, as well as those 6- to 7-year-olds making slower progress (Thompson et al., 2008). For the same word reading ability, these studies showed faster text reading speed by beginners with non-phonics approaches compared with those children taught phonics. The latter were better able to read pseudowords, but were disadvantaged when reading low frequency words, and words that were irregular (Connelly et al., 2001). The speed of text reading advantage for non-phonics approaches was attributed to the greater time made available for text reading. Moreover, with the speed advantage, for an equal amount of time, beginners are exposed to more print words (and associated meanings).

Processing differences were also found among skilled adult readers of equivalent reading ability after nearly two decades beyond their initial reading instruction (Thompson et al., 2009). The adults who had initial instruction in phonics performed better on metalinguistic and letter-sound tasks, but similar to the children in the previous studies, they made more errors (regularizations) on contextually dependent pseudowords, and some low-frequency words than those without such instruction. It was suggested that the initial years of phonics reading instruction left a cognitive bias in processing associated with non-lexical phonological recoding that did not attenuate, or become superseded over time.

While the cross-sectional studies are intriguing because the results suggest a degree of long-term dynamic stability in the processing of print from different teaching approaches, they do not directly address questions of developmental change over time, or the stability of procedures for reading unfamiliar words. The purpose of this study was to examine these issues with regard to the learning and stability of lexicalized phonological processing in reading acquisition during development, alongside the learning of explicit phonological and letter-sound skills. This study examined Maxine's reading development at age 10 and 14 years making comparisons with earlier published reports with the same tasks for evidence of changes in reading processes. Comparisons were made, where appropriate, with published results of skilled adult readers (Thompson et al., 2009) who, like Maxine, had not received explicit phonics instruction as beginner readers. This reading-level match was used to examine the extent to which the operation of components of her reading system differed from other highly skilled readers.

## SCHOOL EXPERIENCE AND INTERESTS, WORD READING ACHIEVEMENT, AND EXPLICIT PHONOLOGICAL SKILLS

Maxine entered a private intermediate school at the age of 8 years, and graduated from high school at 14 years. She studied the normal New Zealand curriculum, and continued with her musical, and other interests. She particularly enjoyed playing Pokemon on her Gameboy, skiing, and chess. Her school report at the end of intermediate school indicates that she was an exemplary student, with impressive examination grades well above the median in all major subjects. Other comments included her consideration of others, valued contributions to group discussions, and that she was well liked by students and staff. At her high school graduation, she won the class award for English Literature.

Maxine was assessed at two time periods, from 10 years 10 months (10:10) to 10:11, and 13:11 to 14:0 on a range of reading and phonological awareness tests. Informed written consent was obtained from Maxine and her parents, and the study was approved by the Auckland Human Subjects Ethics Committee, as part of a larger study on precocious readers. Standardized word reading tests included the Wide Range Achievement Test 3 – Combined Form (WRAT-3, Wilkinson, 1993) for both oral reading and spelling, and the Nelson-Denny Reading Test [N-D, Brown et al. (1981)], Vocabulary subtest, to assess reading comprehension of single words. These assessments showed that Maxine continued her precocious word reading development, reaching beyond high school levels by 10:11 on the WRAT-3 oral reading subtest, and by 13:11 she was equivalent to the comparison sample of adults on the N-D Vocabulary subtest (Thompson et al., 2009). Spelling was consistent with her reading ability.

The phonological awareness tasks included the Rosner Test of Auditory Analysis Skills (Rosner and Simon, 1971), and the Yopp–Singer phoneme segmentation task (Yopp, 1988). The Rosner Test, consisting of 40 items, assesses the skill of the deletion of phonemes from various positions in words, e.g., saying *man* without the "m" sound. The Yopp-Singer phoneme segmentation task requires the pronunciation of sounds of spoken words in the correct order, e.g., "Tell me all the sounds that you can hear in the word 'dog'" (Three sounds relating to the three phonemes comprise the correct response.) Both tests had been administered when she was 5 and 7 years (but in a 13 item version of the Rosner Test at 5 years, Fletcher-Flinn and Thompson, 2004). The Rosner Test placed Maxine at the Grade 6 (U.S.) ceiling level of the test from 7 years (**Table 1**). On the Yopp-Singer segmentation task, for the same ages, she was within (or close to) +1 standard deviation (SD) of the norms based on children in kindergarten (average age of 5:10) in U.S. schools with some letter-sound instruction. Their average score was 54% (SD 35%). Although she attained 92% correct at 13:11, her performance on this task continued to be underdeveloped relative to her age and reading ability.

**Table 1 | Test age levels for phonological awareness for Maxine at chronological ages 7:3 to 13:11, and mean percentage correct on phoneme awareness and the extended Scarborough task for Maxine and New Zealand university students without phonics instruction (standard deviation in parenthesis).**

| | Chronological age (years and months) | | | |
| | Maxine | | | NZ university students |
| | 7:3–7:11 | 10:11 | 13:11 | |
| --- | --- | --- | --- | --- |
| **Test ages** | | | | |
| Yopp-singer phoneme segmentation | 5 | 5 | 5 | – |
| Rosner deletion | 11 | 11 | 11 | – |
| **Percentage correct** | | | | |
| Phoneme awareness (aural) | –[a] | 53 | 70 | 61 (11) |
| **Extended Scarborough task** | | | | |
| Graphophonemic awareness | – | 63 | 77 | 61 (20) |
| Graphophonemic segments | – | 43 | 53 | 39 (13) |

[a]*Tests not administered.*

As Maxine was reading at the same level on the N-D reading test as the comparison sample of adults, the same phonemic awareness and graphophonemic (extended Scarborough task, Thompson et al., 2009) measures used for them were administered to her. In the phoneme awareness task 30 words were presented in aural form and the task was to count the number of the "smallest sounds" in each word, e.g., four for *socks*. The same words from this task were used in a graphophonemic task in which the words were presented in print form. In this task, the participant must read the word, note the number of sounds in the word (awareness score), and underline the letter or letter sequence belonging to each sound (identity score). At 10:11, Maxine's accuracy on tests of phoneme awareness, graphophonemic awareness, and graphophonemic segments were all within 1 SD of that reported for the adults. She showed continued development of these skills at 13:11, although remaining within about 1 SD of the adult means for these tasks (**Table 1**).

It seems fair to say that as Maxine's word reading advanced, her phoneme awareness skills continued to develop and were not markedly different to skilled adults at the same reading level. However, her explicit skill at segmenting spoken words into phonemes and to recite them in order remained underdeveloped and consistent with her earlier kindergarten level performance on this task. This is interesting because Yopp (1988) found phoneme deletion tasks, like the Rosner Test on which Maxine performed adequately, to be more difficult than phoneme segmentation tasks, such as the Yopp-Singer, for his sample of 5-year-olds. Although modeled occasionally by her parents, the "sounding-out" response heuristic for this task was never used by Maxine (Fletcher-Flinn and Thompson, 2000, p. 184), and is not a skill that would have been taught to the New Zealand adults (Thompson et al., 2009). It is, however, an explicit skill that is required for non-lexical phonological recoding to read unfamiliar words.

## EXPERIMENT 1: EXPLICIT LETTER-SOUND SKILLS

The set of three tasks – letter naming, letter sounds, and digraph sounds – which were employed for Maxine (Fletcher-Flinn and Thompson, 2004), and for the comparison group of adults (Thompson et al., 2009) were administered when Maxine was 10:10 and 13:11 using the same procedure with computer presentation in lowercase. Speeded response instructions were given and there was no correction of responses. In scoring, correct letter-sound responses were those taught in explicit phonics instruction. For the 29 digraphs, (e.g., *ee*, *aw*, *ch*), Maxine was instructed to pronounce the sound associated with the two letters, and scoring was the same as in the earlier studies.

Maxine's mean accuracy (**Table 2**) for giving phonic sounds to letters and digraphs from 5:8 to 13:11 was within (or close to) 1 SD of the adult comparison sample (Thompson et al., 2009). There was no significant increase in her mean percentage accuracy over time using the McNemar test for the significance of changes, from 5:8 to 10:10 for giving phonic sounds to letters, $X^2(1) = 1.54$, and digraphs, $X^2(1) = 1.23$, $p > 0.20$; or, from 10:10 to 13:11 for giving phonic sounds to letters, $X^2(1) = 0.0$, and digraphs $X^2(1) = 0.16$, $p > 0.68$. Similarly, over the longer span of time from 5:8 to 13:11 there was no significant change in percentage accuracy for giving phonic sounds to letters, $X^2(1) = 1.54$, $p = 0.21$, or digraphs $X^2(1) = 2.5$, $p = 0.11$.

A repeated-measures ANOVA over items for Maxine's response times (RTs) on letter names to which she responded accurately showed a significant change over time $F(2,38) = 10.45$, $p < 0.0001$, $\eta^2 = 0.36$. Using paired $t$-tests, her performance at 13:11 was slower than at 5:8, $t(19) = -5.69$, $p < 0.0001$; and 10:10, $t(24) = -4.04$, $p < 0.0001$. There was no change from 5:8 to 10:10, $t(20) = -1.19$, $p = 0.25$. The results were similar for responses to letter sounds, $F(2,22) = 12.01$, $p < 0.0001$, $\eta^2 = 0.52$, with significantly slower RTs from 5:8 to 13:11, $t(12) = -4.65$, $p < 0.001$; and 10:10, $t(12) = -3.87$, $p < 0.002$.

Table 2 | Experiment 1: Mean percentage accuracy and response times (RTs) for lowercase letter names, letter sounds, and digraph sounds for Maxine at 5:8, 10:10, and 13:11, and for the NZ university students without phonics instruction (standard deviation in parenthesis).

|  | Letter names | Letter sounds | Digraph sounds |
|---|---|---|---|
| **Mean percent correct** | | | |
| Maxine at 5:8 | 92 | 65 | 62 |
| Maxine at 10:10 | 100 | 81 | 76 |
| Maxine at 13:11 | 100 | 81 | 82 |
| NZ university students | 100 (0) | 75 (12) | 73 (10) |
| **Mean RTs (ms) for correct responses** | | | |
| Maxine at 5:8 | 522 | 551 | 558 |
| Maxine at 10:10 | 551 | 765 | 683 |
| Maxine at 13:11 | 638 | 651 | 589 |

She was somewhat faster from 10:10 to 13:11, $t(12) = 2.50$, $p < 0.02$. Responses to digraphs also changed significantly over time, $F(2,16) = 9.27$, $p < 0.002$, $\eta^2 = 0.54$, with slower RTs from 5:8 to 10:10, $t(9) = -4.08$, $p < 0.003$, but faster from 10:10 to 13:11, $t(20) = 2.64$, $p < 0.02$. There was no difference in RTs for digraphs at 5:8 and 13:11, $t(8) = -1.99$, $p < 0.08$, although it was in the negative direction, indicating a decrease in speed with age.

In summary, Maxine's mean percentage accuracy for providing phonic sounds to letters and digraphs was similar to the comparison sample of adults, and did not change significantly over time. Similar to Maxine, the adults had not experienced explicit phonics instruction as beginner readers, and their mean accuracy did not exceed 75% (SD = 12). Of interest, Maxine showed a tendency for slower RTs with age for letter names, sounds, and digraphs compared with her performance at 5:8.

## EXPERIMENT 2: NONWORD PRONUNCIATION
Two nonword pronunciation tasks were administered to Maxine. Each task consisted of sets of Regular, body-consistent; Regular, body-inconsistent; and Irregular, body-consistent nonwords presented in a randomized sequence. The first source of these nonwords was from Andrews and Scarratt (1998, Experiment 2), and the second was from Coltheart and Leahy (1992, Task 2). The first category required a regular response for accuracy (e.g., *stell*, *dilt*). The other two categories were heterophonic nonwords. For the category of nonwords with Inconsistent lexical bodies (e.g., *dush* which can be pronounced with –*ush* as in "rush," or "push") either the regular, or irregular pronunciation, respectively, was acceptable. The first two categories of nonwords consisted of 40 and 20 items, respectively, from each source. The third classification consisted of nonwords that always have Irregular lexical bodies (e.g., *thild*) that occur in several real words (e.g., *child*). There were 20 of these nonwords in the Coltheart and Leahy task, and 24 of them in the other source. Andrews and Scarratt (1998) also included 24 Irregular Unique nonwords (e.g., *hourt*, *yign*) with Irregular

lexical bodies having only one real word exemplar. The procedure was the same as in the previous reported experiments (Fletcher-Flinn and Thompson, 2000, 2004; Thompson et al., 2009[1]).

### ANDREWS AND SCARRATT NONWORDS (1998)
**Table 3** shows the percentages of regular and irregular responses for the categories of nonwords on the Andrews and Scarratt (1998) task. Overall mean combined accuracy scores were calculated, which included: (1) the *regular* responses to the Regular Consistent and the Inconsistent nonwords, and (2) the *irregular* responses to the Irregular consistent and Irregular Unique nonwords. (The regular pronunciations to these two categories were excluded as less accurate "regularizations.")

The combined mean percentage accuracy for *regular* responses to the Regular Consistent and the Inconsistent nonwords (**Table 4**) showed that Maxine varied very little (between 92 and 91%) from 5:9 to 14:0. This was within +1 SD of the mean percentage accuracy[1] for the adults without phonics instruction. Accurate combined mean percentage accuracy for *irregular* responses to the Irregular consistent and Irregular Unique nonwords was 82% for Maxine at both 10:10 and 14:0, and at 5:9, it was 73%. She exceeded the mean percentage accuracy of the adults by +2.2 SD. The McNemar test showed no significant change in irregular responses for Maxine from 5:9 to 10:10 and 14:0, $X^2(1) = 1.22$, $p = 0.27$.

Response times were analyzed for those response categories above mean acceptable response rates of 33% or higher. At 5:8, 10:10, and 14:0, Maxine's mean RTs for accurate *regular* responses for the Regular Consistent nonwords were: 549, 497, and 380 ms; and, for the Inconsistent nonwords 555, 491, and 376 ms, respectively.

For accurate regular responses to Regular consistent nonwords, a repeated-measures ANOVA by items showed a significant change in RTs over time, $F(2,68) = 53.97$, $p < 0.0001$, $\eta^2 = 0.61$. Paired *t*-tests indicated that Maxine was significantly faster with each age comparison: 5:9–10:10, $t(37) = 2.29$, $p = 0.03$; 10:10–14:0, $t(35) = 8.17$, $p < 0.0001$; and, 5:9 to 14:0, $t(35) = 11.21$, $p = 0.0001$. A similar pattern was shown by Maxine on correct regular responses to Inconsistent regular nonwords, $F(2,52) = 46.57$, $p < 0.0001$, $\eta^2 = 0.64$ with faster RTs with age: 5:9–10:10, $t(27) = 2.02$, $p = 0.05$; 10:10–14:0, $t(31) = 9.30$, $p < 0.0001$; and, 5:9–14:0, $t(29) = 11.45$, $p = 0.0001$.

For correct *irregular* responses, Maxine's mean RTs for the same ages, for Irregular Consistent nonwords were: 553, 534, and 383 ms; and, for Irregular Unique nonwords: 582, 562, and 374 ms. There was a significant change in RTs over time for irregular responses to Irregular Consistent nonwords, $F(2,34) = 32.24$, $p < 0.0001$, $\eta^2 = 0.66$. She was significantly faster at 14:0 than at 10:10, $t(19) = 9.19$, $p < 0.0001$; and, at 14:0 than at 5:9, $t(18) = 7.43$, $p = 0.0001$. RTs were equivalent at 5:9 and 10:10, $t(17) = 1.29$, $p = 0.22$. The irregular responses to Irregular Unique nonwords showed a similar pattern of change, $F(2,20) = 36.75$,

---

[1]Only items that were acceptable in both standard New Zealand and Scottish accents were scored in Thompson et al. (2009, p. 227), whereas all items were scored for comparison with Maxine in this study.

**Table 3 | Experiment 2: Mean percentage of regular and irregular pronunciations for Andrews and Scarratt nonwords varying in regularity and consistency of body spelling for Maxine at 5:9, 10:10, and 14:0, and for the New Zealand university students without phonics instruction (standard deviation in parenthesis).**

|                              | Regular consistent | Inconsistent | Irregular consistent | Irregular unique |
|------------------------------|--------------------|--------------|----------------------|------------------|
| **Regular pronunciations**   |                    |              |                      |                  |
| 5:9 years                    | 95                 | 88           | 4                    | 25               |
| 10:10 years                  | 98                 | 83           | 0                    | 17               |
| 14:0 years                   | 93                 | 90           | 4                    | 21               |
| New Zealand university students | 92 (4)          | 84 (7)       | 10 (9)               | 37 (10)          |
| **Irregular pronunciations** |                    |              |                      |                  |
| 5:9 years                    | _[a]               | 10           | 83                   | 63               |
| 10:10 years                  | –                  | 15           | 92                   | 71               |
| 14:0 years                   | –                  | 10           | 88                   | 75               |
| New Zealand university students | –               | 10 (6)       | 63 (17)              | 46 (8)           |

[a] Dashes indicate that regular pronunciations do not exist for regular consistent nonwords.

$p < 0.0001$, $\eta^2 = 0.79$. There was no difference in RTs at 5:9 and 10:10, $t(10) = 1.208$, $p = 0.30$, but she was significantly faster at 14:0 than at 10:10, $t(14) = 6.85$, $p < 0.0001$; and, at 14:0 than at 5:9, $t(12) = 7.14$, $p = 0.0001$.

### COLTHEART AND LEAHY NONWORDS (1992)

Maxine's mean percentage accuracy for *regular* responses to regular consistent and inconsistent nonwords, averaged over the two categories of nonwords, at both 10:10 and 14:0 was 95% (**Table 4**), which was within 1 SD of the comparison sample of adults at 83%. Accurate irregular responses to the Irregular Consistent nonwords were 80% at 3:4, and 70% at 5:5, with these changes reported not significant (Fletcher-Flinn and Thompson, 2004). Maxine's accurate irregular responses for the same set of nonwords were 70%

at 10:10, and 75% at 14:0. She was, respectively, +1.5 SD and +1.81 SD more accurate than the comparison sample of adults. The McNemar test for the significance of change for each age comparison was not significant, $X^2(1) < 1$.

Maxine's response times were analyzed for those categories above mean acceptable response rates of 33% or higher. At 5:8, 10:10, and 14:0, respectively, Maxine's mean RTs for regular responses to regular consistent nonwords were: 560, 477, 386 ms, and for ambiguous inconsistent nonwords: 609, 525, 412 ms. Her mean RTs for irregular responses to Irregular Consistent nonwords were: 584, 501, and 409 ms. An ANOVA with two factors (time, category type) over items was significant, for time, $F(2,139) = 78.76$, $p < 0.0001$, $\eta^2 = 0.53$, and category type, $F(2,139) = 4.48$, $p < 0.013$, $\eta^2 = 0.06$. There was no

**Table 4 | Experiment 2: Mean percentage of regular and irregular pronunciations for Coltheart and Leahy nonwords varying in regularity and consistency of body spelling for Maxine from 3:4 to 14:0, and New Zealand university students without phonics instruction (standard deviation in parenthesis).**

|                              | Regular consistent | Inconsistent | Irregular consistent |
|------------------------------|--------------------|--------------|----------------------|
| **Regular pronunciations**   |                    |              |                      |
| Maxine at 3:4                | 100                | 55           | 10                   |
| Maxine at 5:5                | 100                | 65           | 30                   |
| Maxine at 11:10              | 100                | 90           | 10                   |
| Maxine at 14:0               | 100                | 90           | 5                    |
| New Zealand University students | 95 (6)          | 71 (13)      | 28 (12)              |
| **Irregular pronunciations** |                    |              |                      |
| Maxine at 3:4                | _[a]               | 35           | 80                   |
| Maxine at 5:5                | –                  | 30           | 70                   |
| Maxine at 11:10              | –                  | 10           | 70                   |
| Maxine at 14:0               | –                  | 10           | 75                   |
| New Zealand university students | –               | 21 (13)      | 46 (16)              |

[a] Dashes indicate that irregular regular pronunciations do not exist for regular consistent nonwords.

interaction, $F < 1$. Paired $t$-tests with Bonferroni adjustments showed that Maxine was faster at 14:0 compared with her earlier responses at 5:9 and 10:10, MSE $= 4951.90$, $p < 0.0001$; and she gave regular responses to Regular consistent nonwords faster than to Ambiguous inconsistent nonwords, MSE $= 4770.36$, $p < 0.03$. There was no difference in RTs for irregular responses to Irregular consistent nonwords compared with the regular responses for the other two categories, MSE $= 4770.36$, $p \leq 1$.

## DISCUSSION

Although Maxine's ceiling level of performance on the Andrews and Scarratt nonwords from 5:9 did not leave much room for gains in accuracy, she was equivalent to the adult comparison sample for regular responses to the regular consistent and inconsistent nonwords, and she exceeded them on the irregular responses to the irregular consistent and irregular unique nonwords. The speed of Maxine's responses increased significantly over time, and by 14 years her mean RT was 378 ms, over the four categories of items, which exceeds the mean RT of 639 ms of the university students from the Andrews and Scarratt's (1998, Table 8) study. Maxine's accuracy on the Coltheart and Leahy (1992) nonwords showed the same pattern of performance for regular and irregular responses, and the same decrease in RTs with age. Similarly, by 14 years, with a mean RT (over categories) of 402 ms, she exceeded the mean RT of 704 ms for the university students from Coltheart and Leahy (1992, Table 4).

The results for Maxine on the two nonword tasks converge to indicate a high degree of proficiency of phonological recoding, relative to samples of skilled adult readers. However, for the present purposes, the most important aspect of her performance was that she retained the same pattern of category responses over time. Lexical phonological processing was developmentally stable from 3 to 14 years of age, covering word-reading levels from 8 years to skilled adult levels of performance. Concomitantly, her speed of processing nonwords continued to increase over this period of time.

## EXPERIMENT 3: WORD NAMING AND LEXICAL SEMANTIC INFLUENCES

Strain et al. (1995, Experiment 2) showed that for adults the degree of imageability, which is a semantic characteristic of words, contributes to word reading accuracy when the words are of low frequency and irregular in spelling. The same result was found for Maxine at 5:9 years and an 11-year-old matched word-reading level comparison sample of normal progress readers without phonics instruction (Fletcher-Flinn and Thompson, 2004), and for the comparison sample of adults (Thompson et al., 2009).

At 10:10 and 14:0, the words from Strain et al. (1995) were administered to Maxine with the same presentation and scoring procedure as for the previous experiments with the 11-year-olds and adults. The stimuli comprised four categories of low-frequency words, with 16 words in each: irregularly spelt words of high imageability (e.g., *boulder, climb*), irregularly spelt words with low imageability (e.g., *broader, cache*), regularly spelt words of high imageabilty (e.g., *banner, cliff*), and regularly spelt words with low imageability (e.g., *blessing, cleft*).

Maxine showed the same ceiling level of accuracy as the adults for regular words and for words of high imageability by 10:10 (**Table 5**), and for irregular words of low imageability by the age of 14 years. Similar to Maxine's performance at 5 years, using McNemar's test, regular words were read more accurately than irregular words at 10:11, $X^2(1) = 4.94$, $p < 0.03$, but not at 14:0, $X^2(1) < 1$. There was no effect of imageability at either age 10:11 or 14:0, $X^2(1) = 2.52$, $p = 0.11$, and $X^2(1) < 1$, respectively. Similar to her previous performance at 5:9 and the adults, regularization responses accounted for 83 and 100% of her errors on the low imageability words with irregular spellings at 10:10 and 14:0, respectively.

At 14:0, Maxine's mean RTs were 1.8–2.15 SD shorter than the adults across all four categories of words. In a 3-way ANOVA by items, there was a significant effect of age, $F(2,157) = 17.20$, $p < 0.0001$, $\eta^2 = 0.18$, but no effect for regularity or imageability, or any interaction ($p > 0.10$). Paired $t$-tests with Bonferroni adjustments on the main effect of time showed that Maxine was faster at 14:0 compared with her earlier responses at 5:9 and 10:10 (MSE $= 4770.36$, $p < 0.0001$).

These results are consistent with the standardized word reading assessments. She reached adult (university) levels of accuracy by 14 years, with significantly faster RTs to isolated words than the adults. Although there was similarity on this task to some aspects of her earlier performance, the absence of any differences in accuracy or RTs for the effects of regularity or imageability may be attributed to reaching ceiling levels of performance. In that case, the phonological processing of words is too efficient to be assisted by a word's semantic characteristics.

## GENERAL DISCUSSION

Although Maxine continued to develop greater speed and word reading accuracy over time, her pattern of responses for categories of nonwords did not change from when she was much younger. According to Fletcher-Flinn and Thompson (2000), Maxine was able to read nonwords while having underdeveloped explicit skills for word reading when she was 3 years by inducing sublexical relations (ISRs) between orthographic and phonological components in words that are stored in orthographic memory. Through a process of lexicalized phonological recoding, she was able to use these ISRs, in turn, to read unfamiliar words. The current findings suggest that the processes associated with lexicalized phonological recoding apparently become stable very early in acquisition, and are resistant to change. This long-term stability is not within the explanatory range of the standard developmental theories of learning to read that propose phases (e.g., Ehri, 1999, 2005), or shifts in phonological recoding processes (Share, 1995). It is, therefore, a limitation of these theories.

As Maxine continued to develop her precocious reading skills, both word reading and word comprehension, going well beyond the normal attainment for her age, her phonological awareness showed differential success. Her performance was age appropriate on a test of phoneme deletion, and at 10:11 and 13:11 equivalent to the comparison sample of adults without phonics instruction on an aural phoneme awareness task and two

**Table 5 | Experiment 3: Mean percent correct and RTs for words varying in regularity of spelling and in imageability for Maxine at 5:9, 10:10 and 14:0, and New Zealand university students without phonics instruction (standard deviation in parenthesis).**

| | Regular | | Exception | |
|---|---|---|---|---|
| | High imageability | Low imageability | High imageability | Low imageability |
| **Mean percentage correct** | | | | |
| Maxine at 5:9 | 94 | 94 | 88 | 50 |
| Maxine at 10:10 | 100 | 100 | 94 | 63 |
| Maxine at 14:0 | 100 | 100 | 100 | 88 |
| NZ university students | 100 (0.0)[a] | 99 (2.9) | 98 (4.2) | 83 (10.2) |
| **Mean response time** | | | | |
| Maxine at 5:9 | 481 | 506 | 492 | 493 |
| Maxine at 10:10 | 493 | 520 | 454 | 474 |
| Maxine at 14:0 | 418 | 409 | 407 | 444 |
| NZ university students (without phonics instruction) | 561 (79) | 593 (87) | 581 (81) | 642 (110) |

graphophonemic tasks. However, on another test (Yopp-Singer) involving segmentation and pronunciation of phoneme components of words, she only reached the 5-year-level of normal age controls, which is consistent with her underdeveloped performance on this task when much younger (Fletcher-Flinn and Thompson, 2000, 2004). It seems reasonable to conclude that she had not learnt the explicit procedure for non-lexical phonological recoding, and hence, it remained underdeveloped.

Maxine's proficiency in providing the phonic sounds for isolated letters, and sounds for digraphs was comparable to a sample of adults without phonics instruction whose learning was also incomplete, averaging 74% across the two categories (Thompson et al., 2009). In contrast to an increase of speed for word (and nonword) reading, her response to individual letters tended to become slower over time. Her mean response times for letter names, phonic sounds for letters and digraph sounds at 13:11 were 3.67 SD, 2.13 SD, and 1.99 SD longer, respectively, than her mean accuracy (combined over items) for the low frequency words of Strain et al. (1995). For comparison, the mean RTs for Maxine's responses to these words, and the nonwords (combined over items) from Andrews and Scarratt (1998) were within 0.52 SD. The significant differences between Maxine's response times for isolated letters, and word and nonword reading indicate that the reading system formed consists of word representations, and does not include explicit responses to isolated letters, which are considered extrasystem entities (Jackson and Coltheart, 2001, p. 103). Of more importance is the lack of any difference in response times between real words and nonwords, indicating that the source of knowledge for reading the nonwords must be lexical in the form of ISRs.

In summary, although showing more processing efficiency, Maxine's pattern of performance was not different to the comparison group of adult readers, nor had it been different to earlier reading-age matches (Fletcher-Flinn and Thompson, 2000, 2004). The evidence presented on the long-term stability of

lexicalized phonological recoding as shown by Maxine's stable performance over time is indicative of the formation of an attractor state of a dynamic system displaying very fast recall. The reading system formed need not be underpinned by explicit skill knowledge as claimed by standard developmental theories (e.g., Share, 1995; Ehri, 1999, 2005), as lexicalized procedures appear to be sufficient for both the establishment, with minimal skills (Fletcher-Flinn and Thompson, 2000), and the expansion and stability of the emergent system, as shown by these results. These findings support Knowledge Sources theory and converge with earlier cross-sectional studies on the induction of ISRs in normal-progress readers without phonics instruction (Thompson et al., 1996; Fletcher-Flinn and Thompson, 2000, 2004), those with such instruction (Fletcher-Flinn et al., 2004), and the long-term effects of differing approaches to reading in normal-progress readers (Connelly et al., 2001; Thompson et al., 2008), and skilled adults (Thompson et al., 2009).

The accumulating (see Thompson, 2014, for a review) and converging evidence from these studies indicates that non-lexical phonological recoding is not central to acquiring word-specific orthographic knowledge as Share (1995) claims, although if taught, successful recoding may contribute to the addition of new words in the orthographic lexicon (Thompson et al., 1996; Thompson and Fletcher-Flinn, 2006, 2012). It is interesting to speculate that in this case, if the procedural (instructional) heuristic is abstracted along with new word pronunciations, it might explain the irregular word disadvantage of beginner readers with phonics instruction (Connelly et al., 2001). Alternatively, with the emergent reading system achieving early stability, regularization of new words might be due to an excessive exposure to regular words (and pseudowords) from typical phonics programes, thus creating an initial bias in the formation of ISRs. A combination of both would strengthen any tendency toward the regularization of new words, leaving a long-term cognitive bias in processing (Thompson et al., 2009). Consideration of these questions and issues arising from them could

provide the impetus for future research, and may pave the way for changes to best optimize reading instruction and intervention.

If scientific progress is to be made, then our current theories of how children learn to read need better scrutiny, with both intensive longitudinal study of single cases and population samples, to test and delineate their range of application. Ideas from connectionist theory, and more generally from cognitive science, on the formation of dynamic systems can contribute to these new tests.

## ACKNOWLEDGMENTS

## REFERENCES

Andrews, S., and Scarratt, D. R. (1998). Rule and analogy mechanisms in reading nonwords: Hough dou peapel rede gnew wirds? *J. Exp. Psychol. Hum. Percept. Perform.* 24, 1052–1086.

Bronfenbrenner, U. (1979). *The Ecology of Human Development.* Cambridge, MA: Harvard University Press.

Brown, J. I., Bennett, J. M., and Hanna, G. (1981). *Nelson–Denny Reading Test, Form E.* Chicago: Riverside.

Cassidy, S. (1990). When is a developmental model not a developmental model? *Cogn. Syst.* 2, 329–344.

Coltheart, V., and Leahy, J. (1992). Children's and adults' reading of nonwords: Effects of regularity, and consistency. *J. Exp. Psychol. Learn. Mem. Cogn.* 18, 718–729.

Connelly, V., Johnston, R. S., and Thompson, G. B. (2001). The effects of phonics instruction on the reading comprehension of beginning readers. *Read. Writ.* 14, 423–457.

Dufau, S., Lété, B., Touzet, C., Glotin, H., Ziegler, J. C., and Grainger, J. (2010). A developmental perspective on visual word recognition: new evidence and a self-organising model. *Eur. J. Cogn. Psychol.* 22, 669–694.

Ehri, L. C. (1999). "Phases of development in learning to read words," in *Reading Development and The Teaching of Reading*, eds J. Oakhill and R. Beard (Oxford: Blackwell), 79–108.

Ehri, L. C. (2005). "Development of sight word reading: Phases and findings," in *The Science of Reading: A Handbook*, eds M. J. Snowling and C. Hulme (Oxford: Blackwell), 135–154.

Fletcher-Flinn, C. M., Shankweiler, D., and Frost, S. J. (2004). Coordination of reading and spelling in early literacy development: an examination of the discrepancy hypothesis. *Read. Writ.* 17, 617–644.

Fletcher-Flinn, C. M., and Thompson, G. B. (2000). Learning to read with underdeveloped phonemic awareness but lexicalized recoding: a case study of a three-year-old. *Cognition* 74, 177–208.

Fletcher-Flinn, C. M., and Thompson, G. B. (2004). A mechanism of implicit lexicalized phonological recoding used concurrently with underdeveloped explicit letter sound skills in both precocious and normal reading development. *Cognition* 90, 303–335.

Fletcher-Flinn, C. M., and Thompson, G. B. (2010). "Learning to read: teaching approaches and theories of word recognition," in *Lifespan Development: New Zealand Perspectives*, 2nd Edn, eds J. Low and P. Jose (Albany, NZ: Pearson), 72–81.

Fletcher-Flinn, C. M., Thompson, G. B., Yamada, M., and Meissel, K. (2014). Learning of a formation principle for the secondary phonemic function of a syllabic orthography. *Read. Writ.* 27, 875–904.

Glotin, H., Warnier, P., Dandurand, F., Dufau, S., Lété, B., Touzet, C., et al. (2010). An adaptive resonance theory account of the implicit learning of orthographic word forms. *J. Physiol.* 104, 19–26.

Hutzler, F., Ziegler, J. C., Perry, C., Wimmer, H., and Zorzi, M. (2004). Do current connectionist learning models account for reading development in different languages? *Cognition* 91, 273–296.

Jackson, N. E. (1992). "Precocious reading of english: sources, structure, and predictive significance," in *To Be Young and Gifted*, eds P. S. Klein and A. J. Tannenbaum (Norwood, NJ: Ablex), 171–203.

Jackson, N. E., and Coltheart, M. (2001). *Routes to Reading Success and Failure.* Ann Arbor, MI: Taylor and Francis.

McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., et al. (2010). Letting structure emerge: Connectionist and dynamical systems approaches to understanding cognition. *Trends Cogn. Sci.* 14, 348–356.

Munakata, Y., and McClelland, J. L. (2003). Connectionist models of development. *Dev. Sci.* 6, 413–429.

Perry, C., Ziegler, J. C., and Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: the CDP+ model of reading aloud. *Psychol. Rev.* 114, 273–315.

Powell, D., Plaut, D., and Funnell, E. (2006). Does the PMSP connectionist model of single word reading learn to read in the same way as a child? *J. Res. Read.* 29, 229–250.

Ramus, F. (2004). "The neural basis of reading acquisition," in *The Cognitive Neurosciences*, 3rd Edn, ed. M. S. Gazzaniga (Cambridge, MA: MIT Press), 815–824.

Rolls, E. T. (2012). Advantages of dilution in the connectivity of attractor networks in the brain. *Biol. Inspir. Cogn. Architect.* 1, 44–54.

Rosner, J., and Simon, D. P. (1971). The Auditory Analysis Test: an initial report. *J. Learn. Disabil.* 4, 384–392.

Seidenberg, M. S. (2007). "Connectionist models of reading," in *Oxford Handbook of Psycholinguistics*, ed. G. Gaskell (Oxford: Oxford University Press), 235–250.

Seidenberg, M. S. (2011). "Reading in different writing systems: one architecture, multiple solutions," in *Dyslexia Across Languages: Orthography and the Gene-Behavior link*, eds P. McCardle, J. Ren, and O. Tzeng (Baltimore, MD: Paul Brooke Publishing), 151–174.

Share, D. L. (1995). Phonological recoding and self-teaching: the sine qua non of reading acquisition. *Cognition* 55, 151–218.

Snowling, M. J. (2013). Early identification and interventions for dyslexia: A contemporary view. *J. Res. Spec. Educ. Needs* 13, 7–14.

Strain, E., Patterson, K. E., and Seidenberg, M. S. (1995). Semantic effects in single-word naming. *J. Exp. Psychol. Learn. Mem. Cogn.* 21, 1140–1154.

Thompson, G. B. (2014). Discovering and accounting for limitations in applications of theories of word reading acquisition. *Front. Psychol.* 5:579. doi: 10.3389/fpsyg.2014.00579

Thompson, G. B., Connelly, V., Fletcher-Flinn, C. M., and Hodson, S. J. (2009). The nature of skilled adult reading varies with type of instruction in childhood. *Mem. Cogn.* 37, 223–234.

Thompson, G. B., Cottrell, D. S., and Fletcher-Flinn, C. M. (1996). Sublexical orthographic-phonological relations early in the acquisition of reading: the knowledge sources account. *J. Exp. Child Psychol.* 62, 190–222.

Thompson, G. B., and Fletcher-Flinn, C. M. (2006). "Lexicalized implicit learning in reading acquisition: the knowledge sources theory," in *Cognition and Language: Perspectives from New Zealand*, eds C. M. Fletcher-Flinn and G. M. Haberman (Bowen Hills, QLD: Australian Academic Press), 141–156.

Thompson, G. B., and Fletcher-Flinn, C. M. (2012). "Toward better teaching: Revising the fundamentals of learning to read," in *Contemporary Debates on Child Development and Education*, eds S. Suggate and E. Reese (London: Routledge, Taylor and Francis), 250–260.

Thompson, G. B., McKay, M. F., Fletcher-Flinn, C. M., Connelly, V., Kaa, R. T., and Ewing, J. (2008). How is word reading achieved without explicit phonics instruction? Issues of phonological recoding, lexical orthography, and fluency. *Read. Writ.* 21, 505–537.

van Geert, P. (2008). The dynamic systems approach in the study of L1 and L2 acquisition: an introduction. *Modern Lang. J.* 92, 179–199.

Wilkinson, G. S. (1993). *Wide Range Achievement Test 3.* Wilmington, DE: Wide Range.

Yopp, H. K. (1988). The validity and reliability of phonemic awareness tests. *Read. Res. Q.* 23, 159–177.

# Alphabetism in reading science

**David L. Share** *

Department of Learning Disabilities, Faculty of Education, University of Haifa, Haifa, Israel
*Correspondence: dshare@edu.haifa.ac.il

There has been mounting concern among social scientists that conclusions from studies conducted on highly educated populations from affluent European cultures may have limited applicability to human behavior in general (Henrich et al., 2010). Similar reservations have also been voiced in the fields of language (Evans and Levinson, 2009) and literacy (Share, 2008a; Frost, 2012). Reading research, in particular, has been overwhelmingly dominated by work on English, which appears to be an outlier among European alphabets (Seymour et al., 2003; Share, 2008a). I have argued that because spelling–sound relations are so complex in English orthography, much of reading research has been confined to a narrow Anglocentric research agenda addressing theoretical and applied issues with only limited relevance for a universal science of reading and literacy.

My intention here is not to reiterate my 2008 arguments or even expand them, but to move on to another major obstacle to progress. Before moving on, however, I would like to add a note of optimism to the Anglocentrism debate. In recent years, interest in other languages has indeed begun to emerge from the shadows probably because the scientific community of Anglo-American reading researchers has felt itself "come of age" as a substantial body of well-replicated and converging findings has coalesced in recent years, at least on several key topics such as word identification and dyslexia (Vellutino et al., 2004; Snowling and Hulme, 2005; DeHaene, 2009; Rayner et al., 2012). The field is now witnessing important first steps toward universal models of reading (Perfetti, 2003; Perfetti et al., 2005; Ziegler and Goswami, 2005; Frost, 2012) as

well as a growing number of linguistically and grammatologically informed studies emerging outside the confines of English and other European alphabets (Nag and Perfetti, 2014; Saiegh-Haddad and Joshi, 2014; Verhoeven and Perfetti, 2014). It is still the case, nonetheless, that the theoretical and applied frameworks developed for English are all too frequently being applied to other languages and writing systems without due consideration for linguistic and writing system diversity. Almost all publications by English-language researchers continue to omit any "…in English" qualification in the titles of their papers—"A New Whiz-Bang+++ Model of Learning to Read"… in English?—as if the results of studies conducted in English alone enjoy the privileged status of universal applicability, unlike researchers investigating other languages who are obliged to qualify their findings by adding the "…in Chinese/Arabic/Korean etc." disclaimer which automatically demarcates the findings as language-specific and hence not necessarily universally applicable.

Here, I focus on yet another "-ism," which I call "alphabetism"; the belief that alphabetic writing systems are inherently superior to non-alphabetic systems, and which, like Anglocentrism, has also stymied psychologists' and educators' thinking about learning to read across diverse writing systems. Here too, I join other scholars who have also expressed concerns about "alphabetolatry," or alphabetic "supremacism" (e.g., Rogers, 1995). Looking around the globe, it is apparent that most individuals do not acquire literacy in a European alphabet, yet in many parts of the (non-European) world, the belief that alphabetic orthographies are the ideal has led to calls

to alphabetize or discard non-alphabetic scripts. Needless to say, these proposals have profound ramifications for instruction and curriculum.

In the past, many influential Western scholars explicitly argued that alphabets are inherently superior to non-alphabetic writing systems (Taylor, 1883; Gelb, 1963; Havelock, 1982). The shelves of most college libraries abound with volumes whose very titles idealize the alphabet (e.g., Diringer's *The Alphabet: A Key to the History of Mankind;* Moorhouse's *Triumph of the Alphabet*). When reading researchers today seek enlightenment on the subject of writing systems they refer to Gelb—the founding father of the field of "grammatology" (Gelb, 1963). Like Taylor (1883) before him, Gelb (1963) propounded an evolutionary view of writing system history from "primitive" pre-alphabetic systems to alphabetic. Consistent with the "ontogeny recapitulates phylogeny" idea, Gelb's inexorable "*three great steps [logographic-to-syllabic-to-alphabetic] by which writing evolved from the primitive stages to a full alphabet*" (p. 203) was embraced by almost all reading researchers, despite its repudiation by subsequent scholarship in the field of writing systems research (Mattingly, 1985; Olson, 1989; Daniels, 1992, in press; Rogers, 2005; Coulmas, 2009). Foremost among these, perhaps, was Ferreiro in her Piagetian classic *Literacy before Schooling* (Ferreiro and Teberosky, 1979) and, subsequently, a series of stage-oriented theories of reading and writing development (Piagetian and non-Piagetian alike) all referring to pre-alphabetic and alphabetic stages (Gough and Hillinger, 1980; Marsh et al., 1981; Frith, 1985; Ehri, 2005). It needs to be pointed out, however, that the

"culture" of alphabetism, like culture in general, is often "invisible"; its presence more often discernible in acts of omission rather than commission. Nonetheless, this alphabetic bias is ubiquitous and is manifest in;

1. Unqualified generalizations about reading "across languages" and/or "across orthographies" in papers that refer almost exclusively to English or to European alphabets (Ziegler and Goswami, 2005; Goswami, 2010; Ziegler et al., 2010; Caravolas et al., 2013; Ehri, 2014).

2. Implicit or explicit acceptance of Gelb's long-discarded evolutionary theory in leading texts on reading development aimed at educators,... "*Taking the final step toward the creation of a true alphabetic writing system, the Greeks assigned a symbol to each consonant and vowel of their language... In many ways, the individual development of the children who are discovering the alphabetic principle in English writing recapitulates human history,*" (Moats, 2000, pp. 82–83).

3. Even the most up-to-date and authoritative texts on the psychology of reading (e.g., DeHaene, 2009; Rayner et al., 2012) continue to regurgitate Gelb's views.

"*[I]n an evolutionary sense, the alphabet is the "fittest..." p. 37*"; *The history of writing suggests a clear evolutionary trend...These systems evolved to a logographic system, which in turn evolved to syllabic systems and finally to alphabetic systems...Such an evolutionary argument suggests that alphabets are fitter (in the Darwinian sense)...* Rayner et al. (2012, pp. 46–47).

4. Reference to non-alphabetic systems as imperfect or defective (e.g., Hannas, 2003; Rayner et al., 2012) as well as attempts to reframe non-alphabetic systems such as the Brahmi-derived Indic (abugidic/aksharik) scripts as alphabetic (Rimzhim et al., 2014).

...*"The Semitic writing systems...and the languages of India still incompletely represent vowels. p 36... In this sense, many of these scripts are not fully alphabetic."* Rayner et al. p. 37.

"*The Phonecian system, however, was not perfect. It failed to represent all vowels... It was the Greeks who finally created the alphabet as we know it... For the first time in the history of mankind, the alphabet allowed the Greeks to have a complete graphic inventory of their language sounds.*" (DeHaene, 2009, p. 193).

"*The basic difference between Western alphabetic and East Asian syllabic writing acts on several levels to promote or inhibit creativity, particularly that associated with breakthroughs in science... syllabic literacy entails a diminished propensity for abstract and analytical thought... Certain Asian characteristics credited with blocking creativity, such as conservative political and social institutions and group-oriented behavior, derive in part from effects that the orthography has had on the minds of individuals,*" (Hannas, 2003, p. 203).

5. The use of alphabetic terminology (e.g., letters, graphemes) to describe and label the functional architecture (and even the anatomical brain structures) of reading ("letter detectors," "letterbox area," "universal letter shapes," DeHaene, 2009) purported to be universal in reading. Whereas the concept of a letter (or grapheme) is widely used (but not entirely unproblematic) in European alphabets, it has questionable applicability to many writing systems such as Chinese characters, Japanese Kanji, Brahmi-derived Indic aksharas or Mayan glyphs. It has even been suggested that the notion of the "phoneme" as the fundamental unit of analysis of speech may be an artifact of West European alphabetic literacy (Daniels, in press).

Although some initial thoughts have been offered as to when an alphabet may or may not be the appropriate orthography (e.g., Perfetti and Harris, 2013), this topic is new to the agenda of reading science. Some historical background on the alphabet provides a valuable perspective on this issue.

## SOME HISTORICAL BACKGROUND

Contrary to popular belief, the alphabet did not originate among Semitic speakers, or their Egyptian neighbors, but was a uniquely Greek creation invented only once, and probably on the basis of a fortuitous misunderstanding of Phoenician writing (Daniels, 1992). An alphabetic writing system, with full and equal representation of consonants and vowels, was ideally suited to the unique features of Indo-European languages (Diringer, 1948; Taylor, 1883). It added vowel notation to the Phoenician abjad, which was also a segmental/phonemic system but represented (and only needed to represent) consonants alone. Would an alphabet ever have been needed had there been no Indo-European languages in the world? Indo-European languages have a large inventory of complex syllable structures, far too many for a syllabary such as Japanese. And because vowels are essential constituents of root morphemes (bat/bet/bit/but/beet/bite... etc.) the Semitic abjad would have been inadequate.

This uniquely European mutation was first disseminated throughout Europe with the spread of Christianity, then across the globe by European colonizers, traders, and, above all, missionaries who never thought to question whether their own writing systems would be optimal for non-European languages. They took it for granted that the ideal orthography was alphabetic, operating on the principle of one letter for one sound (phoneme) for both consonants and vowels under the motto "consonants as in English, vowels as in Italian." (Gleason, 1996).

But are alphabets optimal? Well, we really don't know. There is, however, evidence suggesting that it cannot be assumed that alphabets are inherently superior and therefore the default choice of script. There are at least four lines of counterevidence converging on the conclusion that syllable-based writing systems are, in many cases, superior to alphabets.

1. Psychoacoustically, syllables are more "real" than phonemes (Liberman et al., 1967). Data from illiterates (Morais et al., 1987), pre-literates (Liberman et al., 1974), or persons literate in purely morpho-syllabic or syllabic systems (Read et al., 1986) confirm that syllables are easier to deal with than phonemes.

2. Historically, syllabaries appeared earlier and more often in ancient times (Rogers, 2005; Gnanadesikan, 2009),

whereas the alphabet was a relative late-comer in the history of writing and appeared only once (Daniels, 1992). All new writing systems invented by non-literates who know that writing exists are syllabaries (Daniels, 1992).

3. Anthropologists have reported widespread literacy among indigenous peoples using syllabic systems in North America (among the Cree, McCarthy, 1995; Cherokee, Walker, 1969); Africa, (Scribner and Cole, 1981); and the Philippines (the Hanuno'o, Kuipers and McDermott, 1996).

4. Quasi-experimental studies suggest that young children are able to learn to read syllabically (abugidically) more easily than phonemically/alphabetically (e.g., Gleitman and Rozin, 1973; Asfaha et al., 2009). Asfaha et al., for example, investigated reading acquisition in four Eritrean languages that use either syllabic (abugidic) (CV) Geèz (Tigrinya and Tigre) or alphabetic Latin-based scripts (Kunama and Saho). Instruction in alphabetic Saho focuses on CV units, whereas alphabetic Kunama is taught alphabetically, i.e., phonemically. Asfaha et al. found that first graders learned to read the non-alphabetic Ge'ez far more easily than the alpha-betic scripts in spite of the larger num-ber of signs. Moreover, the abugidic CV-level teaching of alphabetic Saho produced superior results compared to alphabetic teaching of (alphabetic) Kunama[1]. There are also studies show-ing that children and adults who have struggled to learn to read alphabeti-cally find it easier to learn to read a syllable-based orthography than an alphabetic orthography (Gleitman and Rozin, 1973; Moore et al., 2014).

My aim here is not to show that syl-labic writing systems are superior to alpha-betic systems, but simply that alphabets cannot be assumed a priori to be inher-ently superior to other writing systems. The crucial question (as discussed by Perfetti and Harris, 2013) is the match between language structure and writing system, in particular the size and complex-ity of the syllable inventory.

This issue leads to the more general question, What makes an orthography more or less optimal?

## WRITING SYSTEM EFFICIENCY AND A UNIVERSAL MODEL OF LEARNING TO READ

An efficient writing system must do two things simultaneously: represent sound and meaning (Rogers, 1995; Share, 2008b; Frost, 2012). This is no simple task, because these two aspects of writing must often be traded off against each other. I have termed these two dimensions of orthography *decipherability* and *autom-atizability/unitizability* (Share, 2008b). Orthographies can be regarded as dual-purpose devices serving the distinct needs of novices and experts (see Share, 2008a). Because all words are initially unfamiliar, the reader needs a means of deciphering new letter strings unassisted (see Share, 1995, 2008b, for more detailed discussion, and Ziegler et al., 2014 for an explicit com-putational instantiation of this notion). Here, the representation of recombinant sub-lexical phonological elements (either syllabic, sub-syllabic, or phonemic) is fundamental if a script is destined to be decipherable and learnable (Mattingly, 1985; Unger and DeFrancis, 1995). But the essence of skilled reading (as is the case with all human skills) is speed and effort-lessness. To achieve fluent, automatized reading, the expert-to-be requires unique word-specific or morpheme-specific let-ter configurations that can be "unitized" and automatized for instant access to units of meaning. Here morpheme-level (and probably also word-level) representation is essential[2]. Both morpheme *distinctive-ness* (<rite/right>) as well as morpheme *constancy* (<soft/soften>) are crucial for rapid silent reading (Rogers, 1995).

The corollary to this orthographic dualism is what goes on inside the reader's head. Initially unfamiliar words and morphemes become familiar units, as the novice reader's orthographic lexicon begins to grow. This "unfamiliar-to-familiar" or "novice-to-expert" dualism highlights the developmental transition (common to all human skill learn-ing) from slow, deliberate, step-by-step, unskilled performance to rapid, auto-matic, one-step (i.e., unitized) skilled pro-cessing. And because this broader dualism applies to *all* words in *all* orthographies, it seems a useful platform for developing a universal theory of learning to read.

## REFERENCES

Asfaha, Y. M., Kurvers, J., and Kroon, S. (2009). Grain size in script and teaching: Literacy acquisition in Ge'ez and Latin. *Appl. Psycholinguist.* 30, 709–724. doi: 10.1017/S0142716409990087

Caravolas, M., Lervag, A., Defior, S., Malkova, G. S., and Hulme, C. (2013). Different patterns, but equivalent predictors, of growth in reading in con-sistent and inconsistent orthographies. *Psychol. Sci.* 20, 1–10. doi: 10.1177/0956797612473122

Cardoso-Martins, C. (1991). Awareness of phonemes and alphabetic literacy. *Br. J. Edu. Psychol.* 61, 164–173. doi: 10.1111/j.2044-8279.1991.tb00972.x

Coulmas, F. (2009). Evaluating merit—the evolution of writing reconsidered. *Writ. Syst. Res.* 1, 5–17, doi: 10.1093/wsr/wsp001

Daniels, P. T. (1992). "The syllabic origin of writ-ing and the segmental origin of the alphabet," in *The Linguistics of Literacy*, eds P. Downing, S. D. Lima, and M. Noonan (Amsterdam and Philadelphia, PA: John Benjamins), 83–110. doi: 10.1075/tsl.21.10dan

Daniels, P. T. (in press). "Writing Systems," in *The Blackwell Handbook of Linguistics*, 2nd Edn. eds M. Aronoff and J. Rees-Miller (Malden, MA: Wiley-Blackwell).

DeHaene, S. (2009). *Reading in the Brain*. New York, NY: Penguin.

Diringer, D. (1948). *The Alphabet: a Key to the History of Mankind.* London: Thames and Hudson.

Ehri, L. C. (2005). "Development of sight word read-ing: Phases and findings," in *The Science of Reading: A Handbook,* eds M. J. Snowling and C. Hulme (Malden, MA: Blackwell), 135–154.

Ehri, L. C. (2014). Orthographic mapping in the acquisition of sight word reading, spelling mem-ory, and vocabulary learning. *Sci. Stud. Reading* 18, 5–21. doi: 10.1080/10888438.2013.819356

Evans, N., and Levinson, S. C. (2009). The myth of language universals: language diversity and its importance for cognitive science. *Behav. Brain Sci.* 32, 429–492. doi: 10.1017/S0140525X0999094X

---

[1]This is by no means the first time an alphabetic writing system has been taught syllabically (see, for example, Cardoso-Martins, 1991; Liow Rickard and Lee, 2004). It is worth noting that Noah Webster's "blue-back speller" (first published in 1785) was also a syllable-based method of teaching English.

[2]I gloss over deep and unresolved issues regarding the linguistic and psycholinguistic status of morphemes and words, how these units might change in the course of literacy development, and how they are represented in diverse orthographies.

Ferreiro, E., and Teberosky, A. (1979). *Literacy Before Schooling*. Exeter, NH: Heineman.

Frith, U. (1985). "Beneath the surface of developmental dyslexia," in *Surface Dyslexia,* eds K. E. Patterson, J. C. Marshall, and M. Coltheart (London: Lawrence Erlbaum Associated Ltd), 301–322.

Frost, R. (2012). Towards a universal model of reading. *Behav. Brain Sci.* 35, 263–279. doi: 10.1017/S0140525X11001841

Gelb, I. J. (1963). *A Study of Writing*, 2nd Edn. Chicago, IL: University of Chicago Press.

Gleason, A. (1996). "Christian missionary activities," in *The World's Writing Systems*, eds P. T. Daniels and W. Bright (New York, NY: Oxford University Press), 777–780.

Gleitman, L. R., and Rozin, P. (1973).Teaching reading by use of a syllabary. *Read. Res. Q.* 8, 447–483. doi: 10.2307/747169

Gnanadesikan, A. E. (2009). *The Writing Revolution: Cuneiform to the Internet*. Chichester, UK: Wiley-Blackwell.

Goswami, U. (2010). "A psycholinguistic grain size view of reading development across languages," in *Reading and Dyslexia in Different Orthographies*, eds N. Brunswick, S. McDougall, and P. de Mornay Davies (Hove, UK: Psychology Press), 23–42.

Gough, P. B., and Hillinger, M. L. (1980). Learning to read: an unnatural act. *Bull. Orton Soc.* 30, 179–196. doi: 10.1007/BF02653717

Hannas, W. C. (2003). *The Writing on the Wall: How Asian Orthography Curbs Creativity*. Philadelphia, PA: University of Philadelphia Press.

Havelock, E. A. (1982). *The Literate Revolution in Greece and its Cultural Consequence*. Princeton, NJ: Princeton University Press.

Henrich, J., Heine, S. J., and Norenzayan, A. (2010). The weirdest people in the world? *Behav. Brain Sci.* 33, 61–135. doi: 10.1017/S0140525X0 999152X

Kuipers, J. C., and McDermott, R. (1996). "Insular Southeast Asian scipts," in *The World's Writing Systems*, eds P. T. Daniels and W. Bright (New York, NY: Oxford University Press), 474–484.

Liberman, A. M., Cooper, F. S., Shankweiler, D., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychol. Rev.* 74, 431–461. doi: 10.1037/h0020279

Liberman, I. Y., Shankweiler, D., Fischer, F. W., and Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *J. Exp. Child Psychol.* 18, 201–212. doi: 10.1016/0022-0965(74)90101-5

Liow Rickard, S. J., and Lee, L. C. (2004). Metalinguistic awareness and semi-syllabic scripts: Children's spelling errors in Malay. *Read. Writ.* 17, 7–26. doi: 10.1023/B:READ.0000013833. 79570.de

Marsh, G., Friedman, M., Welsh, V., and Desberg, P. (1981). A cognitive-developmental theory of reading acquisition. *Read. Res.* 3, 199–219.

Mattingly, I. (1985). Did orthographies evolve? *Remed. Special Edu.* 6, 18–23 doi: 10.1177/07419 3258500600605

McCarthy, S. (1995). "The Cree syllabary and the writing system riddle: a paradigm in crisis," in *Scripts and Literacy,* eds I. Taylor and D. R. Olson (Dordrecht: Kluwer), 59–75. doi: 10.1007/978-94-011-1162-1_5

Moats, L. C. (2000). *Speech to Print: Language Essentials for Teachers*. Baltimore, MD: Paul H. Brookes Publishing.

Moore, M. W., Brendel, P. C., and Fiez, J. A. (2014). Reading faces: investigating the use of a novel face-based orthography in acquired alexia. *Brain Lang.* 129, 7–31. doi: 10.1016/j.bandl.2013.11.005

Morais, J., Alegria, J., and Content, A. (1987). The relationships between segmental analysis and alphabetic literacy: an interactive view. *Curr. Psychol. Cogn.* 7, 415–438.

Nag, S., and Perfetti, C. A. (2014). Reading and writing: insights from the alphasyllabaries of South and Southeast Asia. *Writ. Syst. Res.* 6, 1–9. doi: 10.1080/17586801.2014.883787

Olson, D. (1989). *Writing*. Encyclopedia Britannica.

Perfetti, C. A. (2003). The universal grammar of reading. *Sci. Stud. Read.* 7, 3–24.

Perfetti, C. A., and Harris, L. N. (2013). Universal reading processes are modulated by language and writing system. *Lang. Learn. Dev.* 9, 296–316. doi: 10.1207/S1532799XSSR0701_02

Perfetti, C. A., Liu, Y., and Tan, L. H. (2005). The lexical constituency model: Some implications of research on Chinese for general theories of reading. *Psychol. Rev.* 112, 43–59.

Rayner, K., Pollatsek, A., Ashby, J., and Clifton, C. Jr. (2012). *Psychology of Reading*. 2nd Edn. New York, NY: Psychology Press.

Read, C., Zhang, Y. F., Nie, H. Y., and Ding, B. Q. (1986). The ability to manipulate speech sounds depends on knowing alphabetic writing. *Cognition* 24, 31–44. doi: 10.1016/0010-0277(86)90003-X

Rimzhim, A., Katz, L., and Fowler, C. A. (2014). Brāhmī-derived orthographies are typologically Āksharik but functionally predominantly alphabetic. *Writing Syst. Res.* 6, 41–53. doi: 10.1080/17586801.2013.855618

Rogers, H. (1995). "Optimal orthographies," in *Scripts and Literacy,* eds I. Taylor and D. R. Olson (Dordrecht: Kluwer Academic Publishers), 31–43. doi: 10.1007/978-94-011-1162-1_3

Rogers, H. (2005). *Writing Systems: A Linguistic Approach*. Malden, MA: Blackwell.

Saiegh-Haddad, E., and Joshi, M. (eds.). (2014). *Handbook of Arabic Literacy and Orthography*. New York, NY: Springer. doi: 10.1007/978-94-017-8545-7

Scribner, S., and Cole, M. (1981). *The Psychology of Literacy*. Cambridge, MA: Harvard University Press.

Seymour, P. H. K., Aro, M., and Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *Br. J. Psychol.* 94, 143–174. doi: 10.1348/000712603321661859

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

Share, D. L. (2008a). On the Anglocentricities of current reading research and practice: the perils of over-reliance on an "outlier" orthography.

*Psychol. Bull.* 134, 584–616. doi: 10.1037/0033-2909.134.4.584

Share, D. L. (2008b). "Orthographic learning, phonology and the self-teaching hypothesis," in *Advances in Child Development and Behavior*, Vol. 36, ed R. Kail (Amsterdam: Elsevier), 31–82.

Snowling, M. J., and Hulme, C. (2005). *The Science of Reading: A Handbook*. Oxford, UK: Blackwell.

Taylor, I. (1883). *The Alphabet: an Account of the Origin and Development of Letters*, 2nd Edn 1899. London: Kegan Paul, Trench.

Unger, J. M., and DeFrancis, J. (1995). "Logographic and semasiographic writing systems: a critique of Sampson's classification," in *Scripts and Literacy*, eds I. Taylor and D. R. Olson (Dordrecht: Kluwer), 45–58. doi: 10.1007/978-94-011-11 62-1_4

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., and Scanlon, D. M. (2004). Specific reading disability (dyslexia): what have we learned from the past four decades? *J. Child Psychol. Psychiatry* 45, 2–40. doi: 10.1046/j.0021-9630.2003.00305.x

Verhoeven, L., and Perfetti, L. C. (eds.). (2014). *Reading Acquisition: Cross-Linguistic and Cross-script Perspectives*. Cambridge: Cambridge University Press.

Walker, W. (1969). Notes on native writing systems and the design of native literacy programs. *Anthropol. Linguistics* 11, 148–166.

Ziegler, J. C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faísca, L., et al. (2010). Orthographic depth and its impact on universal predictors of reading: a cross-language investigation. *Psychol. Sci.* 21, 551–559. doi: 10.1177/0956797610363406

Ziegler, J. C., and Goswami, U. (2005). Reading acquisition, developmental dyslexia and skilled reading across languages: a psycholinguistic grain size theory. *Psychol. Bull.* 131, 3–29. doi: 10.1037/0033-2909.131.1.3

Ziegler, J. C., Perry, C., and Zorzi, M. (2014). Modelling reading development through phonological decoding and self-teaching: implications for dyslexia. *Philos. Trans. R. Soc. B* 369, 1–9.

# Alphabetism and the science of reading: from the perspective of the akshara languages

## Sonali Nag [1,2] *

[1] Early Childhood and Primary School Programmes, The Promise Foundation, Bangalore, India
[2] Department of Experimental Psychology, University of Oxford, Oxford, UK
*Correspondence: sonalinag@t-p-f.org

A commentary on

**Alphabetism in reading science**
*by Share, D. L. (2014). Front. Psychol. 5:752. doi: 10.3389/fpsyg.2014.00752*

An interesting area of enquiry in reading science is the ways in which different writing systems represent language. Discussions have centered around the adaptations seen between writing systems and languages (Perfetti and Harris, 2013) and the related notions of deservedness of a writing system for a language (Halliday, 1977), optimality (Frost, 2012) and level of orthography-language or "grapholinguistic" equilibrium (Seidenberg, 2011). Among the many ideas of relative goodness of writing systems is also a misplaced superiority assigned to alphabet-based orthographies, which has been critically labeled as "alphabetism" (Share, 2014). Share counters the superiority claim with psychoacoustic, historic, anthropological and preliminary experimental evidence to show that syllable-based writing systems are perhaps the better system, at least for some aspects of the orthography-language relationship. The defining parameters for placing symbol systems in a hierarchy are however, as yet, unclear (see Frost, 2012 for a discussion). It is for this very reason that reading research (and the practice it influences) must be alert to unqualified generalizations made from studies conducted in a single writing system. Evidence from robust cross-orthographic experimentation is the best moderator of such universalism. The burgeoning body of work from the Chinese languages has for example

broadened the field, and perhaps snuffed out "alphabetism" in some domains (e.g., neural bases of reading and the preferred ordering of symbols as linear: Perfetti et al., 2010). Some insights are now also available from experimental work and surveys in Japanese Hiragana (e.g., Fletcher-Flinn et al., 2014). More recently, research in the Indic alphasyllabaries highlights the role of orthography-specific investigations in the quest for a more inclusive reading science (Nag, 2007, 2014).

The orthographies of South and Southeast Asia descend from the ancient script of Brahmi and together may be referred to as the Indic alphasyllabaries. The symbol unit of these orthographies is the *akshara*. The surface organization of each unit is typically a symbol block with one or more phonemic markers. An akshara may represent a vowel (/V/), a consonant (/C/), a consonant with the inherent vowel /a/ or other marked vowels (/Ca/, /CV/), and consonant clusters with either the inherent or marked vowels (e.g., /CCa/, /CCV/, /CCCV/). The mapping of word level phonology to specific akshara is decided by a rule of re-syllabification where post-vocalic consonants form the next akshara. To illustrate with number names from the Indo-Aryan language of Hindi, the akshara in *shunya* (zero) follow the rule of re-syllabification with the second akshara formed by a coda-open syllable concatenation (शुन्य, <CV.CCa> "*shu.nya*," the coda of the first syllable is pinned to the next syllable to make the symbol block "*nya*"). The transcription in the akshara system is typically complete, though mapping to phonology is

variable. For example, *nau* (nine) represents an open syllable (नौ, <CV>, "*nau*"), *das* (10) a body and coda (दस, <CV.C°>, "*da.s*"), and *gyaarah* (11) an open syllable, a body and a coda (ग्यारह, <CCV.Ca.C°>, "*gyaa.ra.h*"). There are further conditional rules in Hindi such as vowel suppression where the akshara-to-phonology representation becomes somewhat opaque. Thus, in *bees*, *thees*, and *chalees* (20, 30, and 40) the word-final /s/ is written with an akshara carrying the inherent vowel /a/ but this vowel is suppressed in pronunciation (i.e., <C°>), thus बीस and तीस, <CV.C°>, and चालीस, <CV.CV.C°>. Similar schwa suppression is also seen in the earlier examples, *das* and *gyaarah*.

Akshara-based orthographies such as Bengali, Gujarati, Lao, Tamil, and Sinhala each have similarly well-defined orthographic principles. Whereas in other phonologically-based writing systems like the alphabet and the abjad, the orthographic representation of one particular sub-lexical level predominates, the mapping to phonology in the akshara-based orthographies is defined by context. If appearing single, then the akshara is typically an orthographic syllable, but if in a string, language-specific rules delimit orthographic representation. Thus, akshara units map to multiple levels of phonology. Given the current state of the science, this psycholinguistic design of the akshara requires greater examination. But what should be immediately clear is that the pre-eminence given to the phoneme in several accounts of orthographic representation (e.g., Katz and Frost, 1992; Ziegler and Goswami,

2005) is an alphabet-centric model. The akshara based psycholinguistic tradition has instead drawn upon the role of orality in literacy development (Patel and Soper, 1987; Patel, 1996, 2004), the articulatory features of single akshara and word-level prosody (Pandey, 2007, 2014), the nature and scope of akshara-language mapping (Sircar and Nag, 2013; Nag, 2014), the cognitive bases of reading acquisition (Prakash et al., 1993; Nag and Snowling, 2012) and the profiles of impairment in adult clinical conditions (Karanth, 2002). What is needed for a universal theory of reading (and spelling) development is a delineation of the cognitive-linguistic mechanisms associated with a writing system that has the facility for multiple levels of sub-lexical representation. Constructs that have shown promise include syllable weight and the *mora*. These constructs pick out the regularities in spelling-sound mapping and hence may be the principle that makes learning of the orthography-language connections secure. Ideas about syllable weight and the *mora* have deep roots in linguistic science but are yet to inform discourse in the reading science.

The symbol set is another case in point. The number of letters in alphabet-based systems is small, and symbol learning is completed within the first year of instruction. In contrast to the small set or a *contained orthography*, are systems with several thousand symbols. The characters for a Chinese language such as Mandarin is one example of an *extensive orthography*. In the Indic alphasyllabaries, the number of akshara that can be hypothetically constructed also run into thousands, with two constraints defining the learning space. First, a manageable set of consonant and vowel phonemic markers aid akshara construction, bringing economy to the learning task. Second, the number of akshara that are phonotactically implausible are far more in number, although the number that are in use and hence encountered in print still runs into hundreds. Not surprisingly, a corollary of an extensive symbol set is that symbol learning continues well into middle school and beyond. If the received wisdom is that children typically always know the alphabet by the end of the first year then it is not hard to see how the pace of learning in the extensive orthographies might be perceived. "Slow" learning

then becomes one reason to invoke "alphabetism," with suggestions that the local orthographies are too difficult for fast paced literacy learning.

Furthermore, a comprehensive theory of literacy learning will have to factor in the learning mechanisms involved in the akshara languages, particularly the role of domains such as visual memory, morphology and syntax, and several other aspects of the orthography. Some of these include non-linear symbol arrangements (Vaid and Gupta, 2002; Kandhadai and Sproat, 2010; Winskel and Perea, 2014), unmarked and inherent symbol features (Nag, 2007; Bhide et al., 2014), visually complex symbol sets (Nag et al., 2014) and word types differing because of symbol characteristics (Nag, 2014; Wijayathilake and Parrila, 2014) or morpho-orthographic characteristics (Rao et al., 2012). A step before the hunt for higher-order universals would be to bring focus in reading science on these kinds of particularities.

## ACKNOWLEDGMENT

## REFERENCES

Bhide, A., Gadgil, S., Zelinsky, C. M., and Perfetti, C. (2014). Does reading in an alphasyllabary affect phonemic awareness? Inherent schwa effects in Marathi-English bilinguals. *Writ. Syst. Res.* 6, 73–93. doi: 10.1080/17586801.2013.855618

Fletcher-Flinn, C. M., Thompson, G. B., Yamada, M., and Meissel, K. (2014). Learning of a formation principle for the secondary phonemic function of a syllabic orthography. *Read. Writ.* 27, 875–904. doi: 10.1007/s11145-013-9473-0

Frost, R. (2012). Towards a universal model of reading. *Behav. Brain Sci.* 35, 263–329. doi: 10.1017/S0140525X11001841

Halliday, M. A. K. (1977). "Ideas about language," in *Aims and Perspectives in Linguistics. Occasional Papers, No. I* (Applied Linguistics Association of Australia), 32–55.

Kandhadai, P., and Sproat, R. (2010). Impact of spatial ordering of graphemes in alphasyllabic scripts on phonemic awareness in Indic languages. *Writ. Syst. Res.* 2, 105–116. doi: 10.1093/wsr/wsq009

Karanth, P. (2002). Reading into reading research through nonalphabetic lenses. *Top. Lang. Disord.* 22, 20–31. doi: 10.1097/00011363-200211000-00006

Katz, L., and Frost, R. (1992). "The reading process is different for different orthographies: the orthographic depth hypothesis," in *Orthography, Phonology, Morphology, and Meaning: Advances in Psychology*, Vol. 94, eds R. Frost and L. Katz (North Holland: Elsevier), 67–84.

Nag, S. (2007). Early reading in Kannada: the pace of acquisition of orthographic knowledge and phonemic awareness. *J. Res. Read.* 30, 7–22. doi: 10.1111/j.1467-9817.2006.00329.x

Nag, S. (2014). Akshara-phonology mappings: the common yet uncommon case of the consonant cluster. *Writ. Syst. Res.* 6, 105–119. doi: 10.1080/17586801.2013.855621

Nag, S., and Snowling, M. J. (2012). Reading in an alphasyllabary: implications for a language-universal theory of learning to read. *Sci. Stud. Read.* 16, 404–423. doi: 10.1080/10888438.2011.576352

Nag, S., Snowling, M., Quinlan, P., and Hulme, C. (2014). Child and symbol factors in learning to read a visually complex writing system. *Sci. Stud. Read.* 18, 1–16. doi: 10.1080/10888438.2014.892489

Pandey, P. (2014). Akshara-to-sound rules for Hindi. *Writ. Syst. Res.* 6, 54–72. doi: 10.1080/17586801.2013.855622

Pandey, P. K. (2007). "Akshara as the minimum articulatory unit," in *The Indic Scripts: Palaeographic and Linguistic Perspectives*, eds P. G. Patel, P. Pandey, and D. Rajgor (New Delhi: DK Printworld), 167–232.

Patel, P. G. (1996). Linguistic and cognitive aspects of the Orality-Literacy complex in Ancient India. *Lang. Commun.* 16, 315–329.

Patel, P. G. (2004). *Reading Acquisition in India: Models of Learning and Dyslexia*. New Delhi: Sage Publishers.

Patel, P. G., and Soper, H. V. (1987). Acquisition of reading and spelling in a syllabo-alphabetic writing system. *Lang. Speech* 30, 69–81.

Perfetti, C. A., and Harris, L. N. (2013). Universal reading processes are modulated by language and writing system. *Lang. Learn. Dev.* 9, 296–316. doi: 10.1080/15475441.2013.813828

Perfetti, C. A., Nelson, J., Liu, Y., Fiez, J., and Tan, L.-H. (2010). "The neural bases of reading: universals and writing system variations," in *The Neural Basis of Reading*, eds P. Cornelissen, M. Kringelbach, and P. Hansen (New York, NY: Oxford University Press), 147–172.

Prakash, P., Rekha, D., Nigam, R., and Karanth, P. (1993). "Phonological awareness, orthography and literacy," in *Literacy and Language Analysis*, ed R. J. Scholes (Hillsdale, NJ: Lawrence Erlbaum Associates), 55–70.

Rao, C., Soni, S., and Chatterjee Singh, N. (2012). The case of the neglected alphasyllabary: orthographic processing in Devanagari. *Brain Behav. Sci.* 35, 302–303. doi: 10.1017/S0140525X12000313

Seidenberg, M. S. (2011). "Reading in different writing systems: one architecture, multiple solutions," in *Dyslexia Across Languages: Orthography and the Brain-Gene-Behavior Link*, eds P. McCardle, B. Miller, J. R. Lee, and O. J. L. Tzeng (Baltimore, MD: Paul H. Brookes), 146–168.

Share, D. L. (2014). Alphabetism in reading science. *Front. Psychol.* 5:752. doi: 10.3389/fpsyg.2014.00752

Sircar, S., and Nag, S. (2013). "Akshara-syllable mappings in Bengali: a language-specific skill for reading," in *South and South-East Asian Psycholinguistics*, eds H. Winskell and P. Padakanayya (Cambridge: Cambridge University Press), 202–211.

Vaid, J., and Gupta, A. (2002). Exploring word recognition in a semi-alphabetic script: the case of Devanagari. *Brain Lang.* 81, 679–690. doi: 10.1006/brln.2001.2556

Wijayathilake, M. A. D. K., and Parrila, R. (2014). Predictors of word reading skills in good and struggling readers in Sinhala. *Writ. Syst. Res.* 6, 120–131. doi: 10.1080/17586801.2013.846844

Winskel, H., and Perea, M. (2014). Can parafoveal-on-foveal effects be obtained when reading an unspaced alphasyllabic script (Thai)? *Writ. Syst. Res.* 6, 94–104. doi: 10.1080/17586801.2013.843440

Ziegler, J. C., and Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychol. Bull.* 131, 3–29. doi: 10.1037/0033-2909.131.1.3

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

# Cognitive flexibility predicts early reading skills

## Pascale Colé[1], Lynne G. Duncan[2] * and Agnès Blaye[1]

[1] Laboratoire de Psychologie Cognitive, UMR-7290, Aix-Marseille University, Marseille, France
[2] School of Psychology, University of Dundee, Dundee, UK

An important aspect of learning to read is efficiency in accessing different kinds of linguistic information (orthographic, phonological, and semantic) about written words. The present study investigates whether, in addition to the integrity of such linguistic skills, early progress in reading may require a degree of cognitive flexibility in order to manage the coordination of this information effectively. Our study will look for evidence of a link between flexibility and both word reading and passage reading comprehension, and examine whether any such link involves domain-general or reading-specific flexibility. As the only previous support for a predictive relationship between flexibility and early reading comes from studies of reading comprehension in the opaque English orthography, another possibility is that this relationship may be largely orthography-dependent, only coming into play when mappings between representations are complex and polyvalent. To investigate these questions, 60 second-graders learning to read the more transparent French orthography were presented with two multiple classification tasks involving reading-specific cognitive flexibility (based on words) and non-specific flexibility (based on pictures). Reading skills were assessed by word reading, pseudo-word decoding, and passage reading comprehension measures. Flexibility was found to contribute significant unique variance to passage reading comprehension even in the less opaque French orthography. More interestingly, the data also show that flexibility is critical in accounting for one of the core components of reading comprehension, namely, the reading of words in isolation. Finally, the results constrain the debate over whether flexibility has to be reading-specific to be critically involved in reading.

Keywords: reading acquisition, reading comprehension, cognitive flexibility, semantics, phonology, executive function

## INTRODUCTION

Reading acquisition has mainly been investigated from a psycholinguistic perspective which has been instrumental in identifying the important developmental impact of linguistic skills such as phonological awareness (Harm and Seidenberg, 1999; Ziegler and Goswami, 2005; Sprenger-Charolles et al., 2006). However, reading can also be viewed as a complex cognitive task, which requires the capacity for the concurrent processing of multiple aspects of print, and which, as a result, may implicate more general cognitive processes, such as executive function (Cartwright, 2002, 2012; van der Sluis et al., 2007).

Executive function (EF) serves as an umbrella term for the control functions that monitor the cognitive processing involved in complex, goal-oriented tasks (Miyake et al., 2000; Best and Miller, 2010). The "unity and diversity" view of EF (Miyake et al., 2000; Miyake and Friedman, 2012), emphasizes a common underlying ability to maintain task goals (unity), together with three distinguishable components (diversity), namely shifting of mental sets, inhibition of prepotent responses and updating of working memory representations.

The focus of the present study will be on shifting, also described as cognitive flexibility. This refers to the ability to select adaptively among multiple representations of an object, perspectives or strategies in order to adjust to the demands of a situation (Chevalier and Blaye, 2009; Cragg and Chevalier, 2012; Diamond, 2013).

Cognitive flexibility is involved in the acquisition of theory of mind (Müller et al., 2005) but it is the role that flexibility is thought to play in academic learning skills (Bull and Scerif, 2001; Bull et al., 2008; Yeniad et al., 2013)[1] that has led to our focus on this aspect of EF in relation to reading acquisition. At present, evidence for a direct link to reading is mixed – although several studies that are largely restricted to the English language have supported a positive association between flexibility and reading (Cartwright, 2002, 2007; Cartwright et al., 2010; Kieffer et al., 2013), other studies have failed to find such a relationship among typical or disabled readers of Dutch and French (van der Sluis et al., 2004, 2007; Monette et al., 2011). The differences between these outcomes will be explored in the sections to follow by examining the tasks used, the type of reading skill and the domain specificity of flexibility skills.

Cognitive flexibility is most often examined using task-switching paradigms, measuring the ease of switching between different sets of sorting rules, which reveal initial successes between the ages of 3 and 5 years (Cragg and Chevalier, 2012), and from 7 to 9 years, an increasing capacity to deal with multiple dimensions in switching tasks (Anderson, 2002). The relatively late emergence of flexibility in task switching has been attributed to partial dependence on other EFs (Davidson et al., 2006; Garon

---

[1] However, see St Clair-Thompson and Gathercole (2006) for contrary evidence.

et al., 2008). Authors have variously emphasized the underlying role of: (1) inhibition, either the inhibition of the previous rule (Kirkham et al., 2003) or the disinhibition of the previously inhibited sorting rule (Müller et al., 2006; Chevalier and Blaye, 2008); and (2) working memory, as part of goal setting and maintenance (Marcovitch et al., 2007).

Other measures of flexibility such as fluency in producing multiple uses for a single object (Diamond, 2013) and matrix classification tasks (e.g., Piaget and Inhelder, 1958), reveal a more specific aspect of flexibility, which is conceptualized theoretically as the difficulty in processing two or more dimensions simultaneously. In the revised Cognitive Complexity and Control model (Zelazo et al., 2003), the processing of dimensions simultaneously is regarded as more complex than switching between dimensions and is thought to be constrained developmentally by the conscious (meta-cognitive) control required (Zelazo, 2004)[2].

Finally, it is evident that considerable overlap exists between cognitive flexibility and the Piagetian concept of decentration in concrete operational thinking (Miller, 2010), since both depend on the ability to focus on more than one dimension of a problem. This comparison with the more intensively researched Piagetian concept highlights interesting questions, in particular, whether flexibility can be considered to be domain-general versus domain specific, a question to which we return in our experimental work. An initial investigation of this question suggests that EF skills do not generalize between verbal and non-verbal stimuli, at least among the kindergartners studied (Foy and Mann, 2013).

Several authors have presented a case for the involvement of cognitive flexibility in the development of reading and reading-related skills. The emergence of meta-linguistic awareness, a key component of beginning reading, has been linked to concrete operational thinking, which shares features with cognitive flexibility (as discussed above). Meta-linguistic awareness entails the switching of attention from word meaning to consider other properties of language such as phonology. Tunmer et al. (1988) reported that Grade 1 phonological awareness was partly dependent on level of operativity in tasks such as matrix classification and class inclusion. More recently, Blair and Razza (2007) used an item-selection task (Jacques and Zelazo, 2001), requiring item representation along two dimensions, to reveal correlations between flexibility and both phonological awareness and letter knowledge among kindergartners. Pre-school associations have also been found between flexibility (Dimensional Change Card Sort task) and emergent literacy skills such as phonological and print awareness (Bierman et al., 2008), as well as between theory of mind (Unexpected Location/Contents and Mistaken Identity tasks), flexibility (Wisconsin Card Sorting task), and rhyming skill (Farrar and Ashwell, 2012).

Flexibility, as measured by matrix classification, has also been found to correlate directly with early word reading and

reading comprehension (Arlin, 1981; Hogan and Whitson, 1984). Berninger and Nagy (2008) account for such findings by proposing that flexibility may be required to establish cross-modal connections between spoken and written language and to acquire and coordinate multiple features of print (phonology, morphology, syntax, semantics) during the development of word recognition. If so, flexibility may also underpin reading comprehension which is thought to be the product of word recognition and oral language comprehension (Simple View of Reading, Gough and Tunmer, 1986; Tunmer and Chapman, 2013). Cartwright (2002, 2007) has further argued that cognitive flexibility will play an even more direct role in reading comprehension due to the requirement to process phonological codes for written word recognition simultaneously with the semantic information for comprehension.

Cartwright (2002) provided evidence for this latter claim by studying the cognitive flexibility of English-speaking second to fourth graders in relation to their reading comprehension. A general flexibility task (Bigler and Liben, 1992) was administered, requiring double classification of sets of line drawings of objects into a 2 × 2 matrix using visual (same color) and semantic (same superordinate category) dimensions simultaneously. Cartwright also examined a form of reading-specific flexibility, which involved classification of written words into a 2 × 2 matrix according to phonological (same initial phoneme) and semantic (same superordinate category) criteria. The results indicated that reading-specific flexibility contributed unique variance to reading comprehension beyond the (significant) contributions of age, general flexibility, pseudo-word naming and oral language comprehension. A second experiment, demonstrated that a group receiving a short training in reading-specific flexibility using the matrix classification task exhibited a significant improvement in reading comprehension at post-test, which was not observed among groups receiving training in general flexibility or in a control task (dominoes).

In a later study, Cartwright et al. (2010) showed that general and reading-specific flexibility both improved between 1st and 2nd grades and that this improvement was not explained by increases in decoding ability. While each type of flexibility correlated with reading comprehension, reading-specific flexibility again proved to be a robust and independent predictor of reading comprehension among these younger children, whereas general flexibility contributed no additional variance beyond reading-specific flexibility. Altogether, Cartwright argues that this set of findings constitutes evidence that cognitive flexibility plays an important role in reading development, and further, that the component most crucial to progress is domain-specific.

Recently, Kieffer et al. (2013) found that flexibility in the Wisconsin Card Sorting Test correlated with reading comprehension but not with performance in a task measuring letter and word identification among their Grade 4 readers from low-income backgrounds. The results of path analyses indicated that flexibility was a significant and independent predictor of reading comprehension beyond the control variables (letter/word identification, language comprehension, working memory, processing speed, phonological awareness). Flexibility also made an indirect contribution to reading comprehension via language comprehension,

---

[2]See Kloo and Perner (2003) and Kloo et al. (2010) for a related account in which flexibility is associated with the realization that a single object can be redescribed in a number of different ways.

which the authors interpreted as indicating that higher levels of flexibility may confer advantages in reading for meaning.

However, relations between flexibility and reading have proved more equivocal in other studies, especially those of reading acquisition in languages other than English. Monette et al. (2011) assessed flexibility among French-speaking kindergarteners' with two tasks: a card sort task requiring a switch between two sorting rules and an adapted version of the Trail-making test (Trails P; Espy and Cwik, 2004). They found that flexibility failed to predict a composite measure of the children's reading and writing skills in Grade 1. Although van der Sluis et al. (2007) did observe that flexibility scores from measures of task-switching efficiency were related to Dutch forth- and fifth-graders' accuracy in a timed word reading task, the relationship found was negative.

Further exploration of this topic is clearly required given the failures to replicate evidence that cognitive flexibility is positively associated with reading progress. Our first objective is to determine whether the flexibility required in considering two dimensions simultaneously primarily applies to learning to read in opaque orthographies like English (Cartwright, 2002; Cartwright et al., 2010). Berninger and Nagy's (2008) analysis points to a greater need for flexibility when mappings between the features of print are complex. Opaque orthographies have many-to-one or one-to-many mappings between orthography and phonology which slows the development of word reading (Seymour et al., 2003) and renders the activation of phonology from print difficult (Share, 2008). This may encourage beginning readers of English to make early use of the variety of information at their disposal (orthographic, phonological, semantic, contextual) and could account for the observed influence of reading-specific flexibility on reading comprehension (Cartwright et al., 2010). French has a more transparent system of grapheme–phoneme correspondences than English (Ziegler et al., 1996; Peereman et al., 2007; Moll et al., 2014), and French second-graders are known to make extensive use of phonological decoding in reading (Sprenger-Charolles et al., 2003). Hence, there may be less need for them to resort to other sources of information, raising the question of whether flexibility is critical for early reading comprehension in more transparent orthographies such as French.

A second, and related, objective is to test whether flexibility influences the reading of words in isolation as suggested by Berninger and Nagy (2008). Developmental models of reading comprehension give a central role to recognition of the written words that make up the sentences, paragraphs and text to be understood (Gough and Tunmer, 1986; Perfetti et al., 2005). Text reading comprehension is engaged by accessing the semantic code of words via visual recognition and the language processing mechanisms assemble these words into messages. The quality of access to word representations is critical within this framework and this dependence on the activation and manipulation of different codes (phonological, orthographic, semantic) makes it seem plausible that flexibility could play a role in this key aspect of reading comprehension. In our study, we attempt to answer this question with a single word reading task that requires activation not only of formal codes (phonological, orthographic) but also

semantic codes. Our word reading task, therefore, allows examination of whether flexibility contributes to reading comprehension via the recognition of words in isolation and access to their meanings,

In relation to our third objective, an important question raised by Cartwright's (2002, 2007) research bears on the domain-specificity of flexibility. Although most developmental research on flexibility does not consider the question of specificity, a few studies demonstrate that flexibility in matching tasks is highly dependent on the conceptual domain in question (Bialystok and Martin, 2004; Blaye et al., 2007; Maintenant and Blaye, 2008; Foy and Mann, 2013). While Cartwright's results could be considered as support for this view, the contrast between her reading-specific and general flexibility tasks were not entirely conclusive. In the general cognitive flexibility task, participants had to sort line drawings of objects by color and by the superordinate category that the objects referred to, whereas in the reading-specific flexibility task, they had to sort words by their initial phoneme and by the superordinate category that the words referred to. That is, two potential sources of difference were confounded: the tasks differ both in terms of sorting criteria (perceptual/semantic versus phonological/semantic) and the kind of stimuli to which these criteria are applied (written words versus pictures). Hence, previous work remains inconclusive about which of the two features (stimuli versus criteria) is related to reading. To overcome this limitation, our study manipulates stimuli while keeping criteria equivalent (phonological/semantic).

In sum, the present study aims to investigate three important questions: (1) Is flexibility necessary in learning to read orthographies that are less opaque than English? (2) Does flexibility play a role in word reading as well as reading comprehension? and (3) Is the flexibility that is associated with reading, domain specific or domain-general?

## MATERIALS AND METHODS
### PARTICIPANTS

The participants were 60 second-graders (36 girls and 24 boys) from five schools with a middle-class catchment area in Aix-en-Provence in France (mean age: 7.63 years; SD = 0.30 years). In line with French Institutional and National regulations, four types of authorization were obtained for participation in this study: (1) written consent from the school authorities (the Inspector of National Education in France) in response to a written description of the research objectives and procedure of the study to be conducted with the child at school; (2) the consent of the head-teacher of the elementary school on the basis of information about the experimental procedure; (3) written informed consent from the child's parents or guardians, in which it is explicitly explained that they can refuse to allow their child to participate without consequence for them or their child; and (4) children's final enrollment was based on their own voluntary participation.

There were three additional inclusion criteria: (1) native speakers of French; (2) a reading level at least at chronological age on the French standardized test, "l'Alouette" (Lefavrais, 1967); and (3) non-verbal reasoning skills above the 25th percentile using the Raven's Colored Progressive Matrices (PM47, Raven

et al., 1995). The Alouette test is standardized for children aged from 5 to 14 years and involves reading aloud a text of 265 words as quickly and accurately as possible. The text contains real words in meaningless but grammatically correct sentences. Performance is converted into a reading age according to a standardized procedure taking account both of total reading time and accuracy.

## MATERIALS
### Reading tasks[3]
***Pseudo-word decoding.*** Sixty pseudo-words between 2 and 6 letters in length (e.g., pirda) were presented on a sheet of paper (10 pseudo-words per line). All were regular with regard to grapheme–phoneme correspondences but 20 contained graphemes whose pronunciation was context-dependent (i.e., s = /s/ or /z/; g = /g/ or /j/; c = /k/ or /s/). The number of pseudo-words read aloud correctly within one minute was recorded.

***Word reading.*** Both the recognition and comprehension of words was assessed by asking children to read a list of 108 words silently and to circle any animal names ($n = 50$). Items were selected from the 1000 most frequent words in Manulex (Lété et al., 2004) and distractors came from semantic categories such as fruits, vegetables, modes of transport, clothes, etc. (e.g., hibou, fusée, balai, loup, zèbre, tapis [English translation: owl, rocket, broom, wolf, zebra, carpet]). The word list was distributed across 18 lines of text (six words per line, each containing 2–4 animal names). Animal names increased in difficulty according to length and regularity of grapheme–phoneme correspondences. The number of animal names circled correctly within one minute was recorded. The error rate was negligible ($M = 0.05$; SD = 0.02).

***Passage reading comprehension.*** Performance was averaged across two tests. The first assessed the comprehension of short passages of text. Children read each sentence aloud and then traced a route on a map (e.g., Je vais du garage à la poste en passant par le parc [English translation: I go from the garage to the post office through the park]). Children could return to the text as often as they needed to. In the second task, children read aloud sentences referring to action sequences and then mimed what they had just read (e.g., Avec l'autre main, je prends le plus petit rond et je le mets sur le sol [English translation: With the other hand, I take the smallest circle and put it on the ground]). This test evaluated comprehension of anaphors (e.g., Je prends le grand carré avec une main et je le mets dans la boîte [English translation : I take the big square with one hand and I put it in the box]) and spatial terms (e.g., Je le pose ensuite entre les deux ronds puis sous la boîte [English translation : Next I put it between the two circles then under the box]). For each of these two tasks, a score was computed as a ratio of the number of correct actions to total time taken (in seconds).

### Flexibility tasks
Two double classification tasks were derived from those used by Cartwright (2002), with the constraint of avoiding the potential confusion between the two types of differences that were present in the original versions of the tasks: (i) Word Flexibility – this was reading-specific as it involved the classification of printed words; and (ii) Picture Flexibility – this required classification of drawings and did not involve reading. Both tasks demanded the simultaneous processing of two dimensions: phonology and semantics. The experimenter first demonstrated the sorting of a set of 12 stimuli into a 4-cell matrix, explaining that sorting could be accomplished in two ways: According to what can be heard at the beginning of the picture name/word (phonological criterion) and according to the sorts of things the drawings/words referred to (semantic criterion). She then double-classified the 12 cards into the matrix, commenting on her performance: As you can see, I'm putting all the things starting with /p/ (pear, peach) into this row; and all the things starting with /b/into this row . . . . But look, in this column, I'm putting all the fruits . . . and in this one, I'm putting all the animals. Children then sorted five new sets of 12 cards and were asked to comment on each double classification.

Two points were awarded for each correct double classification with both criteria described verbally; 1 point for evidence of double classification in either card sorting or verbal justification; and 0 for any other performance[4]. Response time (in seconds) for each sorting trial was also computed. Performance was averaged across the five stimulus sets for each task and a flexibility score was computed as a ratio of accuracy to response time: $(Acc/RT)*10$.

## PROCEDURE
The children were tested in a quiet room within their schools over four sessions as follows: (1) Alouette reading, PM47; (2) word reading, passage reading comprehension; (3) word flexibility, pseudo-word decoding; and (4) picture flexibility. The order of the last two sessions was counterbalanced.

## RESULTS
**Table 1** describes participant characteristics and performance on the reading and cognitive flexibility tasks. Although $z$-scores are used in the regression analyses, untransformed scores are presented here for ease of interpretation. The children's mean reading age ($M = 94.65$ months; SD = 7.34; Range = 85–119) was ahead of chronological age [$t(59) = 2.89$, $p = 0.005$].

Correlations between variables are also reported in **Table 1**. As no significant correlations were observed involving chronological age or PM47, these variables were not entered in the final regression analyses. A preliminary series of regression analyses was also conducted, which established that inclusion of PM47

[3]The authors would like to express their gratitude to Liliane Sprenger-Charolles (personal communication) for generously allowing them to use her tests of word and non-word reading and passage-reading comprehension.

[4]Cartwright's (2002) procedure for item scoring is given here to ease comparison with her work : score = 3, child sorted correctly and provided a correct verbal justification; score = 2, child sorted incorrectly but provided a correct verbal justification for the Experimenter's sort; score = 1, child sorted correctly but gave an incorrect (or no) verbal justification; and score = 0, child sorted incorrectly and gave an incorrect (or no) verbal justification. Note that the scoring system differs slightly in the present study because the Experimenter did not demonstrate the correct sort if a child made an error during the experimental trials.

**Table 1 | Pearson product-moment correlations, means, standard deviations and range for age, PM47 (raw scores), pseudo-word decoding, word reading, passage reading comprehension, picture and word flexibility scores (N = 60).**

| Measures | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| 1. Chronological age | — | 0.178 | −0.134 | −0.001 | 0.031 | 0.174 | 0.202 |
| 2. PM 47 | | — | 0.059 | 0.101 | 0.243 | 0.020 | 0.204 |
| 3. Pseudo-word decoding | | | — | 0.533** | 0.482** | 0.210 | 0.305* |
| 4. Word reading | | | | — | 0.469** | 0.325* | 0.470** |
| 5. Passage reading comprehension | | | | | — | 0.293* | 0.530** |
| 6. Picture flexibility | | | | | | — | 0.642** |
| 7. Word flexibility | | | | | | | — |
| M | 91.55 | 29.07 | 39.15 | 15.95 | 23.10 | 0.56 | 0.67 |
| SD | 3.59 | 4.39 | 7.2 | 3.06 | 5.90 | 0.30 | 0.34 |
| Range | 84–100 | 16–35 | 22–53 | 8–22 | 13–41 | 0.07–2.08 | 0.16–1.90 |

*$p < 0.05$, two-tailed; **$p < 0.01$, two-tailed.*

scores did not alter the pattern of results reported in the final analyses. Word flexibility scores correlated positively not only with word reading and passage reading comprehension but also with pseudo-word decoding; whereas picture flexibility scores did not correlate significantly with pseudo-word decoding, but showed a positive association with the two reading measures that involved the processing of meaning (word reading, passage reading comprehension).

Two hierarchical regression analyses were conducted with passage reading comprehension as the criterion variable. The traditional linguistic predictors, pseudo-word decoding and word reading were entered on the first two steps. In Analysis A (**Table 2A**), word flexibility scores were entered on the third step and picture flexibility on the fourth step. In Analysis B, the flexibility tasks were entered in the reverse order. Altogether these four variables accounted for

nearly 40% of the variance in passage reading comprehension (**Table 2B**).

In each analysis, Word flexibility explained approximately 10% of the concurrent variance in passage reading comprehension over and above the more traditional linguistic predictors, and critically, after controlling for picture flexibility in Analysis B. In contrast, picture flexibility failed to explain any additional variance regardless of entry position.

Two new regression analyses were conducted with word reading as the criterion variable (**Tables 3A,B**). Decoding was entered as the first predictor, accounting for more than 28% of the variance. Picture flexibility contributed 4.8% of additional variance when entered before word flexibility, however, did not add any explanatory variance when entered after word flexibility. Word flexibility explained an additional 10.4% of the variance when entered before picture flexibility and 5.7% of the variance when entered on the final step; hence, confirming the critical role of the reading-specific, word flexibility task.

**Table 2 | Hierarchical regression analyses predicting passage reading comprehension with (A) word flexibility entered before picture flexibility; and (B) with picture flexibility entered before word flexibility.**

| | Predictors | Δ $R^2$ | β |
|---|---|---|---|
| **(A)** | | | |
| Step 1 | Pseudo-word decoding | 0.232*** | 0.482 |
| Step 2 | Word reading | 0.063* | 0.297 |
| Step 3 | Word flexibility | 0.110** | 0.377 |
| Step 4 | Picture flexibility | 0.005 | −0.094 |
| **(B)** | | | |
| Step 1 | Pseudo-word decoding | 0.232*** | 0.482 |
| Step 2 | Word reading | 0.063* | 0.297 |
| Step 3 | Picture flexibility | 0.018 | 0.144 |
| Step 4 | Word flexibility | 0.097** | 0.436 |

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.*

**Table 3 | Hierarchical regression analyses predicting word reading with (A) word flexibility entered before picture flexibility; and (B) picture flexibility entered before word flexibility.**

| | Variables | Δ $R^2$ | β |
|---|---|---|---|
| **(A)** | | | |
| Step 1 | Pseudo-word decoding | 0.284*** | 0.533 |
| Step 2 | Word flexibility | 0.104** | 0.339 |
| Step 3 | Picture flexibility | 0.000 | 0.029 |
| **(B)** | | | |
| Step 1 | Pseudo-word decoding | 0.284*** | 0.533 |
| Step 2 | Picture flexibility | 0.048* | 0.223 |
| Step 3 | Word flexibility | 0.057* | 0.321 |

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.*

## DISCUSSION

Our exploration of the concurrent relationship between cognitive flexibility and early reading had three main objectives: (1) to investigate whether flexibility is involved in learning to read an orthography that was more transparent than English, namely, the French orthography; (2) to examine the type of reading skills that are associated with flexibility, word reading and/or reading comprehension; and (3) to clarify whether domain-general or domain-specific cognitive flexibility mediates any such relationship with learning to read.

Our results show that reading acquisition in French is related to cognitive abilities that are not exclusively language-based. This extends Cartwright et al.'s (2010) findings from English to the French orthography. In other words, the flexible handling of orthographic, phonological, and semantic codes appears important even when reading a more transparent orthographic system. Word reading skills are acquired more rapidly in French (Seymour et al., 2003), which is thought to reflect a greater reliance on phonological decoding due to the level of consistency in the grapheme-to-phoneme mappings present in the orthography (Ziegler et al., 1996; Sprenger-Charolles et al., 2003). While it will be important to confirm this finding in orthographies with even higher levels of transparency such as Spanish or Finnish, our findings imply that flexibility has a role that extends beyond dealing with the complexities surrounding orthographic depth. This outcome is consistent with the growing number of studies that implicate cognitive abilities in reading acquisition (Conners, 2009; Sesma et al., 2009; Kendeou et al., 2014).

Our use of word reading as a predictor of passage reading comprehension allowed direct assessment of the consequences of the activation of semantic information about words during reading. Interestingly, word reading contributed more than 6% of the variance in passage comprehension beyond that contributed by pseudo-word decoding. This is consistent with Perfetti et al.'s (2005) hypothesis that reading comprehension is engaged by accessing the semantic code of words via visual recognition, and is supported by evidence that word meaning participates in single word reading from the initial phases of acquisition (Nation, 2008; Nation and Cocksey, 2009). Nevertheless, flexibility in coordinating phonological and semantic information made a contribution to the prediction of reading comprehension over and above the individual contribution of basic phonological and semantic processing skills. The critical influence of the simultaneous processing of dimensions is in keeping with the importance that has been placed on the coordination of multiple features of print in reading (Cartwright, 2002; Berninger and Nagy, 2008; Conners, 2009).

A novel and interesting result from our study is that flexibility predicts second grade reading for comprehension not only of texts but also of isolated words beyond the classic influence of decoding skills (e.g., Ouellette and Beers, 2010). A small but significant part of word reading was explained by general (picture) flexibility when it was entered before reading-specific (word) flexibility in the regression analysis. The picture flexibility task involves the coordinated use of phonological and semantic information about referents as does word reading.

However, the reading-specific flexibility task, based on written words, still accounts for additional variance in word reading over and above general (picture) flexibility; whereas the reverse is not true. It was also reading-specific (word) flexibility rather than general (picture) flexibility that predicted passage reading comprehension beyond the influence of pseudo-word decoding and word reading. Together these findings support the interpretation that it is not only phonological-semantic rather than perceptual-semantic flexibility that operates in word and passage reading comprehension (Cartwright et al., 2010), but phonological-semantic flexibility in the specific context of written words.

In the present study, steps were taken to be precise about the nature of the link between cognitive flexibility and reading comprehension, especially in relation to the question of domain-specificity. In future work, it will be important to introduce controls for any non-executive demands that were imposed by the matrix classification tasks used as van der Sluis et al. (2004, 2007) have argued that the effects of any EF can only be fully understood after taking into account the implications of "task impurity."

Indeed, the variety of tasks used to measure cognitive flexibility [see Introduction for a brief overview, and Diamond (2013) for a more thorough review], point to possible reasons for inconsistency in the findings regarding a role for flexibility in the development of reading skills. The task used to measure flexibility is one of the major differences between the present study and the other study of the French language by Monette et al. (2011). Monette et al. (2011) chose to use a card sort task and an adapted version of the Trail-making test, both tasks that require children to make a switch between two sorting criteria. This type of demand differs critically from the flexibility required by matrix classification tasks, such as those used in the present study, which require the simultaneous processing of two dimensions. Therefore, in line with the views of Cartwright (2002, 2007) and Berninger and Nagy (2008), our contention is that this simultaneous maintenance of two perspectives may be a critical component of developing reading skills due to the need to coordinate the multiple types of information contained in print.

Of course, in order to conclude that this task difference is critical, it will be important to rule out the influence of other differences between the two studies. Other differences include the reading measures used. Monette et al. (2011) employed a composite score based on word reading, spelling, and reading comprehension items from the French version of the WIAT-II administered in a group setting, whereas our reading tasks were administered individually and included the standardized Alouette reading test and separate assessments of specific literacy skills, namely, word reading, passage reading comprehension, and decoding. Our intention was to obtain as accurate a picture as possible of the literacy skills that were related to flexibility and to exert control for other more well-known predictors of word reading and reading comprehension such as decoding ability; however, how far this objective was achieved remains to be established empirically.

As cognitive flexibility develops relatively late, our future work will include a longitudinal component to examine the coordination of phonological and semantic information in reading in relation to emerging flexibility at key points throughout preschool and elementary school, which should offer some causal insight into the role of flexibility in reading acquisition.

## CONCLUSION

Overall, these data contribute to the recent and rapidly growing field investigating the role of EF in reading acquisition. Flexibility in coordinating the processing of phonological and semantic information emerged here as a significant correlate of second grade word reading and passage reading comprehension in French. However, cognitive flexibility had greatest power as a predictor of comprehension, over and above traditional linguistic skills, when the matrix classification measures involved the manipulation of written words rather than pictures. Further research is required to explore our conclusion that the predictive value of this type of flexibility is a consequence of the need for an orthographic reading procedure that simultaneously generates phonological and semantic codes for subsequent processing to achieve comprehension.

## ACKNOWLEDGMENTS

## REFERENCES

Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychol.* 8, 71–82. doi: 10.1076/chin.8.2.71.8724

Arlin, P. K. (1981). Piagetian tasks as predictors of reading and math readiness in grades K-l. *J. Educ. Psychol.* 73, 712–721. doi: 10.1037/0022-0663.73.5.712

Berninger, V. W., and Nagy, W. E. (2008). "Flexibility in word reading: Multiple levels of representations, complex mappings, partial similarities and cross-modal connections," in *Literacy Processes: Cognitive Flexibility in Learning and Teaching*, ed. K. B. Cartwright (New York: The Guilford Press).

Best, J. R., and Miller, P. H. (2010). A developmental perspective on executive function. *Child Dev.* 81, 1641–1660. doi: 10.1111/j.1467-8624.2010.01499.x

Bialystok, E., and Martin, M. M. (2004). Attention and inhibition in bilingual children: evidence from the dimensional change card sort task. *Dev. Sci.* 7, 325–339. doi: 10.1111/j.1467-7687.2004.00351.x

Bierman, K. L., Nix, R. L., Greenberg, M. T., Blair, C., and Domitrovich, C. E. (2008). Executive functions and school readiness intervention: impact, moderation, and mediation in the Head Start REDI program. *Dev. Psychopathol.* 20, 821–843. doi: 10.1017/S0954579408000394

Bigler, R. S., and Liben, L. S. (1992). Cognitive mechanisms in children's gender stereotyping: theoretical and educational implications of a cognitive-based intervention. *Child Dev.* 63, 1351–1363. doi: 10.1111/j.1467-8624.1992.tb01700.x

Blair, C., and Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Dev.* 78, 647–663. doi: 10.1111/j.1467-8624.2007.01019.x

Blaye, A., Chevalier, N., and Paour, J. L. (2007). The development of intentional control of categorization behaviour: a study of children's relational flexibility. *Cogn. Brain Behav.* 6, 791–808.

Bull, R., Espy, K. A., and Wiebe, S. (2008). Short-term memory, working memory and executive functioning in preschoolers: longitudinal predictors of mathematical achievement at age 7 years. *Dev. Neuropsychol.* 33, 205–228. doi: 10.1080/87565640

Bull, R., and Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: inhibition, switching, and working memory. *Dev. Neuropsychol.* 19, 273–293. doi: 10.1207/S15326942DN1903_3

Cartwright, K. B. (2002). Cognitive development and reading: the relation of reading-specific multiple classification skill to reading comprehension in elementary school children. *J. Educ. Psychol.* 94, 56–63. doi: 10.1037//0022-0663.94.1.56

Cartwright, K. B. (2007). The contribution of graphophonological-semantic flexibility to reading comprehension in college students: implications for a less simple view of reading. *J. Lit. Res.* 39, 173–193. doi: 10.1080/10862960701331902

Cartwright, K. B. (2012). Insights from cognitive neuroscience: the importance of executive function for early reading development and education. *Early Educ. Dev.* 23, 24–36. doi: 10.1080/10409289.2011.615025

Cartwright, K. B., Marshall, T. R., Dandy, K. L., and Isaac, M. C. (2010). The development of graphophonological-semantic cognitive flexibility and its contribution to reading comprehension in beginning readers. *J. Cogn. Dev.* 11, 61–85. doi: 10.1080/15248370903453584

Chevalier, N., and Blaye, A. (2008). Cognitive flexibility in preschoolers: the role of representation activation and maintenance. *Dev. Sci.* 11, 339–353. doi: 10.1111/j.1467-7687.2008.00679.x

Chevalier, N., and Blaye, A. (2009). Setting goals to switch between tasks: effect of cue transparency on children's cognitive flexibility. *Dev. Psychol.* 45, 782–797. doi: 10.1037/a0015409

Conners, F. A. (2009). Attentional control and the Simple View of reading. *Read. Writ.* 22, 591–613. doi: 10.1007/s11145-008-9126-x

Cragg, L., and Chevalier, N. (2012). The processes underlying flexibility in childhood. *Q. J. Exp. Psychol.* 65, 209–232. doi: 10.1080/17470210903204618

Davidson, M. C., Amso, D., Anderson, L. C., and Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia* 44, 2037–2078. doi: 10.1016/j.neuropsychologia.2006.02.006

Diamond, A. (2013). Executive functions. *Annu. Rev. Psychol.* 64, 135–168. doi: 10.1146/annurev-psych-113011-143750

Espy, K. A., and Cwik, M. F. (2004). The development of a trail making test in young children: The Trails-P. *Clin. Neuropsychol.* 18, 411–422. doi: 10.1080/138540409052416

Farrar, M. J., and Ashwell, S. (2012). Phonological awareness, executive functioning, and theory of mind. *Cogn. Dev.* 27, 77–89. doi: 10.1016/j.cogdev.2011.08.002

Foy, J. G., and Mann, V. A. (2013). Executive function and early reading skills. *Read. Writ.* 26, 453–472. doi: 10.1007/s11145-012-9376-5

Garon, N., Bryson, S. E., and Smith, I. M. (2008). Executive function in preschoolers: a review using an integrative framework. *Psychol. Bull.* 134, 31–60. doi: 10.1037/0033-2909.134.1.31

Gough, P. B., and Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial Spec. Educ.* 7, 6–10. doi: 10.1177/074193258600700104

Harm, M. W., and Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: insights from connectionist models. *Psychol. Rev.* 106, 491–528. doi: 10.1037/0033-295X.106.3.491

Hogan, J. D., and Whitson, G. T. (1984). Perceptual decentering, multiple classification, and reading. *J. Genet. Psychol.* 145, 277–278. doi: 10.1080/00221325.1984.10532274

Jacques, S., and Zelazo, P. D. (2001). The Flexible Item Selection Task (FIST): a measure of executive function in preschoolers. *Dev. Neuropsychol.* 20, 573–591. doi: 10.1207/S15326942DN2003_2

Kendeou, P., van den Broek, P., Helder, A., and Karlsson, J. (2014). A cognitive view of reading comprehension: implications for reading difficulties. *Learn. Disabil. Res. Pract.* 29, 10–16. doi: 10.1111/ldrp.12025

Kieffer, M. J., Vukovic, R. K., and Berry, D. (2013). Roles of Attention shifting and inhibitory control in fourth-grade reading comprehension. *Read. Res. Q.* 48, 333–348. doi: 10.1002/rrq.54

Kirkham, N. Z., Cruess, L., and Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimension-switching task. *Dev. Sci.* 6, 449–467. doi: 10.1111/1467-7687.00300

Kloo, D., and Perner, J. (2003). Training transfer between card sorting and false belief understanding: helping children apply conflicting descriptions. *Child Dev.* 74, 1823–1839. doi: 10.1046/j.1467-8624.2003.00640.x

Kloo, D., Perner, J., Aichhorn, M., and Schmidhuber, N. (2010). Perspective taking and cognitive flexibility in the dimensional change card sorting (DCCS) task. *Cogn. Dev.* 25, 208–217. doi: 10.1016/j.cogdev.2010.06.001

Lefavrais, P. (1967). *Test de l'Alouette.* Paris: Editions du Centre de Psychologie Appliquée.

Lété, B., Sprenger-Charolles, L., and Colé, P. (2004). MANULEX: a grade-level lexical database from French elementary school readers. *Behav. Res. Method Instrum. Comput.* 36, 156–166. doi: 10.3758/BF03195560

Maintenant, C., and Blaye, A. (2008). Développement de la flexibilité catégorielle de 3 à 8 ans: rôle des aspects conceptuels. *Annee Psychol.* 108, 659–698. doi: 10.4074/S0003503308004041

Marcovitch, S., Boseovski, J. J., and Knapp, R. (2007). Use it or lose it: examining preschoolers' difficulty in maintaining and executing a goal. *Dev. Sci.* 10, 559–564. doi: 10.1111/j.1467-7687.2007.00611.x

Miller, P. H. (2010). "Piaget's theory: Past, present, and future," in *Blackwell Handbook of Childhood Cognitive Development*, 2nd Edn, ed. U. Goswami (Malden, MA: Blackwell Publishers).

Miyake, A., and Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: four general conclusions. *Curr. Dir. Psychol. Sci.* 21, 8–14. doi: 10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., and Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. *Cogn. Psychol.* 41, 49–100. doi: 10.1006/cogp.1999.0734

Moll, K., Ramus, F., Bartling, J., Bruder, J., Kunze, S., Neuhoff, N., et al. (2014). Cognitive mechanisms underlying reading and spelling development in five European orthographies. *Learn. Instr.* 29, 65–77. doi: 10.1016/j.learninstruc.2013.09.003

Monette, S., Bigras, M., and Guay, M.-C. (2011). The role of the executive functions in school achievement at the end of Grade 1. *J. Exp. Child Psychol.* 109, 158–173. doi: 10.1016/j.jecp.2011.01.008

Müller, U., Dick, A. S., Gela, K., Overton, W. F., and Zelazo, P. D. (2006). The role of negative priming in preschoolers' flexible rule use on the dimensional change card sort task. *Child Dev.* 77, 395–412. doi: 10.1111/j.1467-8624.2006.00878.x

Müller, U., Zelazo, P. D., and Imrisek, S. (2005). Executive function and children's understanding of false belief: how specific is the relation. *Cogn. Dev.* 20, 173–189. doi: 10.1016/j.cogdev.2004.12.004

Nation, K. (2008). Learning to read words. *Q. J. Exp. Psychol.* 61, 1121–1133. doi: 10.1080/17470210802034603

Nation, K., and Cocksey, J. (2009). Beginning readers activate semantics from subword orthography. *Cognition* 110, 273–278. doi: 10.1016/j.cognition.2008.11.004

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

Peereman, R., Lété, B., and Sprenger-Charolles, L. (2007). Manulex-infra: distributional characteristics of grapheme–phoneme mappings, and infralexical and lexical units in child-directed written material. *Behav. Res. Methods* 39, 579–589. doi: 10.3758/BF03193029

Perfetti, C., Landi, N., and Oakhill, J. (2005). "The acquisition of reading comprehension skill," in *The Science of Reading : A Handbook*, eds M. J. Snowling and C. Hulme (Malden, MA: Blackwell Publishers), 227–247.

Piaget, J., and Inhelder, B. (1958). *The Growth of Logical Thinking: from Childhood to Adolescence.* New York: Basic Books.

Raven, J. C., Court, J. H., and Raven, J. (1995). *Raven Manual: Coloured Progressive Matrices.* Oxford: Oxford Psychologists Press.

Sesma, H. W., Mahone, E. M., Levine, T., Eason, S. H., and Cutting, L. E. (2009). The contribution of executive skills to reading comprehension. *Child Neuropsychol.* 15, 232–246. doi: 10.1080/09297040802220029

Seymour, P. H. K., Aro, M., Erskine, J. M., Wimmer, H., Leybaert, J., Elbro, C., et al. (2003). Foundation literacy acquisition in European orthographies. *Br. J. Psychol.* 94, 143–174. doi: 10.1348/000712603321661859

Share, D. L. (2008). On the anglocentricities of current reading research and practice: the perils of overreliance on an "Outlier" orthography. *Psychol. Bull.* 134, 584–615. doi: 10.1037/0033-2909.134.4.584

Sprenger-Charolles, L., Colé, P., and Serniclaes, W. (2006). *Reading Acquisition and Developmental Dyslexia.* London: Psychology Press.

Sprenger-Charolles, L., Siegel, L. S., Béchennec, D., and Serniclaes, W. (2003). Development of phonological and orthographic processing in reading aloud, in silent reading, and in spelling : a four-year longitudinal study. *J. Exp. Child Psychol.* 84, 194–217. doi: 10.1016/S0022-0965(03)00024-9

St Clair-Thompson, H. L., and Gathercole, S. E. (2006). Executive functions and achievements in school: shifting, updating, inhibition, and working memory. *Q. J. Exp. Psychol.* 59, 745–759. doi: 10.1080/17470210500162854

Tunmer, W. E., and Chapman, J. W. (2013). The simple view of reading redux: vocabulary knowledge and the independent components hypothesis. *J. Learn. Disabil.* 45, 453–466. doi: 10.1177/0022219411432685

Tunmer, W. E., Herriman, M. L., and Nesdale, A. R. (1988). Metalinguistic abilities and beginning reading. *Read. Res. Q.* 23, 134–158. doi: 10.2307/747799

van der Sluis, S., de Jong, P. F., and van der Leij, A. (2004). Inhibition and shifting in children with learning deficits in arithmetic and reading. *J. Exp. Child Psychol.* 87, 239–266. doi: 10.1016/j.jecp.2003.12.002

van der Sluis, S., de Jong, P. F., and van der Leij, A. (2007). Executive functioning in children, and its relations with reasoning, reading, and arithmetic. *Intelligence* 35, 427–449. doi: 10.1016/j.intell.2006.09.001

Yeniad, N., Malda, M., Mesman, J., van IJzendoorn, M. H., and Pieper, S. (2013). Shifting ability predicts math and reading performance in children: a meta-analytical study. *Learn. Individ. Differ.* 23, 1–9. doi: 10.1016/j.lindif.2012.10.004

Zelazo, P. D. (2004). The development of conscious control in childhood. *Trends Cogn. Sci.* 8, 12–17. doi: 10.1016/j.tics.2003.11.001

Zelazo, P. D., Müller, U., Frye, D., and Marcovitch, S. (2003). The development of executive function in early childhood. *Monogr. Soc. Res. Child Dev.* 68, 1–155. doi: 10.1111/j.1540-5834.2003.06803002.x

Ziegler, J. C., and Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychol. Bull.* 131, 3–29. doi: 10.1037/0033-2909.131.1.3

Ziegler, J. C., Jacobs, A. M., and Stone, G. O. (1996). Statistical analysis of the bidirectional inconsistency of spelling and sound in French. *Behav. Res. Methods Instrum. Comput.* 28, 504–515. doi: 10.3758/BF03200539

# Letter knowledge in parent–child conversations: differences between families differing in socio-economic status

**Sarah Robins[1]\*, Dina Ghosh[2], Nicole Rosales[2] and Rebecca Treiman[2]**

[1] Department of Philosophy, University of Kansas, Lawrence, KS, USA
[2] Department of Philosophy, Washington University in St. Louis, St. Louis, MO, USA

When formal literacy instruction begins, around the age of 5 or 6, children from families low in socioeconomic status (SES) tend to be less prepared than children from families of higher SES. The goal of our study is to explore one route through which SES may influence children's early literacy skills: informal conversations about letters. The study builds on previous studies (Robins and Treiman, 2009; Robins et al., 2012, 2014) of parent–child conversations that show how U. S. parents and their young children talk about writing and provide preliminary evidence about similarities and differences in parent–child conversations as a function of SES. Focusing on parents and children aged three to five, we conducted five separate analyses of these conversations, asking whether and how family SES influences the previously established patterns. Although we found talk about letters in both upper and lower SES families, there were differences in the nature of these conversations. The proportion of letter talk utterances that were questions was lower in lower SES families and, of all the letter names that lower SES families talked about, more of them were uttered in isolation rather than in sequences. Lower SES families were especially likely to associate letters with the child's name, and they placed more emphasis on sequences in alphabetic order. We found no SES differences in the factors that influenced use of particular letter names (monograms), but there were SES differences in two-letter sequences (digrams). Focusing on the alphabet and on associations between the child's name and the letters within it may help to interest the child in literacy activities, but they many not be very informative about the relationship between letters and words in general. Understanding the patterns in parent–child conversations about letters is an important first step for exploring their contribution to children's early literacy skills and school readiness.

**Keywords:** home literacy environment, parent-child conversations, socioeconomic status (SES), letter knowledge, Preschool children

## INTRODUCTION

The early years of formal schooling are devoted to teaching children how to read and write. There are differences among children in their preparedness for this instruction, with some children entering school with more knowledge about letters and print than others. In particular, children from families low in socioeconomic status (SES) tend to be less well prepared for literacy instruction and to perform less well in school than children from more privileged backgrounds (Duncan et al., 1998; McLoyd, 1998; Arnold and Doctoroff, 2003; Ryan et al., 2006). The goal of our study is to explore one route through which SES may exert its influence on children's early knowledge about letters and writing: informal conversations about letters that occur at home. We provide a detailed description of the talk about letters that occurs between U.S. parents and their preschool children, asking whether and how patterns in this talk differ as a function of family SES.

Previous studies have found some SES differences in the early home literacy environment of U.S. children. Much of this research focuses on books and book reading, showing that children in lower SES households have less exposure to books in the home (Vernon-Feagans et al., 2001; Roberts et al., 2005) and are less likely to be read to by their parents (Feitelson and Goldstein, 1986; Lee and Burkam, 2002). Even when book reading does occur between lower SES parents and children, there are differences in the quality of parent behavior during this activity (Whitehurst and Lonigan, 1998; Phillips and Lonigan, 2009). However, a range of activities—beyond book reading—contribute to the home literacy environment and could contribute, in turn, to children's ability to benefit from the reading instruction that is provided at school. Indeed, Phillips and Lonigan (2009, p. 147) recommended that measures of the home literacy environment be expanded to include "literacy artifacts, functional uses of literacy, verbal references to literacy, library use, parental encouragement and value of reading, parental teaching of skills, child interest, parent modeling of literacy behaviors, parental education, and parental attitudes toward education." The present study is a response to that request. We select one of those recommended activities—verbal references

to literacy—and examine whether it differs across families with different SES backgrounds.

We focus on parent–child conversations about letters, in part because previous studies have shown that activities that promote children's focus on letters improve their early literacy skills (Sénéchal et al., 1998; Evans et al., 2000; Hood et al., 2008; Martini and Sénéchal, 2012). Conversations about letters, which can occur across a range of everyday activities, may be an important means by which such activities have their influence. Although parents and children may talk about letters while reading books, they may also do so during activities that are not directly focused on literacy, for example while making dinner or doing chores.

We are encouraged in this line of inquiry by studies showing that patterns in parent speech influence children's learning in domains outside of literacy. For example, when *three* is used in reference to apples, days of the week, and toys, it can prompt children to search for how these disparate sets are similar, encouraging them to think about numerical equivalence and thereby improving their mathematical knowledge once they arrive at school (Mix et al., 2002; Levine et al., 2011). A further motivation for the present study is that interventions designed to promote talk about letters appear to improve children's understanding of written language. For example, when parents and teachers are trained to include more explicit references to print during literacy-related activities, like book reading and joint writing, children's overall letter knowledge improves (Lovelace and Stewart, 2007; Justice et al., 2008).

General differences in the conversational patterns of families differing in SES prompt us to consider whether these patterns will influence how parents and their young children talk about letters. Previous studies have shown differences in both the quantity and quality of mothers' talk to children as a function of SES (Hart and Risley, 1995; Hoff and Naigles, 2002). For example, higher SES parents are more likely to talk to their children in ways that elicit and encourage conversation from the child, whereas lower SES parents are more likely to speak to their children in ways that are focused on directing behavior (Farran and Haskins, 1980; Heath, 1983).

For the present study, our interest is in whether there are further differences in parent–child conversations as a function of SES, specifically, differences in the kind of information that is provided in talk about letters. To explore this question, we must first establish the nature of these conversational patterns. Studies that describe the role of parent–child conversations in the home literacy environment tend to employ one of two methods. Some studies use questionnaires, asking parents about the frequency of certain conversational topics, such as rhyming and alphabet games (Phillips and Lonigan, 2009), or about their approach to talking with their child (Umek et al., 2005). The parents in such studies report that they engage their young children in conversations about letters and print (Phillips and Lonigan, 2009). Other studies document patterns in parent–child talk about letters more directly, carrying out case studies with a single family or a small number of families (Neumann et al., 2008; Edwards, 2012). Such studies reveal that parents offer informative statements about letters such as *That's the letter M for MILK. The letter M makes a MMM sound* (Neumann et al., 2008) or *Both words purple*

*and pink begin with P; those are both P words* (Edwards, 2012). Although studies using questionnaire methods tend to have large samples, parents' responses may reflect, in part, the behaviors they think they should engage in. Even if parents' responses are honest, they may not be aware of or remember many of the relevant conversations. Case studies provide more detail about the content of conversations, but only for a restricted set of families. Many case studies, like the ones mentioned above, focus on families of high SES, raising questions as to whether these results generalize to families of preschool children more broadly.

In a series of recent studies (Robins and Treiman, 2009; Robins et al., 2012, 2014), we developed a method for describing parent–child conversations that combines the advantages of each of the above approaches and minimizes their respective limitations. To provide a direct and detailed account of the patterns found in literacy-relevant conversations across a wide range of families with preschool age children, we have examined parent–child conversations available in CHILDES: an online repository of spoken language transcripts (MacWhinney, 2000). CHILDES transcripts—most of which were collected, initially, for studies of children's spoken language development—provide an excellent resource for identifying whether and how parents and children talk during everyday activities. Our studies demonstrate that preschool children and their parents talk about letters and writing and that the ways in which they do so change across the preschool years.

Our interest in the present study is to determine whether there are differences in patterns of talk about letters as a function of SES. CHILDES, again, serves as a resource for asking this question. A number of the researchers who submitted transcripts to CHILDES provided information about the SES of their participants, as determined by parent education and income. For example, one set of transcripts came from the Home-School Study of Literacy and Language Development (Dickinson and Tabors, 2001), which focused exclusively on low-income families. Menn and Gleason (1986), in contrast, recruited their participants from "middle-class families in the Boston area," while other researchers included parent–child conversations from both lower and higher SES families (e.g., Hall et al., 1984). Because the researchers who contributed data to CHILDES differed somewhat in the standards they used for classifying families, our inquiries about SES in CHILDES rely on a general distinction between higher and lower SES that is applicable across corpora. Even using such a basic distinction between demographic groups, there is reason to believe that at least some group differences can be detected. Specifically, a previous exploratory analysis of SES differences found some differences in how parents and children talked about letters and pictures as a function of SES (Robins et al., 2012).

More recently, we have conducted detailed investigations of patterns in talk about letters that occurs between U.S. parents and their preschool age children (Robins et al., 2014). This study showed that parents talked to their young children about letters, questioning them about various features of letters, combining letters into sequences, and associating letters with words. Children, too, talked about letters, displaying in their utterances at least a rudimentary understanding of some aspects of letters. Having identified these patterns in parent–child conversations

about letters, we now ask whether the patterns differ as a function of SES. The present study thus extends the analyses presented in Robins et al. (2014), tailoring them to the subset of corpora for which we have information about participant SES. While other studies have focused on differences in the quantity of literacy-related interactions across SES groups (e.g., Vernon-Feagans et al., 2001; Umek et al., 2005), our approach focuses on the quality of these interactions, asking about the nature of these conversational patterns in lower and higher SES families.

Specifically, the present study documents patterns in parent and child talk about letters when children are between the ages of 3;0 (years; months) and 5;0, as found in transcripts of parent–child conversations that are available in CHILDES (MacWhinney, 2000). Our study consists of five separate analyses, each of which explores one conversational pattern. We selected patterns that were previously identified in our studies of CHILDES (Robins et al., 2014) and that we hypothesize may differ as a function of family SES. Analysis 1 explores questions asked about letters. Analysis 2 examines utterances that mention associations between letters and words, and Analysis 3 looks at utterances that feature letters in sequence. In our final two analyses we look at the letter names used in these conversations in greater detail, asking about the frequency with which individual letters (monograms) and two-letter combinations (digrams) are used and whether patterns in their use are predicted by the SES of the speakers.

## ANALYSIS 1: QUESTIONS

Our first analysis focuses on one highly interactive form of conversation: questions about letters. Previous studies have shown that parents and children sometimes ask questions about letters and print (Yaden et al., 1989), and studies of preschool classrooms have shown that teachers vary the type and complexity of questions posed to students (Massey et al., 2008). Our previous survey of parent–child conversations in CHILDES (Robins et al., 2014) revealed that parents and children asked a number of questions about letters, using these questions to inquire about many features of the letters. Here we ask whether the quality of these questions varies as a function of SES.

### METHODS
#### Utterances for analysis

For this and all subsequent analyses, we used the same 12 corpora of parent–child transcripts that were included in the previous SES analyses described in Robins et al. (2012). The previous study by Robins et al. (2012) used these transcripts to compare talk about writing and about drawing; here our focus is on talk about letters. All 12 corpora included conversations recorded at home between U.S. parents and children, although two of them also included sessions that occurred in a laboratory setting. Families were classified as either lower SES or higher SES, using the demographic data made available by the researchers in CHILDES (MacWhinney, 2000). Given the lag between collecting data and making it available on CHILDES, many of the corpora include conversations that took place before 2000. Date of transcription could not be included in the formal analyses, however, because individual corpora differed in how they reported it (e.g., by date of recording or date of publication). We included all transcripts of conversations

from these corpora that took place between parents and children between the ages of 3;0 and 5;0.

For Analysis 1, we examined all parent and child utterances, defined as a line in the recorded transcript, which included the word *letter* or a specific letter name (as indicated by an @l code in the transcript—e.g., *that's a T@l*). We excluded utterances of *letter* that referred to mailed correspondence. We found utterances that met these criteria in the transcripts from 111 of the 158 parent–child pairs in the 12 corpora. Our searches yielded a total of 3074 utterances—1481 for higher SES families (533 for parents, 948 for children) and 1593 for lower SES families (550 for parents, 1043 for children).

#### Coding

For each statement in the transcript that included the word *letter* or a letter name, we asked whether it was a question. We then distinguished between questions that asked about letter-related skills and those that did not. Non-skill questions included those that mentioned letters while asking about some other topic (e.g., *Do you like your ABC soup?*). Skill questions were in turn coded as either elaborative or basic. Elaborative questions were those that required the respondent to provide letter name or sound, identify a letter shape, complete a sequence, or some combination of these skills. Basic questions required only a yes or no answer. For example, *what is the letter that your name starts with?* was coded as an elaborative question, whereas *Is this an A?* was coded as basic.

For this and the following analyses, a second coder analyzed approximately 5% of the utterances, randomly selected from the full set. Inter-rater agreement was never below 88% for any feature, and the two coders agreed 94% of the time overall. To ensure that this agreement was higher than expected by chance, we calculated the Cohen's κ coefficient for each coding. All κ scores were above 0.75.

#### Statistical analysis

As in Robins et al. (2012) and Robins et al. (2014), analyses were carried out with multilevel models, using the lme4 software package (Bates, 2009). By treating corpus and child as random factors, we were able to determine whether patterns in questions were predicted by variables of interest while statistically controlling for the undue influence of any particular corpus or parent–child pair. We ran two multilevel models for each of the coded features of an utterance, as described above (i.e., whether the utterance as a question, whether the question was a skill question, and whether the skill question was elaborative). The first or non-SES model included the factors of child age (in months) and speaker (child or parent), as well as the interaction of these two factors. The second or SES model included the variables from the first model, as well as SES (higher or lower) and the interaction of SES with the previous variables. Child age was centered in each model. Each model included a random intercept for each child and for each corpus. The two models were statistically compared to determine whether the second accounted for significantly more variance than the first. When the SES model predicted significantly more variance and included a significant main effect or interaction involving SES, we report the results of this model. Otherwise, we report results from the non-SES model.

## RESULTS

The results of our question analyses are displayed in **Table 1**. Although age was treated as a continuous variable in the analyses, the results are broken down into 2 year-long age groups in **Table 1** in order to illustrate the findings. A first statistical analysis was carried out to examine the factors that may help to predict whether an utterance that included a letter name was a question. The SES model performed significantly better than the non-SES model ($p < 0.001$). Overall, 16% (487 of 3074) of all statements including letter names or the word *letter* were questions. The percentage of utterances that included a letter name that were questions was lower for lower SES parents and children than for higher SES parents when children were younger. The percentage of questions increased for lower SES families as children grew older, such that the SES groups showed similar percentages of questions when children were between 4;0 and 5;0. These trends were confirmed by the main effect of SES ($p < 0.001$) and an interaction between age and SES ($p < 0.001$) in the SES model. All other effects were non-significant.

Nearly all questions—95% (463 of 487)—were classified as skill questions. An analysis designed to predict whether a question was a skill question showed no significant effects in either the non-SES or SES models, and the SES model did not perform significantly better than the non-SES model ($p = 0.244$).

There were, however, differences between SES groups in the types of skill questions asked. In analyses designed to predict whether a skill question was an elaborative question, the SES model predicted more variance than the non-SES model ($p < 0.001$). Overall, 36% (168 of 463) of skill questions were elaborative, requiring the respondent to say something beyond yes or no in order to answer the question. The percentage of skill questions that were elaborative was higher in parents, 49%, than in children, 27%. Collapsing across parents and children, the percentage of skill questions that were elaborative was substantially larger in higher SES families, 53%, than in lower SES families, 17%. Also, higher SES families tended to ask more elaborative questions at the older child ages, whereas lower SES families tended to ask fewer. These trends are supported by the SES model, which showed an interaction between child age and SES ($p = 0.005$), as well as main effects of age ($p < 0.001$), speaker ($p = 0.002$), and SES ($p = 0.039$).

## DISCUSSION

The results of Analysis 1 confirm previous observations that parents and children sometimes ask questions about print (Yaden et al., 1989; Robins et al., 2014). In the later preschool years, a number of these questions not only mention letters but ask about the features of letters directly. There are differences in this highly interactive form of conversation as a function of family SES. Of the letter names they uttered, lower SES parents and children had a smaller proportion that were in questions, and the questions that they asked tended to require less detailed responses. The overall difference in proportion of questions is consistent with previous studies that suggest there are SES differences in the kinds of conversations parents have with their young children (e.g., Farran and Haskins, 1980; Heath, 1983). The further discovery that lower SES families have a smaller percentage of elaborative questions than higher SES families do tempers our previous finding (Robins et al., 2014) that, across the preschool years, the questions that parents ask their young children change from simple questions such as *Where is the I?* to more complex ones such as *Dog starts with D—what letter comes next?* While some parents do this, a change toward more elaborate questions may not happen equally for all children. Given studies that stress the importance of questions for promoting children's interest in and understanding of letters and print (Justice et al., 2008; Massey et al., 2008), our results suggest that lower SES children may be at a disadvantage by having fewer of these interactions.

## ANALYSIS 2: ASSOCIATIONS

Even when parents and their preschool children are not asking questions about print, their conversations may still promote young children's knowledge about print if they involve statements about the connections between letters and words. Case studies suggest that parents make such letter–word associations, as when the parent in the Neumann et al. (2008) study said *that's the letter M for MILK,* or when the parent in the Edwards (2012) study said *Both words purple and pink begin with P.* In our previous study using CHILDES (Robins et al., 2014), we found that associations between letters and words were common for both parents and children, but that the types of words used in these associations differed as a function of the child's age. Specifically, parents of younger children focused on associations between letters and proper names, and the majority of these letter–name associations involved the child's name.

A child's own name may serve as an important entry point for directing the child's attention toward letters. Given children's interest in their own names, focusing on this association when children are younger may give them an incentive to learn about the connections between letters and words more broadly (Aram and Levin, 2004; Both-de Vries and Bus, 2010). We asked in Analysis 2 whether there were differences in the relative frequency of associations with the child's name between lower and higher SES families.

## METHODS

### Utterances for analysis

All utterances of individual letter names and all utterances of the word *letter* that came from both parents and the target child were included in this analysis. This yielded a total 6169 utterances of letter names and *letter*—1804 from parents and 4365 from children.

**Table 1 | Proportion of letter statements in analysis 1 that included different types of questions, by SES and child age.**

| Child age | Letter questions as a proportion of all letter utterances | | Elaborative letter questions as a proportion of all letter skill questions | |
|---|---|---|---|---|
| | Lower SES | Higher SES | Lower SES | Higher SES |
| 3;0–4;0 | 0.07 | 0.18 | 0.25 | 0.50 |
| 4;0–5;0 | 0.19 | 0.18 | 0.14 | 0.55 |
| Overall | 0.14 | 0.18 | 0.17 | 0.53 |

### Coding

First, we coded each utterance of *letter* or individual letter name for whether it was associated with a word. To qualify as associated, both the letter and the word to which it referred needed to be explicitly stated in the same line of the transcript. For example, *D is for dog* and *This is the first letter of your name* were coded as associated, but *d-o-g* was not. For all of the utterances that involved associations, we distinguished between those that were associated with the child's name and those that were associated with other words. Name associations included statements like *Your name starts with J* and *J is for Jason*. Not all corpora provided the first names of the children involved in the study, so this coding method may not have identified all associations with names.

### Model

The analyses were carried out on each individual utterance of a letter name. Corpus and child were incorporated into the model as random factors, and non-SES and SES models were compared, as described in Analysis 1.

### Results

A first analysis was carried out to examine the factors that may help to explain whether a letter was associated with a word. There were no influences of SES, as confirmed by comparison of the SES and non-SES models ($p = 0.276$). Parents and children often associated letters with words throughout the age range studied; approximately 3 out of every 10 utterances (1876 of 6169) of letters were associated with a word. Letters were more likely to be associated with words as the child grew older, and parent utterances of letters were more likely to be associated with words than children's. Further, parents' proportion of letter–word associations remained fairly constant across the 3;0–5;0 age range, whereas children's proportion of letter–word associations increased after age 4;0. The model showed main effects of age ($p < 0.001$) and speaker ($p < 0.001$), as well as an interaction between age and speaker ($p < 0.001$).

Of particular interest were the types of words with which the letters were associated. For this analysis of the proportion of associations that were made with the child's name, the SES model predicted more variance than the non-SES model ($p < 0.001$). Associations of a letter with the child's name constituted 22% of all letter-word associations (409 of 1876). There were SES influences on the proportion of associations that involved the child's name, as **Table 2** shows. Specifically, we found a higher proportion of child name associations in lower SES families than in higher SES ones. The focus on associations with the child's name was especially strong for lower SES families at the younger ages. These trends were supported by the SES model, which showed an interaction between child age and SES ($p < 0.001$), and main effects of child age ($p = 0.042$) and SES ($p < 0.001$). Collapsing across SES groups, there was a higher proportion of letter–word associations with the child's name for parents than for children. Parents' relative proportion of these associations decreased across the 3;0–5;0 age range more quickly than children's did, as reflected in the main effect of speaker ($p = 0.002$), which was modified by an interaction between child age and speaker ($p < 0.001$).

**Table 2 | Proportion of child name associations out of all letter–word associations in analysis 2, by SES, speaker, and child age.**

| Child age | Lower SES | | | Higher SES | | |
|---|---|---|---|---|---|---|
| | **Parent** | **Child** | **Overall** | **Parent** | **Child** | **Overall** |
| 3;0–4;0 | 0.66 | 0.52 | 0.59 | 0.36 | 0.65 | 0.47 |
| 4;0–5;0 | 0.23 | 0.35 | 0.32 | 0.05 | 0.46 | 0.38 |
| Overall | 0.54 | 0.41 | 0.46 | 0.21 | 0.48 | 0.40 |

### DISCUSSION

The results of Analysis 2 confirm our previous finding that parents often talk to their young children about letters as being associated with words (Robins et al., 2014). In the present study, more than a third of parent utterances of letters involved associations between letters and words. For both parents and children, many of these associations featured the child's name. The emphasis on the child's name was particularly strong in lower SES families. That is, while the proportion of letter–word associations was similar for lower and higher SES families, lower SES families were particularly likely to make associations with one particular word: the child's name. While this association may draw the child's interest, serving as a critical starting point for making letter–word associations more broadly, persisting with this particular association may not be highly informative. For example, some studies have suggested that children may treat their own names as special, failing to generalize from this association to others (Drouin and Harmon, 2009).

### ANALYSIS 3: SEQUENCES

One important feature of letters is that they come in sequences. Uttering letters in sequence may provide information that letters form a class of symbols. Some sequences of letters, however, are more informative than others. The alphabetic sequence helps children learn the letter names, but the order of this sequence is unrelated to the order with which letters appear in words or to other characteristics of the letters, such as the nature of the sounds that they symbolize. Our previous study (Robins et al., 2014) showed that parents and children often used letters in sequences and that during the later preschool years—which are the focus of the present study—parents and children increasingly focus on the sequences of letters that make up words over those that make up the alphabetic order sequence.

Questionnaires surveying parents about the home literacy environment indicate that many parents consider reciting the alphabet to be an important literacy-related activity and that do this often (e.g., Phillips and Lonigan, 2009). There may, however, be differences among families in the extent and duration of this focus on alphabetic order. In our previous analyses of SES effects in CHILDES (Robins et al., 2012), we found preliminary indications that lower SES families emphasized the alphabetic sequence more than higher SES families. Lower SES families were, for example, more likely to talk about the letters A, B, and C as belonging to children (e.g., asking *do you know your ABCs?*). As the Robins et al. (2012) study focused on comparisons between writing and drawing, the possibility of SES differences in letter

sequences was not explored further in those analyses. In the present analysis, we asked which utterances of letter names were made as part of a sequence, and further, what kinds of sequences were used. We asked whether lower and high SES families differed in these regards.

## METHODS

### Utterances for analysis

This analysis included only uses of individual letter names, leaving out use of the word *letter*. There were yielded 5899 uses of individual letter names—1654 for parents and 4245 for children.

### Coding

Each letter name was first coded for whether it was part of a sequence. A sequence was defined as any instance of two or more symbols in a row, where symbols could be either letters or numbers, separated at most by *and*. For example the letter names in the utterances *2L*, *AB and C*, *XX42J*, and *D-O-G* would all be coded as being in a sequence, whereas the letters in the utterances *I put A on top of B* and *I see two Ds* would not. Then, we asked about the length of each sequence, counting each letter name or number as a token. Then, for all of the letters that were in a sequence, we asked whether the sequence was in alphabetic order. All sequences featuring consecutive letters of the alphabetic order sequence met this criterion, even if they began in the middle of the alphabet (e.g., *lmnop*).

### Model

The analyses of letter sequences were carried out on individual uses of letter names. Corpus and child were incorporated into the model as random factors, as in Analyses 1 and 2.

## RESULTS

In an analysis designed to predict whether a letter occurred in a sequence, the SES model predicted more variance than the non-SES model ($p < 0.001$). Letters were more likely to be uttered in sequence than not—65% (3813 of 5899) of all letter names were said as part of a sequence. The proportion of letter names in sequence for the different groups is shown in **Table 3**. The children had more sequences than parents and, while children's proportion of sequences remained relatively constant across the 3;0–5;0 age range, parents' proportion of sequence utterances increased at the older child ages. Collapsing across parents and children, there were also SES differences in the frequency of sequence utterances. Higher SES families had a higher proportion of letter sequence utterances than lower SES families, and this was especially due to the relatively small proportion of sequences for lower SES families from 3;0–4;0. These results are supported by main effects of child age ($p = 0.025$), speaker ($p < 0.001$), and SES ($p = 0.020$) in the SES model, as well as interactions between child age and speaker ($p < 0.001$) and child age and SES ($p < 0.001$).

Overall, 42% (1212 of 2918) of children's sequences were in alphabetic order, whereas only 27% (239 of 895) of parent sequences were in alphabetic order. There were influences of SES on the proportion of sequences that were in alphabetic order, as reflected in the better performance of the SES model relative to the non-SES model ($p < 0.001$). These results are displayed in

**Table 3 | Proportion of letter names that occurred in sequences in analysis 3, by SES, speaker, and child age.**

| Child age | Lower SES | | | Higher SES | | |
|---|---|---|---|---|---|---|
| | **Parent** | **Child** | **Overall** | **Parent** | **Child** | **Overall** |
| 3;0–4;0 | 0.38 | 0.66 | 0.58 | 0.51 | 0.72 | 0.66 |
| 4;0–5;0 | 0.61 | 0.67 | 0.65 | 0.64 | 0.71 | 0.69 |
| Overall | 0.50 | 0.66 | 0.62 | 0.58 | 0.71 | 0.68 |

**Table 4 | Proportion of letter sequences that are alphabetic order sequences in analysis 3, by SES and child age.**

| Child age | Lower SES | Higher SES |
|---|---|---|
| 3;0–4;0 | 0.48 | 0.56 |
| 4;0–5;0 | 0.36 | 0.20 |
| Overall | 0.41 | 0.35 |

**Table 4**. Lower SES families had a higher proportion of alphabetic order sequences than higher SES families, and this effect was particularly due to the use of such sequences after age 4;0. That is, while higher SES families showed a decline in the proportion of sequences that were in alphabetic order at the older child ages, lower SES families did not. These trends were supported by the SES model, which showed main effects of speaker ($p = 0.004$) and SES ($p < 0.001$), modified by an interaction between child age and SES ($p < 0.001$).

There were no differences across speaker or SES in the length of sequences that were uttered, and the SES model did not perform better than the non-SES model in predicting sequence length ($p = 0.233$). Overall, the average sequence length was 4.43 letters. Sequences tended to be shorter at the older child ages, shrinking from an average length of 4.79 from 3;0–4;0 to 4.20 from 4;0–5;0, as supported by a main effect of age ($p = 0.015$). The shortening of sequences and the lack of a difference between parents and children, while initially surprising, may reflect changes in the type of sequence uttered across the 3;0–5;0 age range. At the younger ages, many of the sequences uttered were alphabetic order sequences, and children had a higher relative proportion of alphabetic order sequences than parents.

## DISCUSSION

Analysis 3 confirms our previous findings (Robins et al., 2014) that U.S. parents and their young children often use letters in sequences and that many of these sequences feature letters in alphabetic order. The present study also extends and refines those results, showing an influence of SES on both letter sequence utterances. While all parents used sequences of letters, lower SES parents uttered a lower proportion of letters in sequences than did higher SES parents and more letters individually. Moreover, of the letter sequences that parents used, lower SES parents had a higher proportion of alphabetic order sequences, especially when their children were older than 4 years of age. Our finding that lower SES families place more emphasis on memorizing the alphabet in order supports and extends the results of previous studies (Baker et al., 1998; Robins et al., 2012).

Learning the alphabetic sequence is enjoyable for young children, particularly when it is done through songs. It may help to draw children's attention to letters, promoting an interest in learning to read and write. But learning how to read and write requires an understanding of how letters combine to form words, and this is not information that can be gleaned from memorizing letter names in alphabetic order. The fact that lower SES children hear many alphabetic sequences, even during the later preschool years, suggests that they may be at an informational disadvantage relative to higher SES children.

## ANALYSIS 4: MONOGRAMS

The previous three analyses establish SES differences in the questions that parents and children ask about letters, the types of associations they make between letters and words, and how they combine letters into sequences. Having established these differences between SES groups, we use the present analysis to step back and ask a more basic question: Are there differences in the individual letters that are used? Using a method developed in our previous study of parent–child letter talk (Robins et al., 2014), we asked whether some individual letter names (monograms) are used more often than others and whether these differences in frequency of use reflect various features of a letter, such as its position in the alphabet and frequency in English words. In our initial study, we found that parents and children often used the letters *A*, *B*, and *C*, but that with older children the frequency of letters used increasingly reflected the frequency with which individual letters occur in English words. We build on that earlier analysis in Analysis 4, asking whether these general patterns in frequency of monogram use differ as a function of SES.

### METHODS
#### Utterances for analysis
This analysis used the same 5899 uses of individual letter names identified in Analysis 3.

#### Coding and analysis procedure
Letter name utterances were pooled into 2 year-long age groups: 3;0–4;0 and 4;0–5;0. We ran separate regression analyses to predict the number of utterances of each letter name by parents and by children. The dependent variable was the frequency of each letter's use, log transformed in order to make the distribution more normal. The predictor variables included child age group, SES, position in the alphabet, and frequency in words. Our position measure, which we label ABC, distinguished the three letters at the beginning of the alphabet—the ones that are often used to label the sequence—from the remaining 23 letters. The frequency variable reflects how often particular letters occur in English words. It was measured here as the number of occurrences of the letter across the 6231 words that appear in Zeno et al. (1995) survey of written materials for kindergarten and first-grade children. Because this variable showed moderate positive skew, we performed a square root transformation. In addition to the variables of child age, ABC, and letter frequency, the analyses included the interactions between child age and ABC and between child age and letter frequency. All continuous variables were centered in the analyses.

## RESULTS
### Parents
There were no SES influences on parent monogram utterances, nor did SES interact with any of the other variables. For parent utterances we found a significant effect of ABC ($p < 0.001$). Parents used the letter names A, B, and C significantly more often than expected on the basis of other factors. Indeed, 24% (394 of 1654) of parents' monogram utterances were one of these three letters. Parents' rate of A, B, and C utterances did not differ across the age range studied, as indicated by the lack of an interaction between the child age and ABC variables. Parent monogram utterances were also predicted by the frequency of these monograms in English words ($p < 0.001$), and this variable also did not interact significantly with child age.

### Children
There was no effect of SES on children's monogram utterances. The results for children's monogram utterances were similar to those for parents: we found significant effects of ABC ($p < 0.001$) and the frequency of the letter in English words ($p < 0.001$). The first three letters of the alphabet constituted 22% (954 of 4.245) of children's monogram utterances throughout the 3;0–5;0 age range. The emphasis on A, B, and C continued across this period, as indicated by the lack of an interaction between child age and ABC. The frequency of the letter in English words was also a significant predictor of children's monogram use, and there were no interactions with child age.

## DISCUSSION
The letters that U.S. parents and children most often talk about are those at the beginning of the alphabet, which are the ones that are often used as a label for the alphabet sequence, and those that frequently occur in English words. These results align with those of our previous study of monogram use (Robins et al., 2014). Given that the present study examines a subset of the corpora used in that initial analysis, the similarity of these findings is to be expected. What is of interest is that we found no differences in the frequency of monogram use as a function of SES. Given that Analyses 1–3 demonstrate several ways in which the patterns of letter talk differ between higher and lower SES families, one might have suspected that these differences would extend to basic letter use as well. While families of different SES backgrounds differ in the questions that they ask about letters, the ways in which they associate letters with words, and the sequences of letters that they use, there were no effects of SES—as a main effect or interaction—for parents or for children in the individual letters that they used.

## ANALYSIS 5: DIGRAMS
Analysis 3 revealed differences in the sequences of letters that parents and children use as a function of SES, but Analysis 4 revealed no influence of SES on the factors that influence use of individual letter names. For our final analysis, we explore the possibility of SES differences in the factors that influence the frequency of letter use at an intermediate level—two-letter combinations, or digrams. Our previous study of parent–child letter use (Robins et al., 2014) revealed that digrams can serve as an informative

unit of analysis. In that study we found that parents and children used some pairs of letters more often than others and, further, that these differences reflected properties of the digrams themselves, above and beyond properties of the individual letters within them. Here we take that inquiry a step further, asking whether there are SES differences in the letters that parents and children combine into basic sequences.

## METHODS
### Coding and analysis procedure
For Analysis 5, we used the set of letter name sequences identified in Analysis 3 and identified each digram in the sequence. For example, if a child said D-O-G, the utterance was coded as involving two digrams, D-O and O-G. The 26 letters of the alphabet can be combined to create 676 distinct digrams, and we kept track of how often each digram occurred in parent and child speech for each year group. The transcripts analyzed contained 2960 digrams—671 from parents and 2289 from children.

Our analysis of the factors that influence parent and child digram use includes 12 factors. First, we used the monogram variables from Analysis 4, applied separately to each letter in the digram (Letter 1, Letter 2): child age, SES, Letter 1 ABC, Letter 2 ABC, Letter 1 frequency, and Letter 2 frequency. We then added the set of digram-level variables used in Robins et al. (2014): digram ABC, digram alphabet, digram frequency, and digram repeat. The digram ABC variable distinguished the digrams involved in the ABC sequence—A-B and B-C—from the remaining 674 digrams. The digram alphabet variable coded each digram for whether the two letters were in alphabetic order, as in A-B, J-K, and X-Y. Digram frequency was calculated using the same set of words from children's books used to analyze monograms in this and the previous analysis. Finally, the digram repeat variable distinguished between digrams that repeated the same letter (e.g., J-J and P-P) and those that did not (e.g., E-F, B-L).

## RESULTS
### Parents
The results of the regression analyses for parent digram utterances are shown in **Table 5**. Parent digram utterances were influenced by a range of factors, many of which echo the findings of Robins et al. (2014). First, parent utterances of two-letter sequences were influenced by several features of the individual letters in those sequences, including the frequency of each letter in English words (Letter 1: $p < 0.05$, Letter 2: $p < 0.01$) and whether the first letter of the digram was A, B, or C ($p < 0.05$). Parents showed a tendency to use digrams that were in alphabetic order ($p < 0.001$), and of the alphabetic order digrams, parents were most likely to utter A-B or B-C ($p < 0.001$). Further, parents' digram use was predicted by repetition ($p < 0.01$) and by the frequency of the digram in English words ($p < 0.001$). There was, in addition, a main effect of age ($p < 0.001$), reflecting the presence of more letters uttered in combination after child age 4.

SES alone did not predict the digram use for parents, but it did interact with some of the other variables. There was an interaction between SES and digram frequency ($p < 0.01$). Higher SES parents were more likely than lower SES parents to utter digrams that

reflected pairs of letters found in English words. There was also an interaction of SES and alphabetic order digrams ($p < 0.001$), but this trend was in the opposite direction: parents from lower SES families were more likely than parents from high SES families to utter alphabetic digrams.

### Children
The results of children's digram utterances are also displayed in **Table 5**. There were several monogram-level influences on children's digram utterances, echoing the patterns identified in Robins et al. (2014)—the frequency of each letter in words of the language (Letter 1: $p < 0.001$, Letter 2: $p < 0.001$) and whether the first letter of the digram was A, B, or C ($p < 0.01$). At the digram level, there were main effects of each variable: frequency ($p < 0.001$), digram ABC ($p < 0.001$), alphabetic order ($p < 0.001$), and repeat ($p < 0.001$). Many of children's digram utterances involved the sequences at the beginning of the alphabet. Children's digram utterances were also significantly predicted by the child's age ($p < 0.001$), with more digram utterances for children after age 4.

There was no main effect of SES on children's digram use. SES did, however, interact with another variable of interest: digram repeat ($p < 0.05$). This interaction qualifies the main effect of repeated digrams identified above. Children were significantly more likely to utter the same letter twice than would have been expected on the basis of other factors, and this was especially true for lower SES children.

## DISCUSSION
The results of this final analysis provide further insight into SES differences in talk about letters. Although there are no differences among parents and children of higher and lower SES backgrounds in the factors that influenced their use of individual letter names, there are SES differences in how they put letters together. Most of the differences we identified came from the parents' use of letters. Although many of the parents we studied used sequences of letters in their conversations with their young children, the kinds of sequences they used differed as a function of SES. Lower SES parents had a higher proportion of alphabetic order sequences. In contrast, higher SES parents showed a stronger tendency to use sequences of letters that that are common in words of the language. This marks an important difference in the input children receive about how letters go together.

We identified only one influence of SES on children's digram use: lower SES children were more likely than their higher SES peers to repeat the same letter. While letters are occasionally doubled in the spelling of words, this tendency to repeatedly say a single letter name may reflect a focus on naming letters rather than combining them to spell words.

## GENERAL DISCUSSION
There are well-established differences in children's preparedness for reading and writing instruction at the beginning of formal schooling as a function of SES. Children from lower SES backgrounds arrive at school with less understanding of letters and how they can be combined to form words, and this gap only widens over the subsequent years (Duncan et al., 1998; Arnold

**Table 5 | Summary of regression analyses for predictors of child and parent digram utterances in analysis 5.**

| Predictor | Child B | Child SE B | Parent B | Parent SE B |
|---|---|---|---|---|
| SES | 0.027 | 0.120 | −0.014 | 0.082 |
| Child age | 0.114*** | 0.024 | 0.065*** | 0.016 |
| × SES | 0.002 | 0.034 | 0.014 | 0.023 |
| **LETTER 1 PREDICTORS** | | | | |
| Letter ABC | 0.105** | 0.038 | 0.067* | 0.026 |
| × SES | −0.007 | 0.054 | 0.026 | 0.037 |
| Frequency | 0.0005*** | 0.00009 | 0.0001* | 0.00006 |
| × SES | −0.00007 | 0.0001 | −0.00004 | 0.00009 |
| **LETTER 2 PREDICTORS** | | | | |
| Letter ABC | 0.066 | 0.038 | 0.044 | 0.026 |
| × SES | −0.070 | 0.054 | −0.009 | 0.037 |
| Frequency | 0.0005*** | 0.0001 | 0.0002** | 0.00007 |
| × SES | 0.00003 | 0.0001 | −0.00005 | 0.00009 |
| **DIGRAM PREDICTORS** | | | | |
| Letter ABC | 1.382*** | 0.235 | 1.215*** | 0.162 |
| × SES | −0.263 | 0.332 | −0.059 | 0.229 |
| Frequency | 0.021*** | 0.004 | 0.016*** | 0.003 |
| × SES | 0.012 | 0.006 | 0.012** | 0.004 |
| Alphabetic order | 1.781*** | 0.066 | 0.832*** | 0.046 |
| × SES | 0.098 | 0.094 | −0.586*** | 0.064 |
| Repeat | 0.701*** | 0.062 | 0.125** | 0.042 |
| × SES | −0.225* | 0.088 | −0.020 | 0.060 |

$N$ = number of utterances. Overall model for children, $F_{(19,2684)}$ = 143, $p < 0.001$, $R^2$ = 0.500. Overall model for parents, $F_{(19,2684)}$ = 57.99, $p < 0.001$, $R^2$ = 0.286.
*$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

and Doctoroff, 2003; Ryan et al., 2006). There is thus a strong interest in understanding the nature of the child's home environment prior to formal schooling and in identifying factors that may contribute to these differences. Previous studies have found general differences in the conversational patterns of parents and children as a function of SES (e.g., Hart and Risley, 1995; Hoff and Naigles, 2002), as well as differences in the quantity and quality of literacy-related activities such as book reading (Vernon-Feagans et al., 2001; Roberts et al., 2005). The present study examined a feature of the home literacy environment at the intersection of these two activities: parent–child conversations about letters.

The present study builds on a prior investigation of parent–child conversations about letters in CHILDES (Robins et al., 2014) which identified five general patterns in these conversations: questions about letters, associations between letters and words, types of letter sequences, as well as the frequency with which individual letters (monograms) and two letter sequences are used. In that previous study, we found that parents and their young children asked questions about letters, made associations between letters and words, and used letters in sequences, although the kinds of questions, associations, and sequences changed across the preschool years. In the current study, we used demographic information about the parent–child pairs—made available by the researchers who submitted their transcripts to CHILDES—to explore whether these previously established

patterns differed as a function of SES. We found some important influences of SES on the previously established patterns. While parents and children in both higher and lower SES groups talk about letters during everyday activities, there are SES differences in the features of these conversations that influence how engaging and informative the interactions are for young children.

One way that parents can engage their young children is by asking them questions. Questions are considered a highly interactive form of conversation (Massey et al., 2008), and we found differences in the prevalence of this form of conversation as a function of family SES. When lower SES parents and children talked about letters, a smaller proportion of those utterances were questions than was the case for higher SES parents and children. Further, of the questions lower SES families asked, a lower proportion required a detailed response. By using a lower proportion of their utterances about letters to query their children about the letters' features, lower SES parents may do less to draw their children's attention to print in environment. This, in turn, may lead children to be less inquisitive about the letters and words that they see printed on such things as toys, signs, and food boxes. Higher SES parents appear to take better advantage of impromptu opportunities to incorporate information about letters into everyday activities.

Parents can, of course, engage their children in other ways, for example by focusing on topics of interest to the child. With

regards to reading and writing, two such topics are the alphabetic sequence and the child's name. Singing the alphabet song and writing or orally spelling the child's name are enjoyable activities that may help to motivate children to attend to print (Aram and Levin, 2004; Both-de Vries and Bus, 2010). Our study shows that both activities occur in the parent–child pairs we examined. Although these activities are valuable, understanding how letters function to produce words requires going beyond these initial activities to discuss sequences other than the alphabet and words other than the child's name. We found SES differences in how parents extend their discussions of letter sequences and associations. Lower SES parents appear to persist in the focus on the alphabetic sequence and simple associations between the child's name and letters of the alphabet for longer than their higher SES counterparts. So, although both high and low SES children receive information about sets of letters and connections between letters and words, the information they can glean from these conversations differs. Higher SES children appear to have more opportunities to learn about how letters can combine to form a range of words. Our findings are consistent with those of previous studies in which lower SES mothers report believing that helping children with basic letter-related skills is important (Fitzgerald et al., 1991; DeBaryshe, 1995). Lower SES parents may be getting the message that it is important to teach their young children about letters, but they may need further guidance on the range of content these interactions should include.

Documenting features of the home literacy environment and how they vary across families is important because it can suggest routes via which we could intervene to improve the literacy outcomes for lower SES children. As ours was a descriptive study, and because it did not follow children as they entered school, we are unable to draw conclusions about the relationship between the conversational patterns we identified and the later literacy achievements of this specific group of children. A further limitation of our approach is that we could only make a brute distinction between higher and lower SES families. Although we found significant differences between these two groups, we encourage further studies that explore finer-grained distinctions between SES groups and seek to disentangle the various demographic factors that contribute to SES classification. Moreover, because recording and transcribing conversations for inclusion in CHILDES takes time, there is the potential for a gap between the patterns observed and current home literacy practices. Nonetheless, we are confident that our study offers important insight into how parent–child conversations can be studied and into the nature of those conversations. By taking advantage of the information available in CHILDES, we were able to examine a much larger sample of conversations, everyday activities, and families than we could have otherwise. Our sample was larger than that of most previous studies of the home literacy environment, and our analyses more detailed. Our study provides an important and previously unavailable baseline for further studies that explore the nature of the home literacy environment in the U. S. and other countries, how it varies across families, and how it influences children's progress when formal literacy instruction begins at school.

## REFERENCES

Aram, D., and Levin, I. (2004). Mother–child joint writing and storybook reading: relations with literacy among low SES kindergartners. *Merrill Palmer Q.* 48, 202–224. doi: 10.1353/mpq.2002.0005

Arnold, D. H., and Doctoroff, G. L. (2003). The early education of socioeconomically disadvantaged children. *Annu. Rev. Psychol.* 54, 517–545. doi: 10.1146/annurev.psych.54.111301.145442

Baker, L., Fernandez-Fein, S., Scher, D., and Williams, H. (1998). "Home experiences related to the development of word recognition," in *Word Recognition in Beginning Literacy*, eds J. Metsala and L. Ehri (Mahwah, NJ: Erlbaum), 263–288.

Bates, D. (2009). *Computational Methods for Mixed Models.* Available online at: http://cran.r-project.org/web/packages/lme4/vignettes/Theory.pdf

Both-de Vries, A. C., and Bus, A. G. (2010). The proper name as starting point for basic reading skills. *Read. Writ.* 23, 173–187. doi: 10.1007/s11145-008-9158-2

DeBaryshe, B. D. (1995). Maternal belief systems: linchpin in the home reading process. *J. Appl. Dev. Psychol.* 16, 1–20. doi: 10.1016/0193-3973(95)90013-6

Dickinson, D. K., and Tabors, P. O. (eds.). (2001). *Beginning Literacy with Language: Young Children Learning at Home and School.* Baltimore, MD: Paul Brookes Publishing.

Drouin, M., and Harmon, J. (2009). Name writing and letter knowledge in preschoolers: incongruities in skills and the usefulness of name writing as a developmental indicator. *Early Childh. Res. Q.* 24, 263–270. doi: 10.1016/j.ecresq.2009.05.001

Duncan, G. J., Yeung, W. J., Brooks-Gunn, J., and Smith, J. R. (1998). How much does childhood poverty affect the life chances of children? *Am. Sociol. Rev.* 63, 406–423. doi: 10.2307/2657556

Edwards, C. M. (2012). Maternal literacy practices and toddlers' emergent literacy skills. *J. Early Childh. Lit.* 14, 53–79. doi: 10.1177/1468798412451590

Evans, M. A., Shaw, D., and Bell, M. (2000). Home literacy activities and their influence on early literacy skills. *Can. J. Exp. Psychol.* 54, 65–75. doi: 10.1037/h0087330

Farran, D. C., and Haskins, R. (1980). Reciprocal influence in social interactions of mothers and three year old children from different socioeconomic backgrounds. *Child Dev.* 51, 780–791. doi: 10.2307/1129465

Feitelson, D., and Goldstein, Z. (1986). Patterns of book ownership and reading to young children in Israeli school-oriented and non-school-oriented families. *Read. Teach.* 39, 924–930.

Fitzgerald, J., Spiegel, D. L., and Cunningham, J. W. (1991). The relationship between parent literacy level and perceptions of emergent literacy. *J. Lit. Res.* 23, 191–213. doi: 10.1080/10862969109547736

Hall, W. S., Nagy, W. E., and Linn, R. (1984). *Spoken Words: Effects of Situation and Social Group on Oral Word Usage and Frequency.* Hillsdale, NJ: Erlbaum.

Hart, B., and Risley, T. R. (1995). *Meaningful Differences in the Everyday Experience of Young American Children.* Baltimore, MD: Paul H Brookes.

Heath, S. B. (1983). *Ways with Words.* Cambridge: Cambridge University Press.

Hoff, E., and Naigles, L. (2002). How children use input to acquire a lexicon. *Child Dev.* 73, 418–433. doi: 10.1111/1467-8624.00415

Hood, M., Conlon, E., and Andrews, G. (2008). Preschool home literacy practices and children's literacy development: a longitudinal analysis. *J. Educ. Psychol.* 100, 252–271. doi: 10.1037/0022-0663.100.2.252

Justice, L. M., Pullen, P. C., and Pence, K. (2008). Influence of verbal and nonverbal references to print on preschoolers' visual attention to print during storybook reading. *Dev. Psychol.* 44, 855–866. doi: 10.1037/0012-1649.44.3.855

Lee, V. E., and Burkam, D. T. (2002). *Inequality at the Starting Gate: Social Background Differences in Achievement as Children Begin School.* Washington, DC: Economic Policy Institute.

Levine, S. C., Suriyakham, L., Rowe, M., and Huttenlocher, J., and Gunderson, E. A. (2011). What counts in the development of children's number knowledge? *Dev. Psychol.* 46, 1309–1313. doi: 10.1037/a0019671

Lovelace, S., and Stewart, S. R. (2007). Increasing print awareness in preschools with language impairment using non-evocative print referencing. *Lang. Speech Hear. Serv. Sch.* 38, 16–30. doi: 10.1177/1468798411416581

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk, 3rd Edn.* Mahwah, NJ: Lawrence Erlbaum Associates.

Martini, F., and Sénéchal, M. (2012). Learning literacy skills at home: parent teaching, expectations, and child interest. *Can. J. Behav. Sci.* 44, 210–221. doi: 10.1037/a0026758

Massey, S., Pence, K. L., Justice, L. M., and Bowles, R. P. (2008). Educators' use of cognitive challenging questions in economically disadvantaged preschool classroom contexts. *Early Educ. Dev.* 19, 340–360. doi: 10.1080/10409280801964119

McLoyd, V. C. (1998). Socioeconomic disadvantage and child development. *Am. Psychol.* 53, 185–204. doi: 10.1037/0003-066X.53.2.185

Menn, L., and Gleason, J. B. (1986). "Baby talk as a stereotype and register: adult reports of children's speech patterns," in *The Fergusonian Impact.* Vol. I, eds J. A. Fishman et al. (Berlin: Mouton de Gruyter), 111–125.

Mix, K. S., Huttenlocher, J., and Levine, S. C. (2002). Multiple cues for quantification in infancy: is number one of them? *Psychol. Bull.* 128, 278–294. doi: 10.1037/0033-2909.128.2.278

Neumann, M. M., Hood, M., and Neumann, D. L. (2008). The scaffolding of emergent literacy skills in the home environment: a case study. *Early Child. Educ. J.* 36, 313–319. doi: 10.1007/s10643-008-0291-y

Phillips, B. M., and Lonigan, C. J. (2009). Variations in the home literacy environment of preschool children: a cluster analytic approach. *Sci. Stud. Read.* 13, 146–174. doi: 10.1080/10888430902769533

Roberts, J., Jurgens, J., Burchinal, M., and Graham, F. (2005). The role of home literacy practices in preschool children's language and emergent literacy skills. *J. Speech Lang. Hear. Res.* 48, 345–359. doi: 10.1044/1092-4388(2005/024)

Robins, S., and Treiman, R. (2009). Talking about writing: what we can learn from conversations between parents and their young children. *Appl. Psychol.* 30, 463–484. doi: 10.1017/S0142716409090237

Robins, S., Treiman, R., and Rosales, N. (2014). Letter knowledge in parent–child conversations. *Read. Writ.* 27, 407–429. doi: 10.1007/s11145-013-9450-7

Robins, S., Treiman, R., Rosales, N., and Otake, S. (2012). Parent–child conversations about letters and pictures. *Read. Writ.* 25, 2039–2059. doi: 10.1007/s11145-011-9344-5

Ryan, R. M., Fauth, R. C., and Brooks-Gunn, J. (2006). "Childhood poverty: implications for school readiness and early childhood education," in *Handbook of Research on the Education of Young Children*, eds B. Spodek and O. N. Saracho (Mahwah, NJ: Erlbaum), 323–346.

Sénéchal, M., LeFevre, J. A., Thomas, E., and Daley, K. (1998). Differential effects of home literary experiences on the development of oral and written language. *Read. Res. Q.* 32, 96–116. doi: 10.1598/RRQ.33.1.5

Umek, L. M., Podlesek, A., and Fekonja, U. (2005). Assessing the home literacy environment: relationships to child language comprehension and expression. *Eur. J. Psychol. Assess.* 21, 271–281. doi: 10.1027/1015-5759.21.4.271

Vernon-Feagans, L., Hammer, C., Miccio, A., and Manlove, E. (2001). "Early literacy in low income and bilingual children," in *Handbook on Research in Early Literacy*, eds S. Neuman and D. Dickinson (New York, NY: Guilford Publications), 192–210.

Whitehurst, G. J., and Lonigan, C. J. (1998). Child development and emergent literacy. *Child Dev.* 69, 848–872. doi: 10.1111/j.1467-8624.1998.00848.x

Yaden, D. B., Smolkin, L. B., and Conlon, A. (1989). Preschoolers' questions about pictures, print conventions, and story text during reading aloud at home. *Read. Res. Q.* 24, 188–214. doi: 10.2307/747864

Zeno, S. M., Ivenz, S. H., Millard, R. T., and Duvvuri, R. (1995). *The Educator's Word Frequency Guide.* Brewster, NY: Touchstone Applied Science Associates.

# The effect of phonics-enhanced Big Book reading on the language and literacy skills of 6-year-old pupils of different reading ability attending lower SES schools

**Laura Tse[1] and Tom Nicholson[2]***

[1] School of Curriculum and Pedagogy, The University of Auckland, Auckland, New Zealand
[2] Institute of Education, Massey University, Auckland, New Zealand

The purpose of this study was to improve the literacy achievement of lower socioeconomic status (SES) children by combining explicit phonics with Big Book reading. Big Book reading is a component of the text-centered (or book reading) approach used in New Zealand schools. It involves the teacher in reading an enlarged book to children and demonstrating how to use semantic, syntactic, and grapho-phonic cues to learn to read. There has been little research, however, to find out whether the effectiveness of Big Book reading is enhanced by adding explicit phonics. In this study, a group of 96 second graders from three lower SES primary schools in New Zealand were taught in 24 small groups of four, tracked into three different reading ability levels. All pupils were randomly assigned to one of four treatment conditions: a control group who received math instruction, Big Book reading enhanced with phonics (BB/EP), Big Book reading on its own, and Phonics on its own. The results showed that the BB/EP group made significantly better progress than the Big Book and Phonics groups in word reading, reading comprehension, spelling, and phonemic awareness. In reading accuracy, the BB/EP and Big Book groups scored similarly. In basic decoding skills the BB/EP and Phonics groups scored similarly. The combined instruction, compared with Big Book reading and phonics, appeared to have no comparative disadvantages and considerable advantages. The present findings could be a model for New Zealand and other countries in their efforts to increase the literacy achievement of disadvantaged pupils.

**Keywords: spelling, phonemic awareness, reading comprehension, Big Book reading, phonics, achievement gap, shared book, math**

## INTRODUCTION

The main reason for this study was to address the literacy needs of lower socioeconomic status (SES) pupils. These students start school with lower levels of pre-reading skills (Nicholson, 1997, 2003; Foster and Miller, 2007; Reardon, 2011, 2013), make slower gains in reading skills in their first years of school (Nicholson, 1997, 2003; Claessens et al., 2009), and make up more of those pupils who receive remedial tuition in Reading Recovery, 18% in lower SES schools as against 11% in higher SES schools (Cowles, 2013). Since this is the case, an important goal is to teach reading more effectively in lower SES schools so that pupils in those schools make more progress than they do at the moment.

The idea behind this study was to find out if the present text-centered or book reading approach used in most classrooms in New Zealand schools could be modified to increase its effectiveness. The text centered approach includes Big Book reading, reading of a wide range of graded readers, shared book reading with the teacher in small groups, as well as oral language and writing activities. One way to enhance the effectiveness of the text-centered approach would be to combine Big Book reading with explicit phonics to find out if this combination would be more effective in raising achievement than Big Book reading or phonics

on its own. Enhancing Big Book reading with explicit phonics and phonemic awareness, both well known for their effectiveness (Gough, 1996; National Reading Panel, 2000; Ezell and Justice, 2005; Tunmer and Nicholson, 2011) could add an additional source of information to classroom instruction that helps disadvantaged pupils learn to read more quickly and increase their reading achievement.

### BIG BOOKS

Big Book reading (Holdaway, 1982) is a technique to enable the teacher to interact with the class so that they pay more attention to text print as well as attending to illustrations and enjoying the story. Big Book reading involves enlarging the size of the reading material so that a whole class can see the print clearly and engage with it not just in terms of meaning but also in terms of looking at printed words and mentally figuring out how letters in words correspond to sounds in speech.

With Big Books (Ministry of Education, 2003) the teacher reads an enlarged copy of a graded reader so that a whole class can see the print clearly and engage with it not just in terms of meaning but also in terms of word reading. When Big Books first started, teachers made their own books, copying the text onto

large pieces of paper but nowadays Big Books are produced commercially. Following the initial reading, pupils may re-read the Big Book with the teacher either that day or during later readings (Ministry of Education, 2003) based on the principle of teachers reading books to the class, then with the class, and finally the class reading the book by themselves.

The teacher reads the same book aloud to the class usually once a day from Monday to Thursday before moving to a new Big Book the following week. To encourage pupils to focus more on the text and less on illustrations, the teacher, while reading to beginner pupils, often follows the line of text with their finger or with a pointer and stops the reading at times to explain language features including unfamiliar vocabulary, punctuation (such as upper case letters or speech marks), or to discuss with the class some decoding aspect of the text, such as a consonant blend.

A feature of Big Book reading is that it does not teach explicit phonics. Pupils learn phonological recoding implicitly and incidentally in the context of reading. The teacher points out letter sound relationships, e.g., *sun* starts with s but phonological recoding is not taught explicitly as in "s-u-n." Instead the teacher usually encourages pupils to use the initial letter or letters of the word plus sentence cues or illustrations to work out the unfamiliar word. Big Book reading therefore does not teach phonics as sounding out words in full as most phonics handbooks suggest (for example, see Nicholson, 2005) but it does encourage use of initial letter sounds and consonant blends (e.g., gr, st, sp) in conjunction with other contextual cues to predict unknown words without focusing on letter-by-letter sounding out. In this way, pupils are given hints as to how to decode words with phonics but are not directly taught to sound out the entire word (Ministry of Education, 2003). The theory is that pupils use the initial letters of the word plus contextual cues and illustrations to work out the meaning of the word but as they continue with reading of Big Books they will infer the phonological rules of decoding especially through acquisition of sub-lexical knowledge through frequent exposure to text. For a review of the acquisition of word reading and implicit phonological recoding in a text-centered way of teaching reading, see Fletcher-Flinn and Thompson (2010) and Thompson (2014).

### VOCABULARY

Big Book reading also seems to improve vocabulary. Students learn new words when listening to stories (Elley and Mangubhai, 1983; Nicholson and Dymock, 2010). They also learn words when reading stories on their own (Suggate et al., 2013). There are individual differences in vocabulary learning from Big Book reading in that there are greater vocabulary gains for those pupils who are from higher socioeconomic (SES) backgrounds (McBride-Chang, 2012; Reese, 2012), or who have higher initial vocabulary knowledge (Robbins and Ehri, 1994), or who are better readers (Nicholson and Whyte, 1992).

### PHONICS AND PHONEMIC AWARENESS INSTRUCTION

The value of phonemic awareness and phonics instruction is well known. The results of meta-analyses indicate that phonemic awareness (Bus and Ijzendoorn, 1999; Ehri et al., 2001b) and phonics are effective especially for lower SES pupils (National

Reading Panel, 2000; Ehri et al., 2001a; Jeynes, 2008; Hattie, 2009; Suggate, 2010; Arrow and Tunmer, 2012).

A theoretical rationale for teaching phonemic awareness and phonics is code-cipher theory. Gough and Hillinger (1980) argued that beginner readers will learn to read if they have: (a) alphabet knowledge, (b) phonemic awareness, (c) cipher intent, that is, where the pupil attempts to recode letters in words according to their phonemes, and (d) data, that is, printed-spoken pairings of words where the pupil sees the word and hears it at the same time. Phonemic awareness and phonics instruction provides a, b, and c and may provide d if the teacher uses text material for pupils to read. Big Book reading definitely provides d but a, b, and c are not taught explicitly so that pupils who lack skill in these areas may not learn to read as quickly as those who have these skills when they start school, skills that higher SES pupils do tend to have more of when they begin school (Nicholson, 2003).

### COMBINED INSTRUCTION

Pressley (2006) has been an influential voice in favor of balanced reading instruction that combines text centered reading instruction (including Big Books) with phonics and phonemic awareness skills. To illustrate the value of balanced instruction, Pressley made an analogy with two different ways of training children to play little league baseball. Learning to read with the book reading approach would be like training for Little League only by playing games. The downside of learning to play baseball by playing games is that if pupils go into games not knowing the skills of how to grip a bat, how to connect with the ball, or what direction to run, then playing games will not make them better players. On the other hand, training for Little League only by practicing batting, fielding, and running will not help unless pupils get a chance to play real baseball games. A little league player will do better with a combined training strategy, that is, by learning skills and then applying them in match practice.

### READING ABILITY

The present study took reading ability into account in that previous researchers have found that the effects of reading programs are different depending on reading ability. Juel and Minden-Cupp (2000) and Connor et al. (2004) found that the impact of the classroom reading program depended on the reading level of the pupil in that pupils with lower levels of decoding skill did better with a phonics emphasis while pupils who had higher levels of decoding skills did better in classrooms that had a text-centered reading focus.

### AIMS

In the present study the benefit of balanced instruction was tested empirically by comparing Big Book reading on its own, phonics on its own, and Big Book reading enhanced with phonics (BB/EP). Pressley (2006) and Pressley and Fingeret (2007) argue that text-centered reading instruction and explicit phonics on their own are not enough and that balanced instruction is more likely to benefit pupils yet there is little research that directly compares a combination of Big Book reading and phonics with Big Book reading and phonics on their own. The present study aimed to fill this gap.

The aims were twofold, first to discover whether enhancing Big Book reading with phonics and phonemic awareness activities leads to measurable improvements in reading, spelling, phonemic awareness, and receptive vocabulary over and beyond that achieved with either Big Book reading or phonics on their own, and second, to measure whether phonics-enhanced Big Book reading achieves greater changes across different levels of reading ability compared with phonics and Big Book reading on their own.

Thus, there were two research questions:

1. Would a group of children who received combined instruction (BB/EP) make more progress than a group who received only Big Book reading, a group who received only explicit phonics, or a control group who received only math instruction?
2. Would the effects of each reading treatment, BB/EP, Big Book (BB) reading, and explicit phonics (P) vary for children with different levels of reading ability?

## MATERIALS AND METHODS

### PARTICIPANTS

Participants were 96 grade 2 (6-year-old) pupils who attended three low- SES primary schools in South Auckland, New Zealand (children in New Zealand start school when they turn five). The schools in the study had a decile 1 rating which is the lowest SES classification (Norris et al., 1994) used by the Ministry of Education in New Zealand. The Ministry uses census data to rank schools on a 1–10 basis (called deciles) based on SES related variables, such as household incomes and occupation of parents. Schools in the lowest categories receive government assistance.

There were 55 boys and 41 girls. Average age at the start of the study was 6 years and 3 months. Nearly all pupils in the study were Maori (42.7%) or from the Pacific Islands (56.3%). English-only was spoken at home by nearly half of the participants (46.9%). Other languages spoken at home in addition to English were Maori (15.6%), Pacific Island languages (36.5%), and for one child, Vietnamese (1%). None of the students received Reading Recovery tuition during the study, which is individual reading tuition available from the government for 6-year-old students not responding to the regular classroom program. All students had already completed a year of formal reading instruction.

### RESEARCH DESIGN AND PROCEDURE

#### Design

The research plan employed a mixed factorial design. The between-subjects factors included two fixed-effect factors, Ability (High, Middle and Low), and Treatment (Combined, Phonics, Big Book, and Math Control). Within each of these combinations were two Teaching Groups of four students each. Teaching Group is a random-effect factor nested within Ability and Treatment, with Students a random-effect factor nested within Teaching Group. The between-subjects design is shown in **Table 1**. Pre-Post was a repeated-measure factor crossed with the between-subjects design.

**Table 1 | Design of the experiment showing the number of children in each subgroup according to conditions and level of reading ability (N = 96).**

| Reading Ability | Teaching subgroups | Conditions | | | | Total |
| --- | --- | --- | --- | --- | --- | --- |
| | | Control (Math) $n = 24$ | Big Book and phonics $n = 24$ | Big Book $n = 24$ | Phonics $n = 24$ | $n = 96$ |
| Higher | Group1 | 4 | 4 | 4 | 4 | 32 |
| | Group2 | 4 | 4 | 4 | 4 | |
| Middle | Group3 | 4 | 4 | 4 | 4 | 32 |
| | Group4 | 4 | 4 | 4 | 4 | |
| Lower | Group5 | 4 | 4 | 4 | 4 | 32 |
| | Group6 | 4 | 4 | 4 | 4 | |
| Total | | 24 | 24 | 24 | 24 | 96 |

#### Procedure

The 96 students were divided into three ability groups based on their scores on the Burt Word Reading Test (Gilmore et al., 1981). Within each ability group, pupils were randomly assigned to four treatment groups: Combined (BB/EP), Big Book only, Phonics only, and Control (this group received alternative instruction in math). Within each treatment-by-ability combination, pupils were divided into two teaching groups.

There was no difference in chronological age among the four treatment groups. Chronological age for each group was: Combined (6.29 years), Phonics (6.28 years), Big Book (6.25 years), and Control (6.31 years), $F_{(3, 92)} = 0.27$, $p > 0.05$.

Burt word reading ages for the three ability groups were: higher (6.46 years); middle (5.75 years); lower (5.29 years). Only students in the higher ability group were reading at their chronological age.

All students completed pretest and posttest assessments of word reading, reading accuracy, reading comprehension, basic word decoding skills, phonemic awareness, receptive vocabulary, word spelling and math computation. One of the authors administered all the assessments. It took 4 weeks to complete the pre-assessments in May and 5 weeks to complete the post-assessments in November. All scoring was cross-checked with another marker until there was 100% agreement.

Teaching interventions ran for 12 weekly sessions, with one 30-min lesson each week taught to each of the 24 subgroups of four students, a total of 24 lessons per week. The tutor always taught the students in small groups of four. As each subgroup consisted of students with either lower, middle, or higher reading ability levels, the phonics lesson plan, the Big Books, and the Math exercises were different for each ability level. All groups received the same amount of time for instruction.

At the end of the study each subgroup had received 12 lessons. There were two school holidays during the training period (a total of 4 weeks) which lengthened the intervention time period.

Within each of the four training groups, there were three different levels of ability for reading or math and each of these

subgroups received a different package of lessons. The Phonics group worked on a different phonics rule each week. The Big Books group worked on four Big Books over the 12 lessons, re-reading each Big Book across three lessons. The first author was the tutor for all lessons. **Figure 1** shows the differences in instruction for the three reading groups.

*Phonics (P).* Students learned and revised letter-sound rules for 25 min (Nicholson, 2005). The lessons followed the sequence of rules of Anglo-Saxon words in English (Calfee and Patrick, 1995)—**Table 2** indicates the scope and sequence of phonics rules covered. Pupils were taught how to analyze printed words according to their sound patterns—for an example of phonics work during the lesson see **Figure 2**. There was no book reading in the lessons. Each lesson also included letter sound training based on the strategy of Turtle Talk (Gough and Lee, 2007). Turtle Talk involves stretching out the sounds in a word to make them more salient, e.g., "s-u-n." The Turtle Talk activity involved the tutor saying the individual sounds in a word slowly, one after the other, with students attempting to guess the word. It was explained to pupils that turtle talk was a way of saying words slowly just as a turtle walks slowly. This activity is called Turtle Talk because the tutor talks slowly at the speed of a turtle, which was the hypothetical explanation given to pupils.

In addition to the oral language form of Turtle Talk the tutor also printed words on a whiteboard and pointed to the letters in the words while the pupils were turtle talking. The tutor was modeling how to decode words according to their letter-sounds. This was not part of the original Turtle Talk activity but was added to the lesson to get a message across to pupils that they can apply Turtle Talk to the decoding of words.

*Big Book (BB).* The students in the Big Book only group read Big Books that were slightly above their instructional reading level—**Table 2** shows the scope and sequence of Big Book lessons. Ten of the Big Books were published by the Ministry of Education and two by a commercial publisher. Big Books are almost 40 cm high and 30 cm wide in dimension, and illustrated with large print.

The tutor used concepts and ideas from *Ready to Read: Teacher support materials* (Ministry of Education, 2001). Each story lasted for three reading sessions. During the lesson, the tutor read the Big Book several times to and with the students. The tutor read the text with the students as choral reading in the first and second reading. In the third reading, the tutor drew students' attention to one or two of the following areas: phonics (e.g., the gr for *greedy* in the *Greedy Cat* story), punctuation (e.g., speech marks, full stops, and capital letters), language features (e.g., opposites - *little* and *big*, *old* and *new*), or asked interactive questions after the reading about the overall meaning of the story including aspects of the text structure such as plot or character. Over the 12 weeks, the tutor read twelve different Big Books, that is, four different Big Books for each ability group.

*Big Books enhanced with explicit phonics (BB/EP).* In the combined group, pupils covered the same Big Books as for the Big Books group and the same phonics and phonemic awareness lessons as for the phonics group but with less depth because it was a shorter time frame to do both sets of activities. **Table 2** shows the scope and sequence of the 12 lessons for the combined group which covered the same phonics rules as the phonics group and the same Big Books as the Big Book group. The Appendix shows an example of a silent e lesson given to the combined Big Book/Phonics reading group where explicit phonics enhanced the Big Book reading (for examples of other BB/EP lessons from this study, see Nicholson and Dymock, 2014).

The tutor started the lesson with a decoding rule and worked on the Big Book that had examples of this rule. The scope and sequence was the same as for the phonics and Big Book lessons but the instruction for each was condensed so as to use both kinds of instruction. As with the phonics lessons, students in the three reading ability groups also engaged in Turtle Talk using words from the story. After the Turtle Talk activity, the tutor wrote on the whiteboard a short list of words that followed decoding rules. The task for students was to associate Turtle Talk phonemes spoken by the tutor with their written representations on the whiteboard. An example of phonics words taken from the Big Book is shown
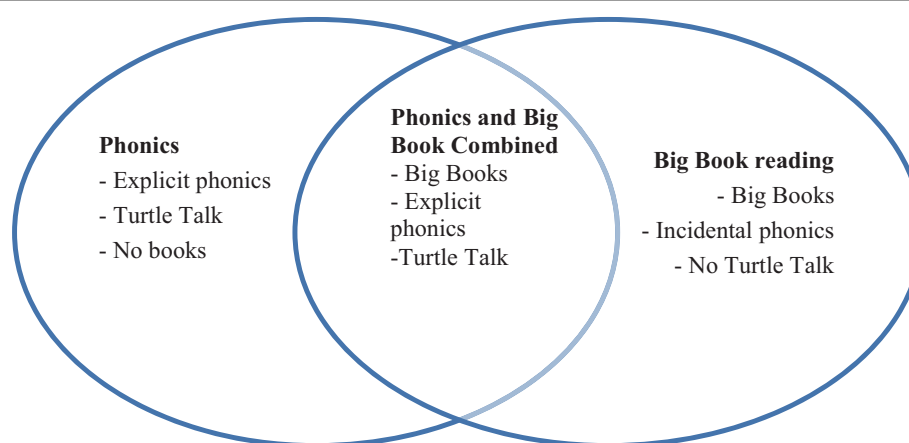


**FIGURE 1 | Diagram of differences among phonics, Big Book, and combined (BB/EP) strategies.**

**Table 2 | The scope and sequence of lesson plans for the combined group (1, first reading; 2, 2nd reading; 3, 3rd reading).**

| Lesson | Lower | Middle | Higher |
|---|---|---|---|
| 1 | Alphabet chart<br>Story: Car shopping<br>(1st reading) | Blends and digraphs<br>Story: Keep trying (1) | Silent e rule (lesson 1)<br>Story: The hole in the King's sock (1) |
| 2 | Consonant blends<br>Story: Car shopping<br>(2nd reading) | Silent e (lesson 1)<br>Story: Keep trying (2) | Silent e rule (lesson 2)<br>Story: The hole in the King's sock (2) |
| 3 | Vowels + silent e (lesson 1)<br>Story: Car shopping (3rd reading) | Silent e (lesson 2)<br>Story: Keep trying (3) | R-affected vowels<br>Story: The hole in the King's sock (3) |
| 4 | Silent e (lesson 2)<br>Story: What does Greedy Cat like? (1) | R- affected vowels<br>Story: Lunch for Greedy Cat (1) | L-affected vowels<br>Story: A good idea (1) |
| 5 | R-affected vowels<br>Story: What does Greedy Cat like? (2) | L-affected vowels<br>Story: Lunch for Greedy Cat (2) | Vowel digraphs<br>ai/ay and oi/oy<br>Story: A good idea (2) |
| 6 | L-affected vowels<br>Story: What does Greedy Cat like? (3) | Vowel digraphs<br>ai/ay and oi/oy<br>Story: Lunch for Greedy Cat (3) | ee<br>Story: A good idea (3) |
| 7 | Vowel digraphs<br>ai/ay and oi/oy<br>Story: Greedy Cat's door (1) | ee<br>Story: Hissing Bush (1) | ie<br>Story: Earthquake (1) |
| 8 | ee<br>Story: Greedy Cat's door (2) | ie<br>Story: Hissing Bush (2) | oa and ew<br>Story: Earthquake (2) |
| 9 | ie<br>Story: Greedy Cat's door (3) | oa and ew<br>Story: Hissing Bush (3) | au and aw<br>Story: Earthquake (3) |
| 10 | oa and ew<br>Story: Keep Trying (1) | au and aw<br>Story: Magnetic Max (1) | ea<br>Story: Firefighter (1) |
| 11 | au and aw<br>Story: Keep Trying (2) | ea<br>Story: Magnetic Max (2) | oo and ou<br>Story: Firefighter (2) |
| 12 | ea<br>Story: Keep Trying (3) | oo and ou<br>Story: Magnetic Max (3) | Syllable breaking (CVC/CVC)<br>Story: Firefighter (3) |

in **Figure 3**. The tutor wrote the words *her, after, purr, lunch, gave, home, came, and still* on the whiteboard. As in the phonics lessons, when doing the phonemic awareness activity the tutor asked students to listen carefully when she slowly said the sounds in the word, e.g., "keh-ay-m" (for *came*), to blend the sounds together in their minds, then to say the word aloud, and point to the correct answer on the whiteboard. Students also performed this activity in reverse (e.g., what word is "m-ay-keh").

**Control group (M).** Students in the control group received Math instruction and the same amount of instructional time as the other treatment groups. This condition controlled for placebo effects, that is, the effects of receiving special attention. Students learned about numbers and the quantities they stand for, specifically, counting, comparing numbers, addition, subtraction, and multiplication. All students practiced and computed math questions at different ability levels. For example, the lower ability

reading group learned basic one-digit addition, the middle ability reading group learnt one- and two-digit addition, and the higher ability reading group learned at a more advanced level for addition.

### Weekly quizzes

The purpose of having quizzes was to assess learning of phonics rules for the Phonics and Combined groups—see **Table 3** for the scope and sequence of quizzes and **Figure 4** for an example of a quiz. The quizzes were given to all four groups each week, at the end of each lesson, except for the first lesson. Each quiz had five questions. The paper-and-pencil quiz took 5 min to complete and tested different decoding patterns, for example, the silent *e* rule, consonant blends/digraphs, and vowel digraphs. The quizzes covered phonics rules taught in the BB/EP and Phonics group lessons with different quizzes for each reading ability group. The lower group were assessed on single letter sounds, consonant blends
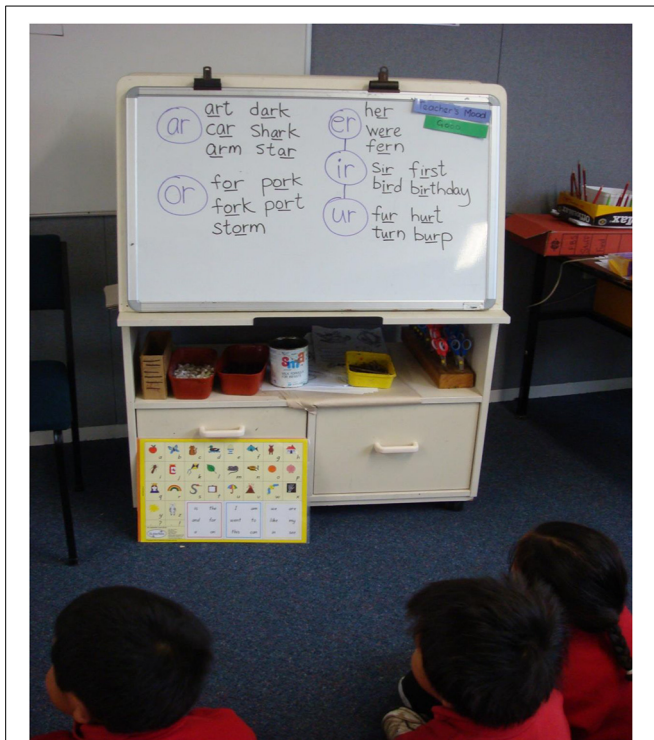
**FIGURE 2 | A segment from a phonics lesson with word patterns written on the whiteboard to illustrate the sounds of r-affected vowels (ar, er, ir, or, ur).**



**FIGURE 3 | A segment from a Big Book/Phonics combined lesson (BB/EP) with word patterns from the Big Book written on the whiteboard to illustrate the sounds of r-affected and l-affected vowels, and the silent e pattern.**

and digraphs, short vowel sounds as in *hop*, the split digraph rule (silent e) as in *hope*, r- and l-affected vowel sounds as in *car*, *wall*, and single-sound vowel digraphs such as ai, ay as in *rain* and *ray*. The middle group was assessed on similar rules but with an additional two-sound vowel digraph tested, ea as in *beach* and *bread*. The higher group was assessed on similar patterns to those of the lower and middle groups but with the addition of two-sound digraphs oo as in *book* and *roof* and ou as in *soup* and *mouse*.

### Measures

**Word reading.** The Burt Word Reading Test (Gilmore et al., 1981) is a norm-referenced test standardized in New Zealand which assesses the ability to read words out of context. Students read words presented on a test card with 110 words printed in different sizes of type and graded in approximate order of difficulty from easy words like *to* and *big* to difficult words like *ingratiating* and *poignancy*. In this test, students read as many words as they can and stop when they make 10 consecutive errors (or miscues). They then look over the remaining words to see if they recognize any other words. The test manual reports high test- retest reliability ($r > 0.95$) and high internal consistency ($r > 0.96$). The reason for using this test is that it is the only New Zealand norm-referenced word reading test standardized for use with 6-year-old pupils. The test-retest correlation in the current study was $r = 0.86, N = 96$.

**Passage reading.** The Neale Analysis of Reading-3rd Edition (Neale, 1999) is a norm-referenced test for pupils aged 6 to 12

years which has two parallel forms. The test assesses passage oral reading accuracy, ability to comprehend passages, and rate of reading. We did not assess rate of reading in this study mainly because children in the lower groups at pretest were reading hardly any words. Pupils completed the green form (Form 2) in the pretest and the yellow form (Form 1) in the posttest. Each form consisted of six passages graded in difficulty. The pupil reads the passages aloud and then answers comprehension questions asked by the examiner. Students cannot look back at the story when answering comprehension questions. The test has a high level of internal consistency with correlations ranging from 0.71–0.96. We chose this test because it is the only norm-referenced measure for 6-year-olds that assesses reading accuracy and comprehension of passages with norms for a similar population to New Zealand (the test was standardized in Australia). The test-retest correlations for this measure were $r = 0.88, N = 96$ for accuracy and $r = 0.67, N = 96$ for comprehension.

**Basic decoding skills.** The Bryant Test of Basic Decoding Skills (Bryant, 1975; reprinted in Nicholson, 2005) is a list of 50 pseudowords read aloud by the student. The test starts with one-syllable consonant-vowel-consonant (CVC) combinations such as *buf*, then moves to silent-e patterns such as *fute*, consonant digraphs such as *thade*, vowel digraphs such as *groy*, and ends with multisyllabic pseudowords such as *vomazful*. Pupils had to pronounce the word correctly as a whole word, not just

**Table 3 | Scope and sequence of the 10 quizzes for ability groups.**

| Quiz | Lower | Middle | Higher |
|---|---|---|---|
| 1 | Single sounds* | Consonant blends and digraphs* | Short and long vowels/silent e [split digraphs][a]* |
| 2 | Consonant blends and digraphs | Short and long vowels/silent e [split digraphs][a] | Short and long vowels/silent e [split digraphs][b] |
| 3 | Short and long vowels/silent e [split digraphs][a] | Short and long vowels/silent e [split digraphs][b] | r-affected vowels |
| 4 | Short and long vowels/silent e [split digraphs][b] | r-affected vowels | l-affected vowels |
| 5 | r-affected vowels | l-affected vowels | ai-ay and oi-oy |
| 6 | l-affected vowels | ai-ay and oi-oy | ee and ie** |
| 7 | ai-ay and oi-oy | ee and ie** | oa and ew |
| 8 | ee and ie ** | oa and ew | au and aw |
| 9 | oa and ew | au and aw | ea |
| 10 | au and aw | ea | oo and ou |

*Quiz 1 started in Week 2 of the intervention.

**Lessons on ee and ie combined in one quiz.

[a,b] The silent e pattern is called a split digraph in England.
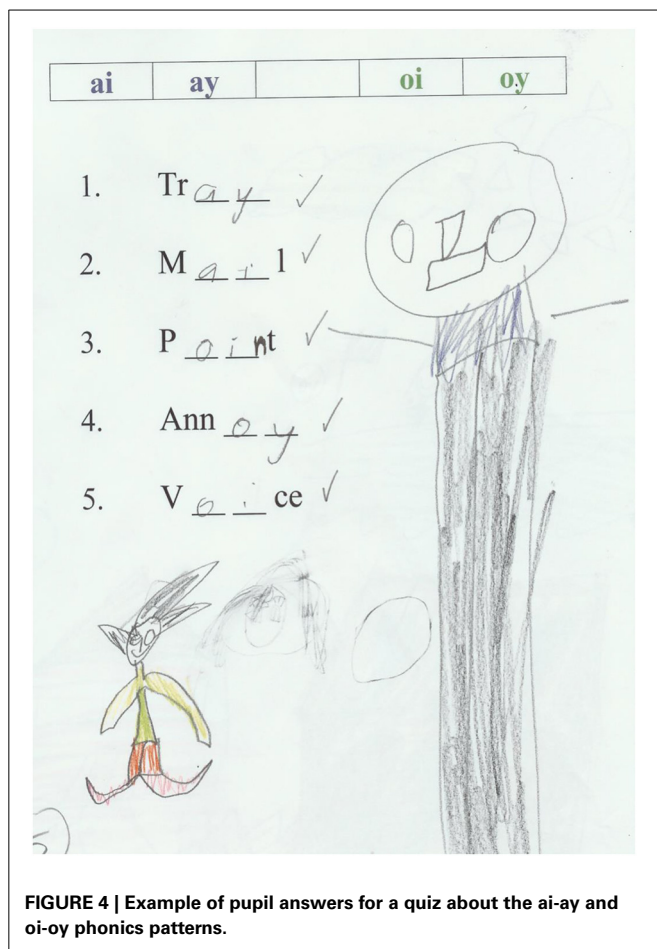


**FIGURE 4 | Example of pupil answers for a quiz about the ai-ay and oi-oy phonics patterns.**

sounding out each letter. When students made 10 consecutive errors, testing stopped and students were encouraged to look at the rest of the list to check if they could read any other words. Juel (1988) reported reliabilities between 0.90 and 0.96 for this test. This test is not norm-referenced. We chose this test because it assessed basic decoding skills and because its scope and sequence of difficulty matched with phonics rules taught in the

study (e.g., the pseudoword *fute* targeted knowledge of the silent e rule). The test-retest correlation in this study was $r = 0.72$, $N = 96$.

*Phonemic awareness.* The Gough-Kastler-Roper (GKR) Test of Phonemic Awareness (Roper, 1984; reprinted in Nicholson, 2005) has 42 items divided into six categories of seven items each assessing a different aspect of phonemic awareness: phonemic segmentation, blending, deletion of initial and final phonemes, and initial and final phoneme substitution. This is an oral assessment measure where students do not see the items. The assessor reads out the questions and the students respond to them verbally (e.g., what are the two sounds in "up"?). The assessor stops after 10 consecutive errors. Roper (1984) reported reliabilities greater than $r = 0.7$ for all subtests of this measure. This test is not norm-referenced. We chose this test because it has been successfully used in other New Zealand studies and it has a range of difficulty. The test-retest correlation in this study was $r = 0.77$, $N = 96$.

*Receptive vocabulary.* The British Picture Vocabulary Scale (BPVS II) (Dunn et al., 1997) is a norm-referenced receptive vocabulary assessment. For example, one of the test pages has four pictures: butterfly, baby, bed and shoe. The pupil points to the picture that represents the word spoken by the examiner (e.g., "bed"). There are 168 target words. The median reliability according to the examiner manual is 0.90. The reason for choosing this measure is that it is suitable for the age group and we wanted to know if the Big Book reading experience had a positive effect on vocabulary learning. The test-retest reliability in this study was $r = 0.67$, $N = 96$.

*Spelling.* The Schonell Spelling Test (Schonell, 1951) is a series of words graded in difficulty. The assessor says the word, says it in a sentence, and says the word again. The pupil then spells the word. The test starts with three-letter words (e.g., net, can, fun) and extends to multi-syllabic words (e.g., irresistible, hydraulic, anniversary). Stevenson et al. (1993) reported high reliability, $r = 0.97$ for the test. The test was suitable for use with young pupils

**Table 4 | Descriptive statistics showing pretest, posttest, and prepost differences for higher, middle, and lower ability pupils in the control, combined, Big Book, and phonics training groups (with minimum and maximum scores indicated for each measure) (N = 96).**

| Reading level | | Math control | | | Combined | | | Big Book | | | Phonics | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Pre- | Post- | Diff | Pre- | Post- | Diff | Pre- | Post- | Diff | Pre- | Post- | Diff |
| **WORD READING (BURT)** | | | | | | | | | | | | | |
| **Min-Max (0–110)** | | | | | | | | | | | | | |
| Lower | M | 3.4 | 14.6 | **11.3** | 2.5 | 19.4 | **16.9** | 4.1 | 11.6 | **7.5** | 4.0 | 11.5 | **7.5** |
| | SD | 4.2 | 6.3 | 5.7 | 2.2 | 5.7 | 5.4 | 3.4 | 5.0 | 2.4 | 3.5 | 8.1 | 6.4 |
| Middle | M | 12.0 | 20.6 | **8.6** | 11.6 | 26.9 | **15.3** | 13.8 | 25.0 | **11.3** | 10.5 | 16.9 | **6.4** |
| | SD | 4.7 | 5.8 | 3.8 | 5.4 | 5.8 | 3.5 | 6.6 | 4.9 | 5.6 | 5.2 | 5.8 | 6.3 |
| Higher | M | 23.3 | 31.4 | **8.1** | 22.8 | 36.8 | **14.0** | 24.6 | 36.6 | **12.0** | 27.3 | 39.3 | **12.0** |
| | SD | 8.3 | 9.7 | 5.7 | 7.2 | 6.8 | 2.6 | 6.3 | 7.7 | 5.3 | 3.7 | 8.6 | 8.2 |
| Across level | M | 12.9 | 22.2 | **9.3** | 12.3 | 27.7 | **15.4** | 14.2 | 24.4 | **10.3** | 13.9 | 22.5 | **8.6** |
| | SD | 10.1 | 10.1 | 5.1 | 9.9 | 9.3 | 4.0 | 10.1 | 11.9 | 4.9 | 10.8 | 14.3 | 7.2 |
| **PASSAGE ACCURACY (NEALE)** | | | | | | | | | | | | | |
| **Min-Max (0–100)** | | | | | | | | | | | | | |
| Lower | M | 0.6 | 6.1 | **5.5** | 0.1 | 5.3 | **5.1** | 0.1 | 3.3 | **3.1** | 0.6 | 4.6 | **4.0** |
| | SD | 1.8 | 4.8 | 3.9 | 0.4 | 5.4 | 5.2 | 0.4 | 3.2 | 3.2 | 1.2 | 5.6 | 4.9 |
| Middle | M | 4.1 | 12.4 | **8.3** | 3.1 | 14.6 | **11.5** | 3.9 | 16.1 | **12.3** | 1.9 | 7.5 | **5.6** |
| | SD | 4.1 | 7.7 | 4.9 | 2.9 | 6.0 | 4.4 | 4.2 | 3.9 | 3.3 | 1.9 | 4.5 | 3.7 |
| Higher | M | 12.1 | 23.9 | **11.8** | 12.9 | 28.6 | **15.8** | 14.5 | 28.9 | **14.4** | 16.5 | 29.0 | **12.5** |
| | SD | 9.2 | 13.4 | 4.6 | 10.1 | 10.1 | 2.8 | 8.2 | 8.9 | 5.5 | 9.5 | 9.5 | 9.4 |
| Across level | M | 5.6 | 14.1 | **8.5** | 5.4 | 16.2 | **10.8** | 6.2 | 16.1 | **9.9** | 6.3 | 13.7 | **7.4** |
| | SD | 7.5 | 11.7 | 5.0 | 8.0 | 12.1 | 6.0 | 8.0 | 12.1 | 6.4 | 9.1 | 12.9 | 7.2 |
| **PASSAGE COMPREHENSION (NEALE)** | | | | | | | | | | | | | |
| **Min-Max (0–36)** | | | | | | | | | | | | | |
| Lower | M | 0.6 | 1.3 | **0.6** | 0.3 | 2.3 | **2.0** | 0.4 | 1.8 | **1.4** | 0.1 | 1.6 | **1.5** |
| | SD | 1.1 | 1.4 | 1.2 | 0.5 | 1.2 | 1.1 | 1.1 | 1.8 | 2.0 | 0.4 | 1.2 | 1.3 |
| Middle | M | 2.0 | 2.5 | **0.5** | 1.5 | 4.4 | **2.9** | 1.4 | 3.0 | **1.6** | 2.4 | 2.3 | **−0.1** |
| | SD | 1.2 | 2.0 | 1.4 | 1.4 | 2.1 | 2.4 | 0.9 | 1.5 | 1.4 | 1.4 | 1.2 | 1.0 |
| Higher | M | 3.5 | 5.0 | **1.5** | 3.5 | 8.4 | **4.9** | 4.1 | 6.1 | **2.0** | 3.9 | 7.1 | **3.3** |
| | SD | 2.5 | 3.9 | 3.0 | 1.8 | 2.1 | 1.6 | 2.0 | 2.2 | 2.5 | 2.9 | 3.7 | 3.2 |
| Across level | M | 2.0 | 2.9 | **0.9** | 1.8 | 5.0 | **3.3** | 2.0 | 3.6 | **1.7** | 2.1 | 3.7 | **1.5** |
| | SD | 2.0 | 3.0 | 2.0 | 1.9 | 3.1 | 2.1 | 2.1 | 2.6 | 1.9 | 2.4 | 3.4 | 2.4 |
| **PHONEMIC AWARENESS (ROPER)** | | | | | | | | | | | | | |
| **Min-Max (0–42)** | | | | | | | | | | | | | |
| Lower | M | 1.4 | 3.4 | **2.0** | 0.1 | 10.4 | **10.3** | 2.3 | 6.5 | **4.3** | 0.0 | 2.9 | **2.9** |
| | SD | 2.9 | 4.3 | 2.2 | 0.4 | 5.4 | 5.4 | 4.5 | 7.9 | 7.4 | 0.0 | 2.5 | 2.5 |
| Middle | M | 1.5 | 11.0 | **9.5** | 5.4 | 22.1 | **16.8** | 1.5 | 6.1 | **4.6** | 3.1 | 12.1 | **9.0** |
| | SD | 2.0 | 11.2 | 10.4 | 8.0 | 11.2 | 8.7 | 2.8 | 3.8 | 4.1 | 3.6 | 9.4 | 6.1 |
| Higher | M | 15.8 | 21.8 | **6.0** | 19.6 | 32.1 | **12.5** | 22.0 | 27.1 | **5.1** | 17.3 | 29.8 | **12.5** |
| | SD | 14.1 | 10.4 | 9.6 | 9.0 | 5.1 | 6.2 | 9.2 | 9.3 | 10.5 | 15.4 | 8.6 | 9.1 |
| Across level | M | 6.2 | 12.0 | **5.8** | 8.4 | 21.5 | **13.2** | 8.6 | 13.3 | **4.7** | 6.8 | 14.9 | **8.1** |
| | SD | 10.6 | 11.7 | 8.5 | 10.7 | 11.7 | 7.1 | 11.3 | 12.2 | 7.5 | 11.6 | 13.4 | 7.4 |
| **BASIC DECODING SKILLS (BRYANT)** | | | | | | | | | | | | | |
| **Min-Max (0–50)** | | | | | | | | | | | | | |
| Lower | M | 0.5 | 1.9 | **1.4** | 0.0 | 5.8 | **5.8** | 0.1 | 0.1 | **0.0** | 0.0 | 1.8 | **1.8** |
| | SD | 1.4 | 3.2 | 3.2 | 0.0 | 7.5 | 7.5 | 0.4 | 0.4 | 0.5 | 0.0 | 3.3 | 3.3 |
| Middle | M | 0.0 | 3.1 | **3.1** | 0.1 | 12.9 | **12.8** | 0.1 | 1.0 | **0.9** | 0.4 | 4.4 | **4.0** |
| | SD | 0.0 | 3.2 | 3.2 | 0.4 | 6.2 | 6.0 | 0.4 | 2.1 | 2.1 | 0.7 | 5.3 | 5.3 |
| Higher | M | 9.4 | 16.0 | **6.6** | 6.9 | 19.5 | **12.6** | 10.1 | 13.3 | **3.1** | 9.5 | 22.3 | **12.8** |
| | SD | 12.9 | 15.4 | 5.7 | 6.9 | 4.9 | 3.6 | 7.4 | 7.6 | 10.7 | 9.8 | 11.1 | 7.7 |
| Across level | M | 3.3 | 7.0 | **3.7** | 2.3 | 12.7 | **10.4** | 3.5 | 4.8 | **1.3** | 3.3 | 9.5 | **6.2** |
| | SD | 8.4 | 11.0 | 4.6 | 5.0 | 8.3 | 6.6 | 6.3 | 7.5 | 6.2 | 7.0 | 11.7 | 7.3 |

*(Continued)*

**Table 4 | Continued**

| Reading level | | Math control | | | Combined | | | Big Book | | | Phonics | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Pre- | Post- | Diff | Pre- | Post- | Diff | Pre- | Post- | Diff | Pre- | Post- | Diff |
| **SPELLING (SCHONELL)** | | | | | | | | | | | | | |
| **Min-Max (0–100)** | | | | | | | | | | | | | |
| Lower | M | 1.9 | 3.1 | **1.3** | 0.1 | 5.3 | **5.1** | 0.6 | 1.6 | **1.0** | 0.1 | 3.3 | **3.1** |
| | SD | 5.3 | 5.2 | 2.9 | 0.4 | 5.5 | 5.5 | 1.8 | 2.6 | 1.4 | 0.4 | 4.4 | 4.3 |
| Middle | M | 2.1 | 8.9 | **6.8** | 3.0 | 12.1 | **9.1** | 0.9 | 7.0 | **6.1** | 4.0 | 4.4 | **0.4** |
| | SD | 2.5 | 7.2 | 5.7 | 4.7 | 4.1 | 5.9 | 2.1 | 4.7 | 4.0 | 3.5 | 4.6 | 3.6 |
| Higher | M | 14.3 | 20.9 | **6.6** | 12.5 | 23.5 | **11.0** | 13.1 | 24.6 | **11.5** | 12.9 | 22.5 | **9.6** |
| | SD | 12.6 | 13.7 | 5.8 | 8.4 | 7.5 | 3.8 | 7.1 | 8.6 | 7.1 | 6.5 | 5.6 | 5.8 |
| Across level | M | 6.1 | 11.0 | **4.9** | 5.2 | 13.6 | **8.4** | 4.9 | 11.1 | **6.2** | 5.7 | 10.0 | **4.4** |
| | SD | 9.7 | 11.7 | 5.4 | 7.6 | 9.5 | 5.5 | 7.3 | 11.5 | 6.3 | 6.8 | 10.2 | 6.0 |
| **BRITISH PEABODY VOCABULARY TEST (BPVT)** | | | | | | | | | | | | | |
| **Min- Max (0–168)** | | | | | | | | | | | | | |
| Lower | M | 44.5 | 52.3 | **7.8** | 41.0 | 49.6 | **8.6** | 45.4 | 50.3 | **4.9** | 42.5 | 51.3 | **8.8** |
| | SD | 11.4 | 14.6 | 9.6 | 3.7 | 7.2 | 6.8 | 10.9 | 6.8 | 12.2 | 6.3 | 8.0 | 9.0 |
| Middle | M | 55.1 | 59.1 | **4.0** | 50.1 | 56.0 | **5.9** | 45.1 | 52.9 | **7.8** | 48.4 | 53.3 | **4.9** |
| | SD | 5.9 | 9.6 | 6.9 | 12.2 | 12.2 | 9.8 | 10.2 | 9.0 | 7.7 | 9.6 | 8.7 | 7.0 |
| Higher | M | 49.3 | 52.8 | **3.5** | 52.1 | 60.8 | **8.6** | 50.9 | 50.9 | **0.0** | 53.5 | 62.4 | **8.9** |
| | SD | 14.2 | 10.3 | 8.6 | 10.6 | 5.3 | 12.1 | 5.9 | 7.8 | 9.4 | 6.6 | 11.0 | 8.2 |
| Across level | M | 49.6 | 54.7 | **5.1** | 47.8 | 55.5 | **7.7** | 47.1 | 51.3 | **4.2** | 48.1 | 55.6 | **7.5** |
| | SD | 11.5 | 11.6 | 8.3 | 10.4 | 9.6 | 9.5 | 9.3 | 7.7 | 10.0 | 8.6 | 10.2 | 8.0 |
| **MATHEMATICS (WRAT 3)** | | | | | | | | | | | | | |
| **Min-Max (0–55)** | | | | | | | | | | | | | |
| Lower | M | 10.3 | 15.6 | **5.4** | 11.5 | 13.8 | **2.3** | 10.3 | 12.8 | **2.5** | 9.5 | 11.0 | **1.5** |
| | SD | 3.3 | 4.1 | 3.2 | 2.6 | 2.4 | 2.1 | 1.5 | 2.2 | 2.3 | 2.1 | 1.7 | 2.7 |
| Middle | M | 12.8 | 20.6 | **7.9** | 12.8 | 15.4 | **2.6** | 11.1 | 12.3 | **1.1** | 12.6 | 13.8 | **1.1** |
| | SD | 1.8 | 2.1 | 3.6 | 1.5 | 2.2 | 2.6 | 2.2 | 3.5 | 3.6 | 2.0 | 2.8 | 3.2 |
| Higher | M | 14.3 | 20.8 | **6.5** | 12.5 | 16.9 | **4.4** | 14.4 | 16.6 | **2.3** | 14.4 | 17.4 | **3.0** |
| | SD | 2.7 | 2.3 | 1.8 | 2.4 | 1.6 | 1.8 | 2.6 | 4.5 | 2.5 | 2.2 | 1.6 | 2.2 |
| Across level | M | 12.4 | 19.0 | **6.6** | 12.3 | 15.3 | **3.1** | 11.9 | 13.9 | **2.0** | 12.2 | 14.0 | **1.9** |
| | SD | 3.0 | 3.8 | 3.0 | 2.2 | 2.4 | 2.3 | 2.7 | 3.9 | 2.8 | 2.9 | 3.3 | 2.8 |

in that the words slowly increase in difficulty. The test-retest correlation in this study was $r = 0.84$, $N = 96$.

*Math.* The WRAT 3 Wide Range Achievement Test (Wilkinson, 1993) is a norm-referenced test of math computation. The test divided into 2 parts. Part 1 was given orally with 15 questions involving counting, identifying numbers, and solving simple oral problems, such as "Read these numbers out loud" and "Which number is more, 9 or 6?" Part 2 was a pencil and paper test with 40 math problems, with questions suitable for this age group, such as $2 + 1 =, 5 - 3 =, 4 \times 2 =$. Students answered as many questions as they could in 15 min. Raw score is the number of questions answered correctly in parts 1 and 2 of the test. The test manual reported reliabilities from 0.87–0.96. We chose this test because it started with very simple calculations and it did not involve reading. The test-retest correlation in this study was $r = 0.56$, $N = 96$.

## DATA ANALYSIS

The pre-post battery and the quizzes were both analyzed by standard factorial ANOVA techniques, augmented by orthogonal contrasts to assess specific questions for the Ability and Treatment factors. For Ability, orthogonal polynomials were used to evaluate the linear and quadratic trends across the three levels. For Treatment, Helmert contrasts (Keppel and Wickens, 2004), also orthogonal, served to answer the following questions from the research problem:

1. Does performance of the Math Control group (C) differ from the average of the other treatment groups?
2. Does performance of the Combined group (BB/EP) differ from the average of the two single-treatment groups?
3. Do the two single-treatment groups, Big Book (BB) and Phonics (P) differ from one another?

The analyses of all measures were based on $N = 96$ except for the spelling and basic decoding skills measures where for each measure one of the children did not complete the assessment as intended. For these measures the analyses were based on $N = 95$.

Effect sizes were measured using the partial omega square statistic ($\omega^2$) which suited the contrast analyses. Keppel and Wickens (2004) recommend this statistic as most suitable for

orthogonal contrasts. Omega square statistics report the amount of variance accounted for by the contrast. A small effect captures about 1% of the variance, a medium effect about 6% of the variance and a large effect about 15% of the variance. Only the omega square statistics for each contrast are reported in **Table 5** since these are the most important effects for this study.

## RESULTS

The pretest, posttest, and difference mean scores, and standard deviations for the eight dependent measures are shown in **Table 4**. The statistical analyses are shown in **Table 5**. The prepost difference raw scores for the four treatment groups are shown in **Figure 5** to make comparisons clearer. The difference scores for ability are presented in **Figures 6, 7** as percent scores in order to show trend differences with a common metric. The percent score was the difference score divided by the maximum score for each measure. We report the findings for the treatment groups first.

### GROUP RESULTS FOR PRETEST AND FOR PRE-POST DIFFERENCE SCORES

At pretest there were no significant group effects not as main effect or as a contrast. This showed that the treatment groups were equivalent at pretest. The results for the pre-post difference scores showed a different pattern altogether. In presenting the pre-post difference results we focus on the questions relating to the contrasts since they were most important in terms of the analysis.

Question 1. Does performance of the Math Control group differ from the average of the other treatment groups (C vs. BB/EP, BB, P)? As can be seen in **Table 5** and **Figure 5**, the contrast between the control group and the other groups (Math/Other) for the language and literacy measures were sometimes not significant mainly due to the control group scoring more highly than the phonics and Big Book groups so that the average of the three groups was similar to the control group. The exceptions were two significant Math/Other effects for reading comprehension and basic decoding skills where the math group scored significantly below the average of the other treatment groups.

The control group (Math/Other) contrast was highly significant for the math measure and with a substantial effect size showing that the control group performed much better than the average of the three reading groups. This was because the control group received alternative math instruction and the other groups did not.

Question 2. Does the performance of the Combined group (BB/EP) differ from the average of the Big Book (BB) and Phonics (P) groups? As shown in **Table 5** and **Figure 5**, the BB/EP group had significantly higher scores than the average mean score of the BB and P groups for word reading, reading comprehension, basic decoding skills, phonemic awareness and spelling. Two of the effect sizes were substantial (word reading and basic decoding skills). For reading accuracy, the BB/EP group was not significantly different to the average mean of the BB and P groups though it was nearly so [$p = 0.053$: BB/EP mean(diff) = 10.8, BB mean(diff) = 9.6, P mean(diff) = 7.4]. There was no significant effect for the contrast of the BB/EP group and the other two groups in relation to the vocabulary and math measures.

**Table 5 | Results of Three-Way ANOVAs for pretest and prepost difference data for each measure with polynomial contrasts for ability and helmert contrasts for group, using a random effects general linear model and partial omega square effect sizes.**

| Variables | df | Pretest F | Pretest $\omega^2$ | Prepost difference F | Prepost difference $\omega^2$ |
|---|---|---|---|---|---|
| **BURT WORD READING** | | | | | |
| Ability | 2 | 117.50** | | 0.41 | |
| Linear | 1 | 232.26** | **0.78** | 0.34 | 0.00 |
| Quadratic | 1 | 2.86 | 0.03 | 0.49 | 0.00 |
| Group | 3 | 0.62 | | 8.37** | |
| Math/other | 1 | 0.20 | 0.00 | 2.92 | 0.04 |
| Combined/BB,P | 1 | 1.61 | 0.00 | 20.98** | **0.29** |
| BB vs. phonics | 1 | 0.03 | 0.00 | 1.19 | 0.00 |
| Ability × group | 6 | 0.54 | | 1.81 | |
| In group team | 12 | 0.61 | | 1.42 | |
| *MS* error | 72 | (30.31) | | (26.84) | |
| **NEALE ACCURACY** | | | | | |
| Ability | 2 | 47.34** | | 35.36** | |
| Linear | 1 | 85.19** | **0.57** | 70.56** | **0.52** |
| Quadratic | 1 | 9.49** | **0.12** | 0.17 | 0.00 |
| Group | 3 | 0.14 | | 2.89* | |
| Math/other | 1 | 0.06 | 0.00 | 0.71 | 0.00 |
| Combined/BB,P | 1 | 0.35 | 0.00 | 3.88# | **0.06** |
| BB vs. phonics | 1 | 0.01 | 0.00 | 4.08* | **0.06** |
| Ability × group | 6 | 0.49 | | 1.47 | |
| In group team | 12 | 0.63 | | 2.98** | |
| *MS* error | 72 | (34.87) | | (19.01) | |
| **NEALE COMPREHENSION** | | | | | |
| Ability | 2 | 33.67** | | 7.15** | |
| Linear | 1 | 67.08** | **0.51** | 9.61** | **0.12** |
| Quadratic | 1 | 0.42 | 0.00 | 4.67* | **0.05** |
| Group | 3 | 0.22 | | 6.25** | |
| Math/other | 1 | 0.06 | 0.00 | 7.56** | **0.12** |
| Combined/BB,P | 1 | 0.49 | 0.00 | 11.16** | **0.17** |
| BB vs. phonics | 1 | 0.12 | 0.00 | 0.05 | 0.00 |
| Ability × group | 6 | 0.38 | | 1.33 | |
| In group team | 12 | 0.48 | | 1.11 | |
| *MS* error | 72 | (2.77) | | (3.89) | |
| **PHONEMIC AWARENESS** | | | | | |
| Ability | 2 | 45.89** | | 3.97* | |
| Linear | 1 | 76.21** | **0.54** | 4.67* | **0.05** |
| Quadratic | 1 | 15.52** | **0.18** | 3.28 | 0.03 |
| Group | 3 | 0.50 | | 5.67** | |
| Math/other | 1 | 0.79 | 0.00 | 2.37 | 0.03 |
| Combined/BB,P | 1 | 0.12 | 0.00 | 12.25** | **0.19** |
| BB vs. phonics | 1 | 0.58 | 0.00 | 0.05 | 0.03 |
| Ability × group | 6 | 0.48 | | 0.85 | |
| In group team | 12 | 0.37 | | 0.37 | |
| *MS* error | 72 | (65.87) | | (59.99) | |
| **SPELLING** | | | | | |
| Ability | 2 | 40.17** | | 19.83* | |
| Linear | 1 | 69.06** | **0.51** | 39.44* | **0.38** |
| Quadratic | 1 | 11.22** | **0.14** | 0.18 | 0.00 |

*(Continued)*

**Table 5 | Continued**

| Variables | df | Pretest | | Prepost difference | |
|---|---|---|---|---|---|
| | | F | ω² | F | ω² |
| Group | 3 | 0.19 | | 3.60** | |
| Math/other | 1 | 0.34 | 0.00 | 2.04 | 0.02 |
| Combined/BB,P | 1 | 0.00 | 0.00 | 7.18** | **0.11** |
| BB vs. phonics | 1 | 0.22 | 0.00 | 1.56 | 0.01 |
| Ability × group | 6 | 0.24 | | 2.09 | |
| In group team | 12 | 0.47 | | 2.18* | |
| MS error | 71 | (36.21) | | (20.21) | |
| **BASIC TEST OF DECODING SKILLS (NON-WORDS)** | | | | | |
| Ability | 2 | 22.22** | | 20.89** | |
| Linear | 1 | 33.52** | **0.34** | 41.34** | **0.39** |
| Quadratic | 1 | 11.36** | **0.14** | 0.50 | 0.00 |
| Group | 3 | 0.15 | | 14.38** | |
| Math/other | 1 | 0.26 | 0.00 | 5.90* | **0.09** |
| Combined/BB,P | 1 | 0.36 | 0.00 | 29.05** | **0.37** |
| BB vs. phonics | 1 | 0.02 | 0.00 | 8.29** | **0.13** |
| Ability × group | 6 | 0.11 | | 1.81 | |
| In group team | 12 | 0.17 | | 2.57** | |
| MS error | 71 | (33.91) | | (20.37) | |
| **BRITISH PEABODY VOCABULARY TEST** | | | | | |
| Ability | 2 | 6.31** | | 0.52 | |
| Linear | 1 | 11.42** | **0.14** | 0.90 | 0.00 |
| Quadratic | 1 | 1.23 | 0.00 | 0.14 | 0.00 |
| Group | 3 | 0.30 | | 0.82 | |
| Math/other | 1 | 0.76 | 0.00 | 0.38 | 0.00 |
| Combined/BB,P | 1 | 0.00 | 0.00 | 0.62 | 0.00 |
| BB vs. phonics | 1 | 0.13 | 0.00 | 1.46 | 0.01 |
| Ability × group | 6 | 0.92 | | 0.67 | |
| In group team | 12 | 0.83 | | 0.52 | |
| MS error | 72 | (91.96) | | (89.13) | |
| **MATHEMATICS** | | | | | |
| Ability | 2 | 19.76** | | 1.52 | |
| Linear | 1 | 39.44** | **0.38** | 2.79 | 0.03 |
| Quadratic | 1 | 0.15 | 0.00 | 0.23 | 0.00 |
| Group | 3 | 0.21 | | 16.18** | |
| Math/Other | 1 | 0.34 | 0.00 | 45.43** | **0.48** |
| Combined/BB,P | 1 | 0.14 | 0.00 | 2.99 | 0.04 |
| BB vs. phonics | 1 | 0.15 | 0.00 | 0.01 | 0.00 |
| Ability × group | 6 | 1.63 | | 1.11 | |
| In group team | 12 | 1.36 | | 1.11 | |
| MS error | 72 | (4.98) | | (7.24) | |

*p < 0.05; **p < 0.01; #p = 0.053.
Math, control group; Combined, Big Book enhanced with phonics; Other, Combined, Big Book and Phonics; BB, Big Book; P, Phonics; In Group Team, small groups.

Question 3. Did the two single-treatment groups differ from one another? As shown in **Table 5** and **Figure 5**, the final contrast between the BB and P groups showed a mixed picture for reading accuracy and decoding. For reading accuracy the BB group performed better than P and had similar scores to the BB/EP group. Thus, for reading accuracy we can infer that the BB/EP

and BB groups made similar progress. For decoding the P group performed better than the BB group and had similar scores to the BB/EP group. Thus, we can infer that for basic decoding skills the BB/EP and P groups made similar progress. On all other measures (word reading, reading comprehension, phonemic awareness, spelling, vocabulary, and math) there was no difference between the BB and P groups.

To summarize the pre-post results for the treatment groups, the Combined BB/EP instruction was more effective than Big Book reading for all literacy measures except reading accuracy where there was no difference between the Combined and Big Book groups. Combined instruction was more effective than phonics for all literacy measures except basic decoding skills where it was equally effective. The control group who received math instruction made significantly more progress in math than the other three groups who did not receive math teaching. In **Figure 8** the results for word reading, reading accuracy, and reading comprehension are expressed as reading ages and spelling as a spelling age to give a more meaningful interpretation of the results. These graphs show that for reading comprehension, word reading, and spelling, the BB/EP instruction brought the reading and spelling ages of these children closer to their chronological age. For reading accuracy, BB/EP and BB instruction both moved children closer to their chronological age.

## ABILITY RESULTS AT PRETEST AND FOR PRE-POST DIFFERENCE SCORES

### Pretest

A trend analysis of pretest scores for word reading, reading comprehension, receptive vocabulary and math showed that the linear coefficient made a significant contribution in explaining the trend (effect sizes were from 0.14 to 0.78) but the quadratic coefficient did not (effect sizes 0.00 to 0.03). As can be seen in **Figure 6**, mean percent score (percent of maximum possible score) decreased similarly in line with reading ability (word reading: high = 22%, middle = 11%, low = 3%; reading comprehension: high = 10%, middle = 5%, low = 1%; vocabulary: high = 31%, middle = 30%, low = 26%; math: high = 25%, middle = 22%, low = 19%).

A trend analysis of pretest scores for reading accuracy, phonemic awareness, spelling, and basic decoding skills, showed that the linear coefficient (effect sizes were from 0.34 to 0.57) and quadratic coefficient (effect sizes were from 0.12 to 0.18) both made a significant contribution in describing the trend of the data, though the linear trend accounted for most of the variance. Although students' scores did decrease in a linear way from the higher group to the middle group, this pattern did not continue for the lower group. As can be seen in **Figure 6**, the middle and lower groups had similar percent scores that were well below those of the higher group (reading accuracy: high = 14%, middle = 3%, low = 1%; phonemic awareness: high = 44%, middle = 7%, low = 2%; spelling: high = 13%, middle = 2%, low = 1%; decoding skills: high = 17%, middle = 1%, low = 1%).

### Posttest

A trend analysis of prepost gain scores for reading accuracy, phonemic awareness, spelling, and basic decoding skills showed that the linear coefficient made a significant contribution to
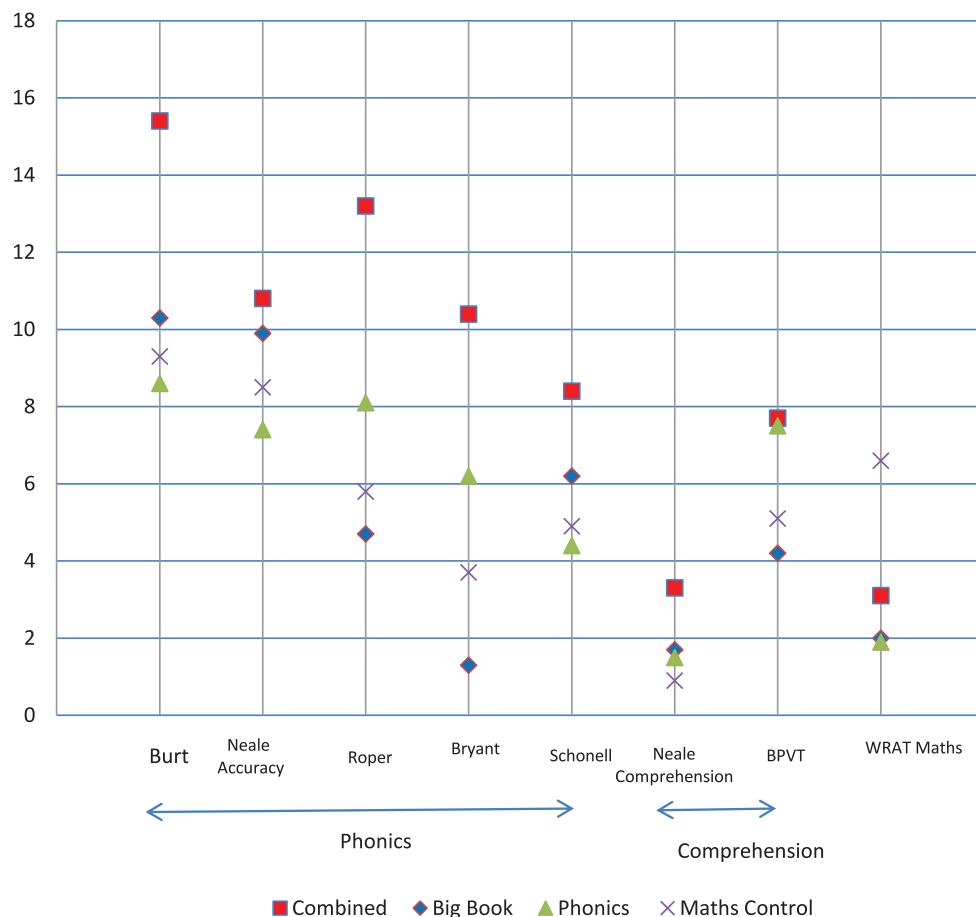
**FIGURE 5 | The pre-post difference raw scores for the four treatment groups for each measure.**

explaining the trend (effect sizes were from 0.05 to 0.52) and that the quadratic coefficient did not (effect sizes were from 0.00 to 0.03). As can be seen in **Figure 7**, mean percent gains decreased similarly in line with reading ability (reading accuracy: high = 14%, middle = 9%, low = 4%; phonemic awareness: high = 21%, middle = 24%, low = 11%; spelling: high = 10%, middle = 6%, low = 3%; decoding: high = 19%, middle = 10%, low = 4%).

A trend analysis of gains for reading comprehension showed that the linear coefficient (effect size was 0.12) and the quadratic coefficient (effect size was 0.05) both made a significant contribution to explaining the trend. As can be seen in **Figure 7**, there was a linear decrease in prepost comprehension gain from the higher to middle group but this pattern did not continue for the lower group whose percent gain was similar to that of the middle group (reading comprehension: high = 8%, middle = 3%, low = 4%).

A trend analysis of prepost difference scores for word reading, receptive vocabulary, and math showed no significant linear or quadratic trends. As can be seen in **Figure 7**, the three ability groups made similar percent gains for these measures (word reading: high = 10%, middle = 9%, low = 10%; receptive vocabulary: high = 3%, middle = 3%, low = 4%; math: high = 4%, middle = 6%, low = 5%).

## INTERACTIONS

Pretest scores showed no significant ability × group interactions, indicating that the treatment groups were equivalent in ability at pretest. Prepost difference scores (gains) showed no significant ability × group interactions, indicating that the three ability groups made similar gains across the four treatment groups.

## IN GROUP TEAM EFFECTS

There were no significant in-group team effects at pretest, indicating that the subgroup teams were equivalent. For pre-post measures there were significant in-group team effects for reading accuracy, spelling and basic decoding skills, indicating some differences among the subgroups. These were random effects, however, and not the focus of this design.

## PHONICS QUIZZES

All groups completed the 10 weekly phonics quizzes. Each quiz had five questions and was marked out of 5. The marks for the 10 different quizzes were averaged to be out of 5 (see **Table 6** for means and standard deviations). Each ability group did different quizzes. The analysis was the same ANOVA design as for the test battery except that it was not possible to include ability as a fixed
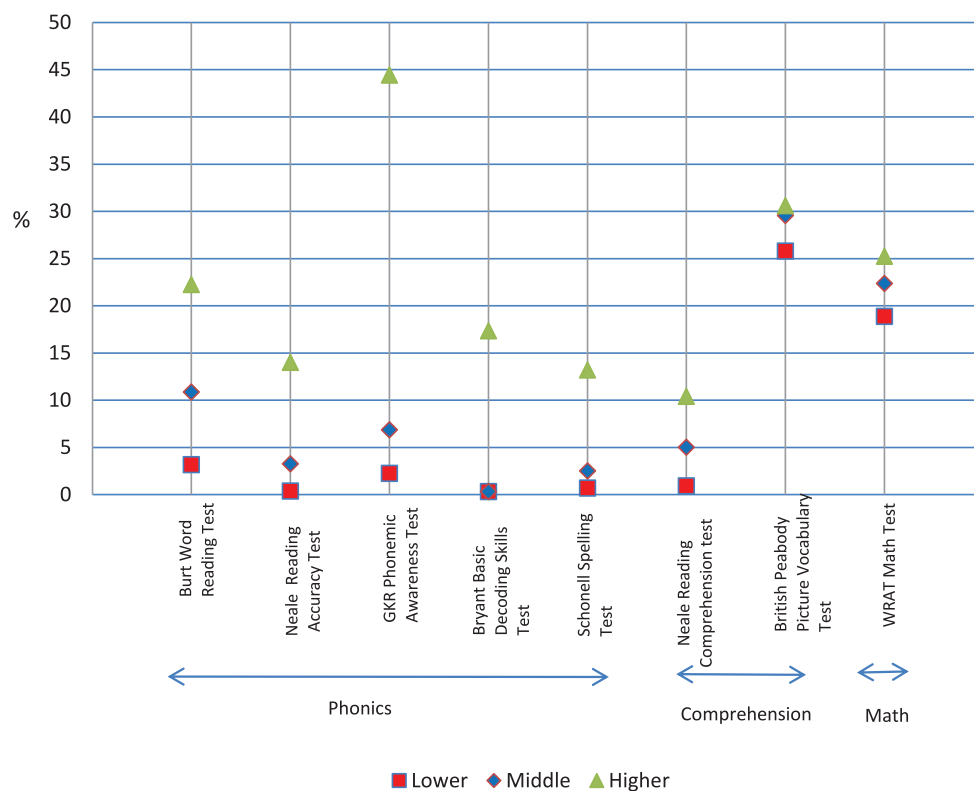
**FIGURE 6 | The pretest scores for the three ability groups expressed as percentage of maximum score for each measure.**
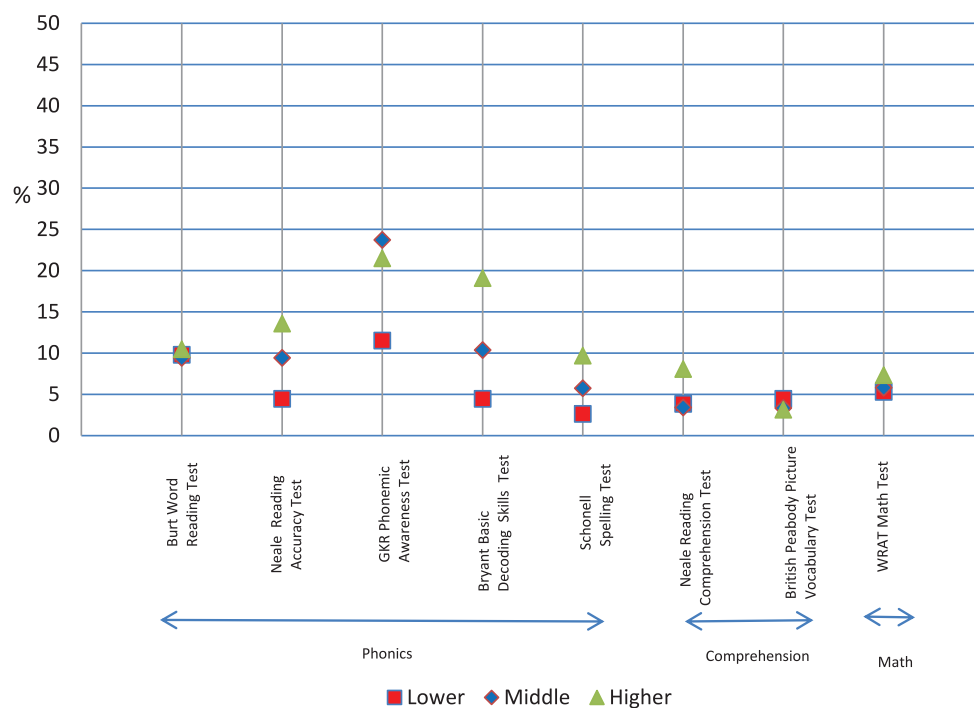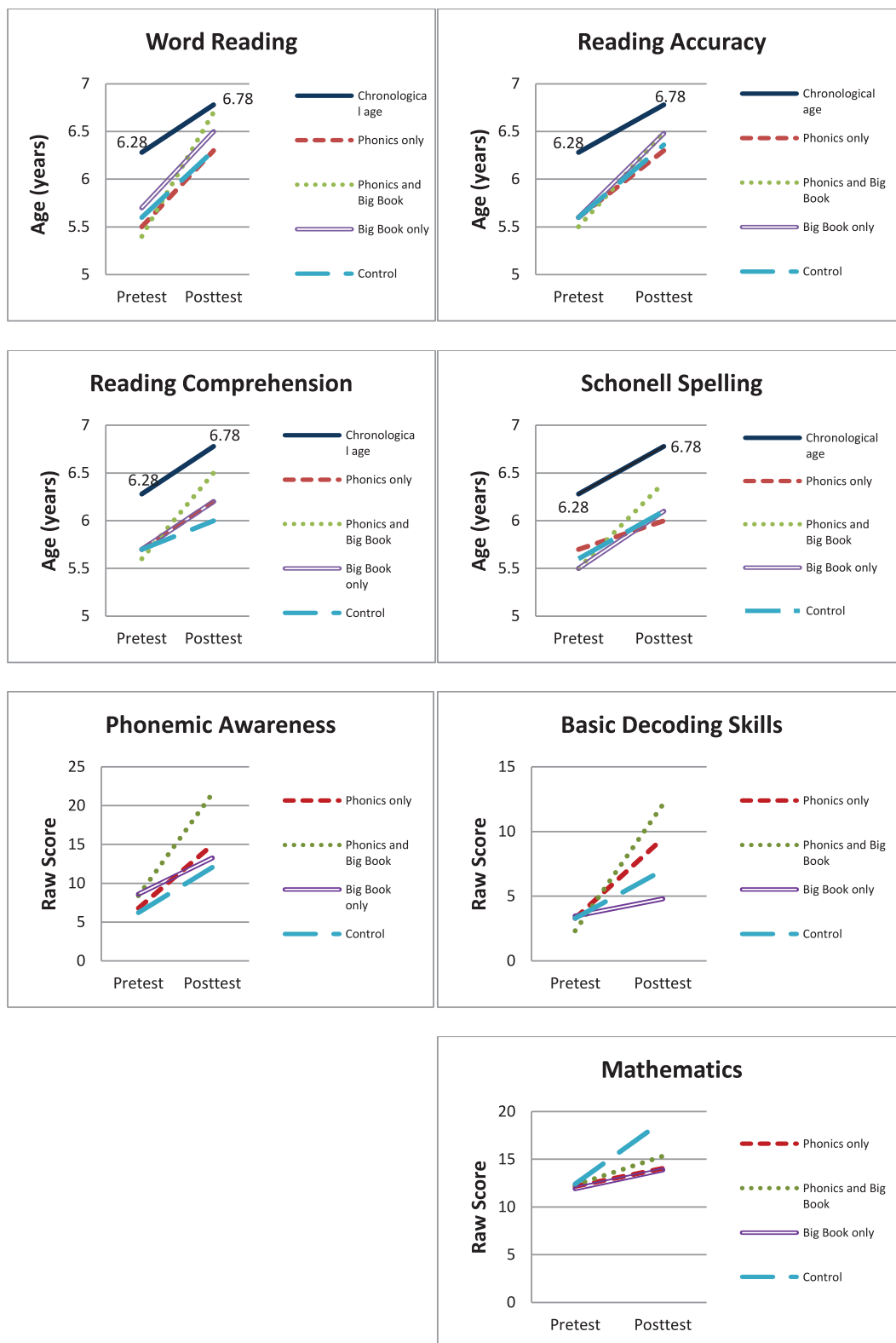


**FIGURE 7 | The pre-post difference scores for the three ability groups expressed as percentage of maximum score for each measure.**

**FIGURE 8 | Mean scores for each of the significant results for each of the training groups pretest to posttest.** Mean Burt and Neale results reported as reading ages; spelling reported as spelling age. Other results reported as raw scores.

**Table 6 | Average quiz scores: Means and Standard Deviations.**

| Reading level | | Control (Math) | Combined (BB/EP) | Big Book | Phonics |
|---|---|---|---|---|---|
| Lower | M | 1.39 | 2.51 | 1.40 | 2.16 |
|  | SD | 0.72 | 0.45 | 0.45 | 1.87 |
| Middle | M | 1.79 | 2.64 | 2.21 | 1.84 |
|  | SD | 0.70 | 0.78 | 0.66 | 0.81 |
| Higher | M | 2.26 | 2.79 | 2.35 | 2.48 |
|  | SD | 1.20 | 1.37 | 0.93 | 0.82 |
| Across levels | M | 1.81 | 2.65 | 1.99 | 2.16 |
|  | SD | 1.20 | 1.37 | 0.93 | 0.82 |

**Table 7 | Average quiz scores: separate ANOVAs for Lower, middle, and higher reading ability groups.**

| Variables | df | F | $\omega^2$ |
|---|---|---|---|
| **READING ABILITY** | | | |
| **Lower** | | | |
| Group | 3 | 9.08** | |
| Math/other | 1 | 8.70** | 0.32 |
| Combined/BB,P | 1 | 10.18** | 0.36 |
| BB vs. phonics | 1 | 8.35** | 0.31 |
| In group team | 4 | 0.57 | |
| MS error | 24 | (0.28) | |
| **Middle** | | | |
| Group | 3 | 2.80# | |
| Math/other | 1 | 2.66 | 0.09 |
| Combined/BB,P | 1 | 4.49* | 0.18 |
| BB vs. phonics | 1 | 1.28 | 0.02 |
| In group team | 4 | 2.68 | |
| MS error | 24 | (0.44) | |
| **Higher** | | | |
| Group | 3 | 0.33 | |
| Math/other | 1 | 0.35 | 0.00 |
| Combined/BB,P | 1 | 0.58 | 0.00 |
| BB vs. phonics | 1 | 0.05 | 0.00 |
| In group team | 4 | 0.58 | |
| MS error | 24 | (1.30) | |

*$p < 0.05$; **$p < 0.01$; #$p = 0.06$.

Math, control group; Combined, Big Book enhanced with phonics; Other, Combined, Big Book and Phonics; BB, Big Book; P, Phonics; In Group Team, small groups.

effects factor because each ability group received different quizzes to match their ability level. The ANOVA results are shown in **Table 7**.

The results for the lower reading ability group showed that the contrast between the control group and the average of the means of the other groups (Math/Other) was significant. The control group mean was considerably below the other groups. The contrast between the combined BB/EP group and the average of the other two reading groups was significant. Inspection of

the mean scores showed that the BB/EP group was higher than the other groups. The contrast between Big Books and Phonics means scores was significant, showing that the Phonics group scores were higher than those of the Big Book group.

The results for the middle group showed that the contrast between the control group and the average of the means of the other groups (Math/Other) was not significant. The control group had the lowest score of the four groups but the Phonics group also had a similarly low score and this probably made the difference non-significant. The contrast between BB/EP and the average mean of the Big Book and Phonics groups was significant. The contrast between Big Books and Phonics means was not significant. From this we can infer that the combined BB/EP group had a higher mean score than did the other two reading groups.

The results for the higher ability group were not significant for any of the three contrasts. This indicated that the treatments did not have differential effects for the higher ability group.

In summary, inspection of the mean scores in **Table 6** confirm the ANOVA results showing that for the lower ability group, the combined BB/EP and Phonics groups had significantly better quiz scores than the Big Book and control groups. For the middle ability reading group the combined group had better quiz scores than the other three groups. For the higher ability group, quiz scores were not significantly different among the four treatment groups.

## DISCUSSION

The model that drove this study was that combining Big Book reading with explicit phonics would have benefits across the board for a range of literacy skills, more so than Big Book reading or explicit phonics on their own. This is what the study found. The findings highlight the importance of combining necessary skills with authentic reading experience to increase literacy achievement for disadvantaged children.

The current study cuts new ground in our understanding of the impact of Big Book reading and phonics on children's literacy development. While many studies have compared Big Book (or shared book) reading with phonics none to our knowledge have compared Big Books enhanced with explicit phonics (BB/EP) with Big Book reading or phonics on their own. Many experienced researchers, such as Pressley (2006), have concluded, based on their reading of the research for each kind of instruction, that balanced instruction using both practices must be more effective than either on their own. This study is the first to show that this conclusion is correct.

### THE LITERACY GAP

A relevant question for this study was whether the treatments were closing the reading ability gap, that is, whether they were increasing the learning rate for the lower/middle ability groups relative to the higher reading ability group. This did not happen. There was no interaction between treatments and reading ability for any of the measures. The lower reading ability groups did not outpace the higher reading ability group in relative gains for any of the treatment groups. Future research could look at refinements to the present study that might help to close the literacy gap.

## SPECIFIC RESULTS

### Word reading

The Combined group did better than the other groups including the control group. In relation to the Big Books and Phonics groups this may have been because of the explicit phonics being applied to particular words from the Big Book text as part of the combined lessons (see the Appendix sample lesson). The combined instruction showed children how to use explicit phonics to help them decode words from their books. Children could see the practical application of phonics to reading in that the lessons would cover phonics aspects of some words from the Big Books before the teacher and the children began to read the Big Books. This focus on words from the books was not addressed in the Big Books group except in an incidental way and was not addressed at all in the Phonics group.

### Reading accuracy

The Combined group did as well as the Big Books group in passage reading accuracy and better than the Phonics and control groups. The results for the control group are explainable in that they did not receive reading instruction. A possible explanation for the phonics group results is that the combined group and the Big Books group both engaged in Big Book reading whereas the explicit phonics group did not engage in reading of text. Thus, the phonics group did not get the opportunity to apply their skills to book reading. Research on phonics indicates that teaching skills in isolation without opportunities to apply these skills while reading will not help them improve in book reading (Compton et al., 2014).

### Reading comprehension

The Combined group did better than the Big Books and Phonics groups in reading comprehension and this may have been because the explicit phonics in the combined instruction improved the word reading skills of children (as can be seen in their improved Burt word reading results) which in turn made the comprehension process easier by enabling the combined group children to focus more of their mental energy on comprehending what they read. There is support for this idea from other research (Tan and Nicholson, 1997) showing that improved word reading skills in a trained group produced better reading comprehension compared with a control group even though there was no difference in passage reading accuracy between the two groups. In other words the combined group did better than the other groups in comprehension because their superior word reading skills enabled them to process words more easily thus releasing more cognitive resources for comprehension.

### Phonemic awareness

The Combined group made better progress in phonemic awareness than the other groups. This was understandable for the Big Books group in that they did not receive any instruction in phonemic awareness. A possible explanation why the combined group did better than the Phonics group who also received phonemic awareness instruction might be that using the Turtle Talk strategy to learn phonemic awareness in the combined group lessons may have been more effective because the phonemic training was on words from the Big Book stories they had read and this may have been more impacting in terms of learning how to read words when reading compared with the phonics phonemic exercises which were on unrelated words that were not part of book reading.

### Basic decoding skills

The Combined and Phonics groups made similar progress in basic decoding skills and made better progress than did the Big Books group. This was understandable in that Big Book reading allows for incidental phonics learning but does not teach basic decoding skills in detail except to make use of initial consonant blends.

### Spelling

The Combined group made better progress than did the other groups. This was understandable for the control group and Big Books group who received no explicit instruction in spelling though the Big Books group may have picked up spelling skills implicitly through reading of Big Books. The phonics group did learn skills useful for spelling but these words may not have been stored as well in memory as compared with the combined group because the words covered in the phonics lessons were not part of a Big Book whereas with the combined group the spelling activities involved words from a Big Book and these words may have been more memorable in terms of storing their component letters in memory.

### Vocabulary

There were no differences among the four groups in receptive vocabulary. It was understandable that there would have been few gains in vocabulary for the phonics and control groups because they did not receive instruction in vocabulary. However there was the possibility that the Combined and Big Books groups might have improved vocabulary since they both focused on meaning and there is a strong body of research to indicate that reading books aloud to pupils improves vocabulary (McBride-Chang, 2012). The reason for the lack of an effect on vocabulary for the Combined and Big Books groups might have been that the Big Book lessons did not have enough complex vocabulary or there might not have been enough discussion of unfamiliar words. To address this issue, future research could look at the effects of adding activities that build more vocabulary and general knowledge instruction into the combined and Big Book lessons (Nicholson and Dymock, 2010; Compton et al., 2014).

### Math

The results for the control group in math showed that small group instruction in mathematics had significant benefits for them in their learning of math skills as compared with the other groups in the study who did not receive this instruction. It was understandable that the other groups would not make similar gains because they received no math instruction. The math result was the strongest in the whole study. In hindsight it would have been interesting to combine math instruction with Big Book reading to see if this would also improve math skills. There are a number of children's books that have a math aspect to them and these could

have been used to teach computation. Future research could look at this possibility.

## CONCLUSION

The findings reveal that we do not have to teach disadvantaged children in an either-or fashion, using either Big Book reading or phonics but we can combine the instruction, integrating them in a meaningful way, and produce better readers and spellers. If teachers included explicit phonics in their Big Book lessons even on a once-weekly basis, the present results indicate that this would have greater long-term benefits across more literacy measures than would Big Book reading or explicit phonics instruction on their own.

The big picture was that the combined instruction was as effective as Big Books for reading accuracy and was superior to Big Books for word reading, reading comprehension, spelling, basic decoding skills, and phonemic awareness. Likewise, the combined instruction was as effective as explicit phonics for basic decoding skills and was superior to phonics for all other measures of literacy.

To conclude, the present study found that Big Books enhanced with phonics, as compared with Big Book reading and phonics on their own, seemed to have no disadvantages and considerable advantages across a range of literacy measures. This type of balanced instruction could be a model for New Zealand and other countries wanting to find more effective ways to teach literacy to disadvantaged children, who are the ones we are very concerned about.

## ACKNOWLEDGMENTS

## REFERENCES

Arrow, A. W., and Tunmer, W. E. (2012). "Contemporary reading acquisition theory. The conceptual basis for differentiated reading instruction," in *Contemporary Debates in Childhood Education and Development,* eds S. P. Suggate and E. Reese (New York, NY: Routledge), 241–249.

Bryant, D. (1975). *Bryant Test of Basic Decoding Skills*. New York, NY: Teachers College Press.

Bus, A. G., and Ijzendoorn, M. H. (1999). Phonological awareness and early reading: a meta-analysis of experimental training studies. *J. Educ. Psychol.* 91, 403–414. doi: 10.1037/0022-0663.91.3.403

Calfee, R. C., and Patrick, C. L. (1995). *Teach our Children Well. Bringing K-12 Education into the 21st Century.* Stanford, CA: Stanford University Alumni Association.

Claessens, A., Duncan, G., and Engel, M. (2009). Kindergarten skills and fifth grade achievement. Evidence from the ECLS-K. *Econ. Educ. Rev.* 28, 415–427. doi: 10.1016/j.econedurev.2008.09.003

Compton, D. L., Miller, A. C., Elleman, A. M., and Steacy, L. M. (2014). Have we forsaken reading theory in the name of quick fix for children with reading disability? *Sci. Stud. Read.* 18, 55–73. doi: 10.1080/10888438.2013.836200

Connor, C. D., Morrison, F. J., and Katch, L. E. (2004). Beyond the reading wars: exploring the effect of child-instruction interactions on growth in early reading. *Sci. Stud. Read.* 8, 305–336. doi: 10.1207/s1532799xssr0804_1

Cowles, S. (2013). *Annual Monitoring of Reading Recovery: The Data for 2012.* Wellington: Research Division, Ministry of Education. Available online at: www.educationcounts.govt.nz

Dunn, L. M., Dunn, L., Whetton, C., and Burley, J. (1997). *The British Picture Vocabulary Scale, 2nd Edn.* Windsor: National Foundation for Educational Research-Nelson.

Ehri, L. C., Nunes, S. R., Stahl, S. A., and Willows, D. M. (2001a). Systematic phonics instruction helps students learn to read: evidence from the national reading panel's meta-analysis. *Rev. Educ. Res.* 71, 393–447. doi: 10.3102/00346543071003393

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., and Shanahan, T. (2001b). Phonemic awareness instruction helps pupils learn to read: evidence from the national reading panel's meta-analysis. *Read. Res. Q.* 36, 250–287. doi: 10.1598/RRQ.36.3.2

Elley, W. B., and Mangubhai, F. (1983). The impact of reading on second language learning. *Read. Res. Q.* 19, 53–67. doi: 10.2307/747337

Ezell, H. K., and Justice, L. M. (2005). *Shared Book Reading: Building Young Pupils' Language and Emergent Literacy Skills*. Baltimore, MD: Paul H. Brookes.

Fletcher-Flinn, C., and Thompson, G. B. (2010). "Learning to read: teaching approaches and theories of word recognition," in *Lifespan Development: New Zealand Perspectives, 2nd Edn.,* eds J. Low and P. Jose (Albany: Pearson), 72–81.

Foster, W., and Miller, M. (2007). Development of the literacy achievement gap: a longitudinal study of kindergarten through third grade. *Lang. Speech Hear. Serv. Sch.* 38, 173–181. doi: 10.1044/0161-1461(2007/018)

Gilmore, A., Croft, C., and Reid, N. (1981). *Burt Word Reading Test: New Zealand Revision*. Wellington: New Zealand Council for Educational Research.

Gough, P. B. (1996). How pupils learn to read and why they fail. *Ann. Dyslexia* 46, 3–20. doi: 10.1007/BF02648168

Gough, P. B., and Hillinger, M. L. (1980). Learning to read: an unnatural act. *Bull. Orton Soc.* 30, 179–196. doi: 10.1007/BF02653717

Gough, P. B., and Lee, C. H. (2007). A step toward early phonemic awareness: the effects of turtle talk training. *Psychologia* 50, 54–66. doi: 10.2117/psysoc.2007.54

Hattie, J. (2009). *Visible Learning: A Synthesis of over 800 Meta-Analyses Relating to Achievement*. London: Routledge.

Holdaway, D. (1982). Shared book experience: teaching reading using favorite books. *Theor. Pract.* 21, 293–300. doi: 10.1080/00405848209543021

Jeynes, W. H. (2008). A meta-analysis of the relationship between phonics instruction and minority elementary school student academic achievement. *Educ. Urban Soc.* 40, 151–166. doi: 10.1177/0013124507304128

Juel, C. (1988). Learning to read and write: a longitudinal study of 54 pupils from first through fourth grades. *J. Educ. Psychol.* 80, 437–447. doi: 10.1037/0022-0663.80.4.437

Juel, C., and Minden-Cupp, C. (2000). Learning to read words: linguistic units and instructional strategies. *Read. Res. Q.* 35, 458–492. doi: 10.1598/RRQ.35.4.2

Keppel, G., and Wickens, T. D. (2004). *Design and Analysis: A Researcher's Handbook*. Upper Saddle River, NJ: Pearson.

McBride-Chang, C. (2012). "Shared book reading. There is no downside for parents," in *Contemporary Debates in Childhood Education and Development,* eds S. P. Suggate and E. Reese (New York, NY: Routledge), 51–58.

Ministry of Education. (2001). *Reading to Read: Teacher Support Material*. Wellington: Learning Media.

Ministry of Education. (2003). *Effective Literacy Practice in Years 1 to 4*. Wellington: Learning Media.

Neale, M. D. (1999). *Neale Analysis of Reading Ability, 3rd Edn.* Hawthorn, VIC: Australian Council for Education Research.

Nicholson, T. (1997). "Closing the gap on reading failure: social background, phonemic awareness and learning to read," in *Foundations of Reading Acquisition and Dyslexia,* ed B. Blachman (Hillsdale, NJ: Lawrence Erlbaum), 381–407.

Nicholson, T. (2003). "Risk factors in learning to read, and what to do about them," in *Preventing and Remediating Reading Difficulties: Bringing Science to Scale,* ed B. Foorman (Timonium, MD: York Press), 165–196.

Nicholson, T. (2005). *The Phonics Handbook*. Chichester: Wiley.

Nicholson, T., and Dymock, S. (2014). *The New Zealand Dyslexia Handbook*. Wellington: New Zealand Council for Educational Research.

Nicholson, T., and Dymock, S. J. (2010). *Teaching Reading Vocabulary*. Wellington: New Zealand Council for Educational Research.

Nicholson, T., and Whyte, B. (1992). "Matthew effects in learning new words while listening to stories," in *Literacy Research, Theory and Practice: Views from Many Perspectives,* eds C. K. Kinzer and D. J. Leu (Chicago, IL: National Reading Conference), 499–503.

Norris, M., Bathgate, M., and Parkin, M. (1994). *Development of a Socioeconomic Indicator for Schools.* Paper presented at the meeting of the New Zealand Association for Research in Education Conference, Christchurch.

Pressley, M. (2006). *Reading Instruction that Works. The case for balanced teaching.* New York, NY: Guilford Press.

Pressley, M., and Fingeret, L. (2007). "What we have learned since the National reading Panel: visions of the next version of reading first," in *Shaping Literacy Achievement: Research we have, Research We Need,* eds M. Pressley, A. K. Billman, K. H. Perry, K. E. Reffitt, and J. M. Reynolds (New York, NY: Guilford Press), 216–245.

National Reading Panel. (2000). *Teaching Pupils to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and its Implications for Reading Instruction.* Bethesda, MD: National Institute of Child Health and Human Development.

Reardon, S. F. (2011). "The widening academic achievement gap between the rich and the poor: new evidence and possible explanations," in *Whither Opportunity? Rising Inequality and the Uncertain Life Chances of Low Income Pupils,* eds R. Murnane and G. Duncan (New York, NY: Russell Sage Foundation Press), 92–115.

Reardon, S. F. (2013). The widening income achievement gap. *Educational Leadership* 71, 9–16.

Reese, E. (2012). "The tyranny of shared-book reading," in *Contemporary Debates in Childhood Education and Development,* eds S. P. Suggate and E. Reese (New York, NY: Routledge), 59–68.

Robbins, C., and Ehri, L. C. (1994). Reading storybooks to kindergartners helps them learn new vocabulary words. *J. Educ. Psychol.* 86, 54–64. doi: 10.1037/0022-0663.86.1.54

Roper, H. D. (1984). *Spelling, Word Recognition and Phoneme Awareness among First Grade Pupils.* Unpublished Doctoral Dissertation, University of Texas, Austin.

Schonell, F. (1951). *Essentials in Teaching & Testing Spelling.* London: Macmillan.

Stevenson, J., Pennington, B. F., Gilger, J. W., DeFries, J. C., and Gillis, J. J. (1993). Hyperactivity and spelling disability: testing for shared genetic etiology. *J. Child Psychol. Psychiatry* 34, 1137–1152. doi: 10.1111/j.1469-7610.1993.tb01779.x

Suggate, S. P. (2010). Why what we teach depends on when: grade and reading intervention modality moderate effect size. *Dev. Psychol.* 46, 1556–1579. doi: 10.1037/a0020612

Suggate, S. P., Lenhard, W., Neudecker, E., and Schneider, W. (2013). Incidental vocabulary acquisition from stories: second and fourth graders learn more from listening than reading. *Language* 33, 551–571. doi: 10.1177/0142723713503144

Tan, A., and Nicholson, T. (1997). Flashcards revisited: training poor readers to read words faster improves their comprehension of text. *J. Educ. Psychol.* 89, 276–288. doi: 10.1037/0022-0663.89.2.276

Thompson, G. B. (2014). Discovering and accounting for limitations in applications of theories of word reading acquisition. *Front. Psychol.* 5:579. doi: 10.3389/fpsyg.2014.00579

Tunmer, W. E., and Nicholson, T. (2011). "The development and teaching of word recognition skill," in *Handbook of Reading Research,* eds P. D. Pearson, R. Barr, M. L. Kamil, and P. Mosenthal (New York, NY: Longman), 405–431.

Wilkinson, G. S. (1993). *Wide Range Achievement Test.* Wilmington, DE: Jastak Associates.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## APPENDIX

Lesson plan example – Lesson 1 for higher reading ability pupils, Combined group (BB/EP), 1st reading of Big Book The Hole in the King's Sock, phonics rule was silent e.



The King had cold feet.
He looked at his socks.
"There's a hole in my sock!" he cried.
"Get me the Royal Menders."
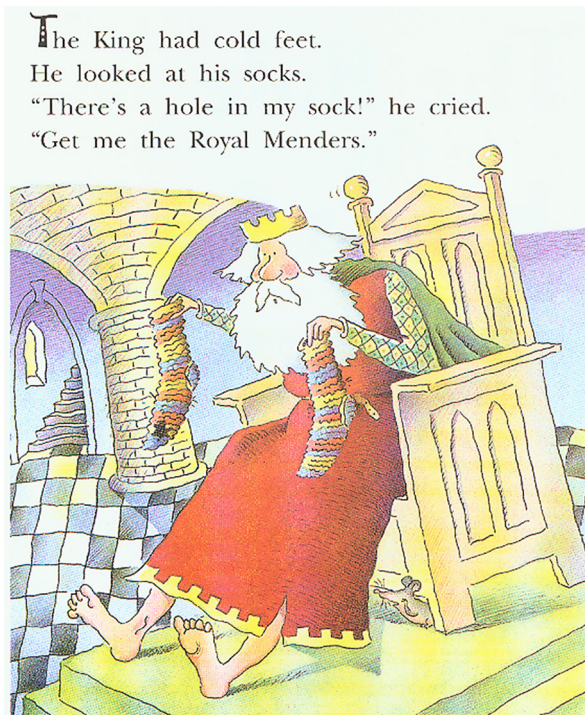
Introduction

1. Focus: Silent e rule
2. Story: *The Hole in the King's Sock* (Level-Orange)

Teacher (T): Hello, we are going to learn a rule which is called the silent e rule, and then read the story about the King who found a hole in his sock.

T: Do you know what a vowel is? In English, we have 5 letters with vowel sounds and each letter stands for two sounds, one long and one short. First of all, the 5 vowels are written as: a, e, i, o, u and sometimes y is also included as well. The sounds of the vowels usually change when there is an e at the end of the word, and we call this the silent e rule. (Then recap that the short sounds are the actual sounds of the letters and the long sounds are the names of the letters).

Lesson: (using whiteboard)

T: The silent e rule for *a_e* means that when you see the word spelled *ate* it says "ate"- the special e makes the vowel says its name. I am going to underline the vowels.



T: Remember, the letter e is silent in "ate", and this e is going to make the other vowel "ay" say its name. Let's say this word together.

T: Well done! Let's have a look at some other examples on the whiteboard.

|            |            |
|------------|------------|
| Short vowel | Long vowel |
| mat        | mate       |
| hat        | hate       |
| pet        | Pete       |
| pin        | pine       |
| cop        | cope       |
| cut        | cute       |

T: Let's have a look at these words from the story - you are going to see them in the story later.

| came | gave     | made     |
|------|----------|----------|
| wove | stitched | dough    |
| knit | knitting | wriggled |
| gold | learn    | thread   |

Students look at the word as a whole first, sounding them out if they do not know the word. They repeat and read the words 2 times.

| came: c-a-me (silent e)                    | gave: g-a-ve (silent e)                    |
|--------------------------------------------|--------------------------------------------|
| made: m-a-de (silent e)                    | wove: w-o-ve (silent e) past tense of weave |
| stitched: st-i-tch-ed (tch ="ch" sound)    | d-ough: irregular word                     |
| knit: kn-i-t (silent k)                    | knitting: kn-i-tt-i-ng                     |
| wriggled: wr-i-gg-l-ed (silent w)          | g-o-l-d                                     |
| l-ear-n                                    | th-r-ea-d                                   |

Activity: Turtle Talk (researcher selects 5–6 words from chart above)

The students listen to the phonemes of the words provided by the researcher e.g., "m-ay-deh" and they have to point out the correct word on the whiteboard. Pupils get a chance to Turtle Talk and say the word and the teacher has to guess what it is.

The teacher explains the silent e rule again when reading words from the story that had the silent e pattern - *came*, *gave*, *made*, *wove*. The word *dough* from the story is an irregular word. The -tch in *stitched* has the ch sound because ch is spelled tch after a short vowel sound. Explain that *knit* and *knitting* both have a silent k; *wriggled* has a silent w.

T: Great, I am going to read you the story of *The Hole in the King's Sock*, and I am going to ask you some questions about what happened in the story afterwards. Before we start, what are socks? Yes, they are covers we put on our feet. Where do you buy your socks from?

Pupils say: the warehouse, the supermarket, two-dollar shop.

T: Well, we will see what happens to the King's sock. Now, please listen carefully to the story (during the reading, encourage students to predict what might happen next).

Comprehension questions (orally)

1. What was the King's problem?
2. Did he find a solution? What was that?
3. Did it work? If not, why did it not work?
4. Was the problem solved at the end?

# Unrecognized ambiguities in validity of intervention research: an example on explicit phonics and text-centered teaching

## G. Brian Thompson *

School of Education, Victoria University of Wellington, Wellington, New Zealand
*Correspondence: brian.thompson@vuw.ac.nz

A commentary on

**The effect of phonics-enhanced Big Book reading on the language and literacy skills of six-year-old pupils of different reading ability attending lower SES schools**
*by Tse, L. and Nicholson,T. (2014). Front. Psychol. 5:1222. doi: 10.3389/fpsyg.2014.01222*

Tse and Nicholson (2014) have tested a small-group instructional intervention that they propose as a modification to enhance reading progress among low attainment 6-year-olds in a "text-centered" teaching approach. The authors (T&N) cite a Ministry of Education (2003) handbook to describe this approach. It has four main components (pp. 91–101): (i) Teacher Reading of texts to listening children, (ii) Shared Reading in which the children engage in watching the text print ("Big Books") as the teacher shows how it matches the spoken text, (iii) Guided Reading in which there is detailed teacher support of the individual children's attempts at reading a text (e.g., for "using word-level information to decode new words" p. 97), (iv) Children's Independent Reading of texts (with minimal errors) by themselves for individual levels and interests. This report of T&N, however, lacked evidence about what the children received of each of these components prior to, and concurrent with, the intervention study. Without such evidence we cannot tell in what way the instructional interventions were the same, different from, or in conflict with other instruction received.

T&N's proposed modification to the Shared Reading component was to combine it with systematically taught explicit phonics (a "sounding out" procedure in which the child pronounces successive sounds of letters of a word to generate an oral reading response). For theoretical justification of this modification, T&N cited some of the claims of Gough and Hillinger (1980) but omitted others, that phonics "gives the child artificial rules ......to learn the real rules" (p. 192), which "are unconscious and implicit" (p. 187). This implies that phonics is a heuristic procedure for initial instruction but subsequently discarded without any disadvantage [although Thompson et al. (2009) found evidence to the contrary]. Neither T&N nor their citation of Gough and Hillinger provide justification for the particular phonics rules (e.g., final $e$-marker of "long" vowels) and corresponding sounds (e.g., for vowel digraphs) selected for instruction (T&N, Table 2) of these 6-year-olds with word reading test ages in the lower half of the normative distribution, and a mean aural vocabulary test age of 4 years 8 months (determined from BPVT norms using raw scores in T&N, Table 4). T&N found no effect of their intervention on the children's aural vocabulary but were silent on why the overall text-centered approach, with their modification, would be suitable for children with an apparent large developmental lag in understanding spoken English.

T&N gave no report of the opportunities that the items of the pre-and post-test measures provided for children to use the taught phonics procedures. Interpretation of results for each measure depends upon the extent to which these opportunities were provided; and for comparison between measures, whether such opportunities were equal or different. For each reading measure the writer determined the percentage of word items that provided this opportunity among items in the applicable reading-level range. For example, this was 34% of items providing opportunities in the decoding skills measure. It was, however, 16% in the isolated word reading, and in this there were also 16% that provided conflicting opportunities because the taught procedures could not work (e.g., final-$e$ marker of "long vowels" in the words *one, love*). The decoding skills items had no conflicting opportunities. Hence, any superior score gains for this measure could be just an artifact of more (workable) opportunities. Another unbalanced feature of the design is noted. The phonics procedures demonstrated to the children were followed up by their individual attempts at weekly "quizzes" (T&N, Table 3; Figure 4). There were no similar individual opportunities involving text reading, which could disadvantage performance on that measure.

The pre- to post-test performance gain of the intervention that combined phonics with shared reading was compared with the mean of the gains of shared reading and explicit phonics interventions, each taught separately. In these comparisons of performance gains, oral reading of isolated words and decoding skill (pseudowords) had substantially greater gains

for the combined intervention than the separate interventions. In contrast, the gain in word accuracy in oral text reading was not greater for the combined intervention, failing to reach a statistically significant difference (T&N, Table 5). This orthogonal contrasts analysis, although relevant to the purpose of the study, was not sufficient for this randomized treatments-versus-control design. It also required statistical comparisons between the performance gains of the combined intervention sample and the (math-only) sample that controlled for gains in reading performance from influences external to the intervention. Without these there is no basis to confirm the T&N interpretation that the combined instructional treatment had some significant effects.

Speed of reading was a score in the test of text reading but was not reported, although relevant to comparison of phonics and text-centered instruction (Thompson et al., 2008). And critically, there was no report of the extent to which the children made successful use of the taught explicit phonics in their word responses in text reading, or any of the other reading outcomes. Without this information we are left to speculate whether T&N's claimed (but unconfirmed) positive intervention effects for isolated words and pseudowords could have been an outcome of the children acquiring implicit sublexical processes (Thompson and Fletcher-Flinn, 2012; Thompson, 2014) from the isolated word exemplars for the taught phonics rather than the children's use of the phonics.

Apart from omission of the required statistical comparisons, the design and its implementation in this study may rate above average on a list of validity criteria such as Troia (1999) but our focus has been mainly on ambiguities in validity not often recognized in research on instructional interventions. Included in these are lack of information and evidence for (i) the context of both prior and concurrent instruction, (ii) how the intervention fits wider teaching goals and other instructional needs of the participants, (iii) the extent, and balance, of opportunities in the outcome measures to use procedures that were taught, and (iv) children's use of those procedures in such opportunities, (v) testing contrary predictions from alternative theories.

## REFERENCES

Gough, P. B., and Hillinger, M. L. (1980). Learning to read: an unnatural act. *Bull. Orton Soc.* 30, 179–196.

Ministry of Education. (2003). *Effective Literacy Practice in Years 1 to 4*. Wellington: Learning Media.

Thompson, G. B. (2014). Discovering and accounting for limitations in applications of theories of word reading acquisition. *Front. Psychol.* 5:579. doi: 10.3389/fpsyg.2014.00579

Thompson, G. B., Connelly, V., Fletcher-Flinn, C. M., and Hodson, S. J. (2009). The nature of skilled adult reading varies with type of instruction in childhood. *Mem. Cogn.* 37, 223–234. doi: 10.3758/MC.37.2.223

Thompson, G. B., and Fletcher-Flinn, C. M. (2012). "Toward better teaching: revising the fundamentals of learning to read," in *Contemporary Debates in Childhood Education and Development*, eds S. Suggate and E. Reese (Abington: Routledge), 250–260.

Thompson, G. B., McKay, M. F., Fletcher-Flinn, C. M., Connelly, V., Kaa, R. T., and Ewing, J. (2008). Do children who acquire word reading without explicit phonics employ compensatory learning? Issues of phonological recoding, lexical orthography, and fluency. *Read. Writ.* 21, 505–537. doi: 10.1007/s11145-007-9075-9

Troia, G. A. (1999). Phonological awareness intervention research: a critical review of the experimental methodology. *Read. Res. Q.* 34, 28–52.

Tse, L., and Nicholson, T. (2014). The effect of phonics-enhanced Big Book reading on the language and literacy skills of six-year-old pupils of different reading ability attending lower SES schools. *Front. Psychol.* 5:1222. doi: 10.3389/fpsyg.2014.01222

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

# What are the criteria for a good intervention study? Response: "Unrecognized ambiguities in validity of intervention research: an example on explicit phonics and text-centered teaching"

Tom Nicholson[1]* and Laura Tse[2]

[1] Institute of Education, Massey University, Auckland, New Zealand, [2] School of Curriculum and Pedagogy, The University of Auckland, Auckland, New Zealand

**A commentary on**

**Unrecognized ambiguities in validity of intervention research: an example on explicit phonics and text-centered teaching**
*by Thompson G. B. (2015). Front. Psychol. 5:1535. doi: 10.3389/fpsyg.2014.01535*

Thompson (2015) has raised several validity issues about our study (Tse and Nicholson, 2014) while acknowledging that it would score well in terms of Troia's (1999) criteria for "What makes a good study?" A response to the critique is detailed briefly below.

Thompson's first point was that the study lacked evidence about what instruction children received prior to and concurrent with the intervention study. Interactions with teachers and the principals of the schools however indicated that reading instruction was similar from one school to the next. Any differences among schools and classrooms were also controlled for in that participants were randomly assigned to groups thus spreading possible effects of differences in instruction across all groups.

The second point was the absence of justification for the phonics rules taught however the article explained that the taught Anglo-Saxon decoding rules were from Calfee and Patrick's (1995) well-known explanation of Anglo-Saxon letter-sound patterns. The intervention followed their scope and sequence in the study. It is not clear why this might be a validity problem in that the study did reference the source of the phonics rules.

The third point was that participants' vocabulary age was low at 4.8 years compared with chronological age of 6.3 years and thus Big Books may have been inappropriate. Their standard score was 86 which is close to the average range (90–110) and there are studies to support Big Book reading with lower SES children such as these (Nicholson and Whyte, 1992; Valdez-Menchaca and Whitehurst, 1992; Whitehurst et al., 1994). The Big Books were also selected so as to be at the reading level of the children who were being taught and given that their reading level was in the beginner range the language should have been understandable for them.

The fourth point was that the article did not discuss whether children had opportunities to use their decoding skills to process the items of the pre and post-test measures. Although not reported our data did confirm that the combined group scored better

on regular words (e.g., went) than irregular (e.g., love). The Bryant Test of Basic Decoding skills also gave opportunities to use decoding skills.

The fifth point was that the phonics group practiced phonics quizzes but the Big Book group did not practice reading of text. This was not completely the case. Children in the Big Book group did get opportunities to practice reading of text through the Big Book lessons. They did three readings of each text and read along with the teacher.

The sixth point was that the orthogonal analysis was not sufficient and needed to compare the performance gains of the combined group with those of the treatment control group (math-only). To do this however risked statistical error so instead of carrying out all possible comparisons among the four groups the decision was to use Helmert contrasts which were pre-planned orthogonal contrasts. This approach offered protection against statistical error (Kwon, 1996; Keppel and Wickens, 2004). As Kuehne (1993) has pointed out, using post-hoc comparisons increases the chance of type 1 error (in the study, to do six *post-hoc* comparisons across four groups would increase the possibility of type 1 error to 26%). The Helmert contrast procedure is common in other disciplines but less common in education. The way the Helmert contrasts worked in the study was that the control group mean was first compared with the overall mean score for the other three groups. Then the phonics enhanced Big Books group mean was compared with the overall mean for the two remaining groups (Big Book and phonics).

Finally the means of the Big Book and phonics groups were compared. It was like peeling an onion. The logic was that if the control group was not better than the mean of the other three groups and if the phonics enhanced group was better than the mean of the combined Big Book and phonics groups, and if there was no difference in the contrast between the Big Book and phonics groups, then it can be inferred that the phonics enhanced group was superior to the other groups. The orthogonal contrast worked just as well as all possible contrasts with less risk of type 1 and 2 error.

The seventh issue was that speed of reading was not reported. Thompson's previous research would suggest a slower reading speed for the phonics enhanced Big Book group but it could counter-wise be argued that they would have gained similar fluency to the Big Book group because they also read Big Books. To answer this question, fluency would be a useful variable for future studies to find out which approach is more effective for fluency.

To conclude, one reviewer commented that the present study could be "a model for how such work might be conducted on a larger scale, which might lead New Zealand and other nations to progress in dealing with the [achievement] gap issue." Replicating and scaling up the present study will clarify further whether enhancing Big Book reading with explicit phonics brings disadvantaged children closer to their expected reading and spelling age in a short time with only a small adjustment to Big Book instruction.

# References

Calfee, R. C., and Patrick, C. L. (1995). *Teach Our Children Well. Bringing K-12 Education into the 21st Century*. Stanford, CA: Stanford University Alumni Association.

Keppel, G., and Wickens, T. D. (2004). *Design and Analysis*. Upper Saddle River, NJ: Pearson.

Kuehne,. C. C. (1993). "The advantages of using planned comparisons over post hoc tests," in *Paper Presented at the Annual Meeting of the Mid-South Educational Research Association* (New Orleans, LA).

Kwon, M. (1996). "The use of planned comparisons in analysis of variance research," *Paper Presented at Annual Meeting of the Southwest Educational Research Association* (New Orleans, LA).

Nicholson, T., and Whyte, B. (1992). "Matthew effects in learning new words while listening to stories," in *Literacy Research, Theory, and Practice: Views from Many Perspectives* eds C. K. Kinzer and D. J. Leu (Chicago: National reading Conference), 499–503.

Thompson, G. B. (2015). Unrecognized ambiguities in validity of intervention research: an example on explicit phonics and text-centered teaching. *Front. Psychol.* 5:1535. doi: 10.3389/fpsyg.2014.01535

Troia, G. A. (1999). Phonological awareness intervention research: a critical review of the experimental methodology. *Read. Res. Q.* 34, 28–52.

Tse, L., and Nicholson, T. (2014). The effect of phonics-enhanced Big Book reading on the language and literacy skills of six-year-old pupils of different reading ability attending lower SES schools. *Front. Psychol.* 5:1222. doi: 10.3389/fpsyg.2014.01222

Valdez-Menchaca, M. C., and Whitehurst, G. J. (1992). Accelerating language development through picture book reading: a systematic extension to Mexican day-care. *Dev. Psychol.* 28, 1106–1104.

Whitehurst, G. J., Arnold, D. S., Epstein, J. N., Angell, A. L., Smith, M., and Fischel, J. E. (1994). A picture book reading intervention in day-care and home for children from low-income families. *Dev. Psychol.* 30, 679–689.

# Developmental differences in masked form priming are not driven by vocabulary growth

*Adeetee Bhide[1], Bradley L. Schlaggar[1,2,3,4]\* and Kelly Anne Barnes[1]*

[1] Department of Neurology, Washington University School of Medicine, St. Louis, MO, USA
[2] Department of Radiology, Washington University School of Medicine, St. Louis, MO, USA
[3] Department of Pediatrics, Washington University School of Medicine, St. Louis, MO, USA
[4] Department of Anatomy and Neurobiology, Washington University School of Medicine, St. Louis, MO, USA

As children develop into skilled readers, they are able to more quickly and accurately distinguish between words with similar visual forms (i.e., they develop precise lexical representations). The masked form priming lexical decision task is used to test the precision of lexical representations. In this paradigm, a prime (which differs by one letter from the target) is briefly flashed before the target is presented. Participants make a lexical decision to the target. Primes can facilitate reaction time by partially activating the lexical entry for the target. If a prime is unable to facilitate reaction time, it is assumed that participants have a precise orthographic representation of the target and thus the prime is not a close enough match to activate its lexical entry. Previous developmental work has shown that children and adults' lexical decision times are facilitated by form primes preceding words from small neighborhoods (i.e., very few words can be formed by changing one letter in the original word; low N words), but only children are facilitated by form primes preceding words from large neighborhoods (high N words). It has been hypothesized that written vocabulary growth drives the increase in the precision of the orthographic representations; children may not know all of the neighbors of the high N words, making the words effectively low N for them. We tested this hypothesis by (1) equating the effective orthographic neighborhood size of the targets for children and adults and (2) testing whether age or vocabulary size was a better predictor of the extent of form priming. We found priming differences even when controlling for effective neighborhood size. Furthermore, age was a better predictor of form priming effects than was vocabulary size. Our findings provide no support for the hypothesis that growth in written vocabulary size gives rise to more precise lexical representations. We propose that the development of spelling ability may be a more important factor.

Keywords: priming, vocabulary, lexical precision, reading, developmental

## INTRODUCTION

Learning to read, unlike learning to speak, requires explicit instruction. Models of reading attempt to account for reading performance across development. However, several unresolved questions prevent models of reading skill acquisition from being further refined. Specifically, what are the mechanisms that allow fluent readers to distinguish between words that are visually similar but have different meanings?

Masked priming paradigms (Forster et al., 1987) provide a means for studying developmental changes in orthographic processing. In masked priming paradigms, a prime is presented briefly (c. 50 ms) and is masked by a row of hash marks that precedes it and a target word that follows it (typically in a different-case font). Participants are typically unaware of the primes because of their rapid and masked presentation, and hence cannot use different strategies for processing the primes. Therefore, this paradigm is particularly useful for studying developmental changes because it can distinguish age-related differences from differences in strategic processing.

Form priming, where the prime and target differ by a single letter (e.g., clee-FLEE), provides a way to measure the precision of orthographic representations. Form priming in adults varies as a function of orthographic neighborhood size (N), the number of words that can be formed from a target word by changing a single letter. For example, the word *echo* has no orthographic neighbors, whereas the word *yell* has 9 orthographic neighbors including *yelp* and *cell*. For adults, masked non-word form primes significantly facilitate lexical decision times relative to unrelated primes (e.g., pilk-FLEE) when the target word has few orthographic neighbors, but not when the target word has many neighbors (Forster, 1987; Segui and Grainger, 1990; Forster and Davis, 1991; Castles et al., 1999).

Competing hypotheses have been proposed regarding the mechanisms underlying neighborhood size effects on adult form priming. The first hypothesis, an entry-opening search model (Forster and Davis, 1984; Forster, 1989), is predicated on the idea that word detectors are more sharply tuned for high than low N words, to minimize confusion with other visually similar words.

This sharper tuning may be accomplished by recoding words into their bodies and antibodies (Forster and Taft, 1994). According to this hypothesis, form primes are not sufficiently close matches to high N words to facilitate response times, resulting in less priming for high than low N words. Another hypothesis, a network framework model based on Rumelhart and McClelland's (1982) interactive activation model, posits that greater levels of competition or inhibition between neighbors for high than low N targets offset the facilitatory effect of the prime. A recent study from Andrews and Hersch (2010) found that for adults who were "above average" spellers, form primes preceding high N targets led to a significant slowing of lexical decision response time (i.e., the prime inhibited rather than facilitated response times). Such inhibitory effects are not reconcilable with search models, but can be accommodated by models postulating inhibitory links.

Masked form priming for high N words has been shown to vary across development, suggesting experience, maturation, or an interaction between the two alters the mechanisms that yield this behavioral effect. Unlike adults, children show masked form priming for both low and high N words (Castles et al., 1999, 2003, 2007). Castles et al. (1999) initially hypothesized that written vocabulary growth, and its effects on either lexical tuning or lexical competition (Castles et al., 2007), is related to developmental differences in neighborhood density effects on form priming. Since children have smaller written vocabularies than adults, they may not know all of the neighbors of a high N target, rendering the target effectively low N. As neighbors of a particular word are learnt, vocabulary growth would either initiate the recoding of high N words or increase the number of inhibitory links associated with that particular word. Some studies support this theory, documenting attenuation of form priming with age (Castles et al., 2007). Other studies have continued to show large priming effects during developmental periods when written vocabulary was presumed to increase. For example, Castles et al. (1999) reported that children continued to show facilitation from 2nd grade until 6th grade, when vocabulary testing revealed that the high N targets were effectively high N for the 6th graders. This finding led to the hypothesis that a neighborhood density threshold has to be reached before lexical detectors begin to narrow their tuning. This hypothesis was supported by the observation that the 6th graders with the highest sight vocabularies showed no form priming for high N targets (Castles et al., 1999).

The purpose of this paper is to test the hypothesis that differences in written vocabulary size underlie developmental differences in form priming. To test this hypothesis, we controlled for, and quantified, effective N. We used very low N (0–1 neighbors) and very high N (≥10 neighbors) targets, which afforded us two potential advantages. First, the use of very high N stimuli provided an opportunity to test whether we could replicate Andrews and Hersch's (2010) finding of significant inhibition for high N form priming in adults, as such high N stimuli would be expected to generate substantial lexical competition. Second, the use of very high N stimuli increased the likelihood that these stimuli would be high N for young children, who might know only a fraction of the neighbors. We measured children's knowledge of these neighbors to determine whether individual differences in form priming related to individual differences in neighbor knowledge, as would

be expected if vocabulary size related to high N form priming. As a final test, we created a "matched" set of stimuli whose average N equaled the estimated effective N for the youngest children. These matched stimuli were shown only to the adults to control for potential differences in effective N between children and adults. If significant differences were seen between the matched N stimuli in adults and the high N stimuli in children, then the likelihood is markedly reduced that vocabulary differences are the cause of developmental differences in priming. In the first analysis, we used linear mixed effects modeling to calculate the expected reaction time to stimuli that were preceded by different prime types as a function of age. We expected everyone to show facilitation due to repetition primes (e.g., flee-FLEE) and form primes preceding low N targets. However, we expected only the younger participants to show facilitation when form primes preceded high N targets. The second analysis also entailed a linear mixed effects model, but the matched N targets were inputted for the adults to control for effective neighborhood size. If we found no differences between the adults' matched N words and the children's high N words, then the results would support the written vocabulary growth hypothesis. In contrast, if we found significant differences between the two age groups, wherein children were facilitated in the high N form priming condition but adults were not facilitated in the matched N form priming condition, then the results would not support the written vocabulary growth hypothesis. In our final analysis, we tested whether age or vocabulary size was a better predictor of high N form priming effects in children and adolescents. Vocabulary being the better predictor would support the written vocabulary growth hypothesis, whereas the opposite result would not.

A second goal of this study was to map the developmental trajectory of form priming. Previous studies have primarily examined discrete age groups (e.g., testing children in 2nd, 4th, and 6th grade, as well as adults). This study is the first priming study, to our knowledge, to test age as a continuous variable from childhood through adulthood.

## METHODS

### PARTICIPANTS
Twenty-seven adults (18–27 years old, 11 males), 26 adolescents (13–17 years old, 13 males), and 38 children (7–12 years old, 15 males) were recruited from Washington University in St. Louis and the St. Louis metropolitan area. All participants were monolingual, native English speakers, with normal or corrected-to-normal vision and no history of neurological or psychiatric disorders. Adult participants and parents of child and adolescent participants provided informed consent and child and adolescent participants provided informed assent. Participants were compensated $15/h. All aspects of the study were performed with the approval of the Washington University Human Studies Committee.

Subject data were included in the study if: (1) accuracy in each condition of the lexical decision task was above chance (50%), and (2) the participant did not report being able to read the primes. Thirteen children and 2 adolescents did not meet the accuracy criteria. One adult reported being able to see the primes. The final sample comprised 26 adults (18–23 years old, 11 males),

24 adolescents (13–17 years old, 13 males), and 25 children (8–12 years old, 12 males). Although we did not collect IQ and reading measures from all children, the available IQ and reading data did not differ between included and excluded participants. Additional demographic and psychometric data are reported in **Table 1**.

Thirteen out of 38 children had to be removed from the analysis due to low accuracy. This issue does not appear unique to our cohort. For example, Castles et al. (1999) stated that "a number" of grade 2 children (mean age = 7 years, 10 months) needed to be removed from the analysis due to low accuracy, but did not state how many. Fifteen percent of the grade 3 children (mean age = 8 years, 6 months) tested in the Castles et al. (2007) study had to be removed due to low accuracy. The neighborhood size of stimuli used in the present study may explain why this experiment was more challenging than previous studies. Low N words and high N non-words are the most challenging stimuli to correctly classify. The low N word targets in the Castles et al. (1999) study had a mean N of 1.3, whereas in the present study they had 0 neighbors. The high N non-words in the Castles et al. (1999) study had a mean N of 8.9, whereas in the present study they had a mean N of 12.7 (range: 10–19).

## DESIGN AND STIMULUS MATERIALS
### Design
The lexical decision task contained both word and non-word stimuli; only words were analyzed. The words and non-words varied by Orthographic Neighborhood Size (high, low) and Prime Type (repetition, form, unrelated). An additional Orthographic Neighborhood Size condition (matched N) was presented to adult participants as a control condition. Children and adolescents completed a neighbor knowledge test to measure their effective Ns. Adolescents also completed psychometric testing; children and adults did not, but some psychometric data were available from previous studies in the lab. We describe three discrete age groups in the Methods section, as the testing procedures were slightly different for each age group. However, in the analyses, age is treated as a continuous variable.

**Table 1 | Participant information.**

|  | *N* | Age in years | *N* | Age in years | IQ | Reading standard score |
|---|---|---|---|---|---|---|
| Children | **38** | **10.31 (1.39)** | 25 | 10.57 (1.14) | 124* | 118.62* |
| Adolescents | **26** | **15.10 (1.24)** | 24 | 15.22 (1.20) | 111 | 103.25 |
| Adults | **27** | **20.98 (1.21)** | 26 | 20.91 (1.18) | 126* | N/A |

*Bold values pertain to all tested participants. Non-bold data pertain to the participants who met the inclusion criteria. The asterisk (\*) signifies that the data is an estimate based on a subset of the included population. The Vocabulary and Matrix Reasoning Subtests of the Wechsler Abbreviated Scale of Intelligence were used to calculate the IQ scores. The Letter-Word ID, Word Attack, and Reading Fluency subtests of the Woodcock Johnson (WJ) Tests of Achievement were used to calculate the reading standard scores. Data are presented as means with standard deviations (SD) in parentheses, except where noted.*

### Stimulus materials: lexical decision task
Stimuli were white letter strings displayed in the center of the screen in Courier font on a black background. Stimuli subtended 0.57 visual degrees vertically and up to 1.64 visual degrees horizontally. The mask had a contrast value of 0.47 and the other stimuli had similar contrast values.

Target items were 210 4–5 letter English words and 210 4–5 letter legal non-words compiled using the e-Lexicon database (Balota et al., 2007). All the word targets shown to children and adolescents had a 3rd grade frequency $\geq 1$ (Zeno et al., 1995). The target words had one of three orthographic neighborhood sizes: 70 were high N (10–19 orthographic neighbors), 70 were low N (0 orthographic neighbors), and 70 were matched/medium N (8–9 orthographic neighbors) (Balota et al., 2007). The target non-words had the same characteristics, save that the low N non-words had 1 orthographic neighbor. The orthographic neighborhood size for the matched N stimuli, shown to adults as a control for children's smaller vocabularies, was selected to approximate the expected effective N of the high N list for the youngest children. Specifically, stimuli in matched N list had 8–9 orthographic neighbors, which corresponded to the average number of neighbors per high N word target with a 3rd grade frequency $\geq 1$ (Zeno et al., 1995).

Lexical properties of the target stimuli are displayed in **Table 2**. There were no effects of Orthographic Neighborhood Size or Lexicality on letter string length, and no effect of Orthographic Neighborhood Size on HAL (Hyperspace Analog to Language) frequency (Lund et al., 1995)[1], HAL log frequency, and number of syllables for words, all *p*s > 0.10. The HAL frequency of the orthographic neighbors for the matched N words and high N words did not differ, $p = 0.92$. The 3rd grade frequency of the high N and low N words also did not differ, $p = 0.39$.

Each trial began with a forward mask consisting of a row of Xs, matched in length to the number of letters in both the prime and target (e.g., XXXX for a four-letter prime/target). Although an "x" is an English letter, it was only found in three targets. The forward mask was presented for 800 ms. Next, a prime was presented in lowercase font for 66.66 ms. This prime duration was chosen to closely approximate the prime duration in Castles et al. (1999) (57 ms) given our monitor refresh rate of 13.33 ms. Although the prime duration is slightly longer than usual, all of our participants except one reported either being completely unaware of the prime or of simply seeing a flicker on the monitor. Despite suggestions that prime visibility has a minimal impact on behavioral effects (Schmidt, 2013), we elected to exclude that participant because she occasionally made lexical decisions to the prime instead of the target stimulus. Then, a target was presented in uppercase font for 800 ms or until a lexical decision was made. Participants were instructed to determine whether the target was a real English word or a "made-up" word and to indicate their response by either the left or right button on a button box with the corresponding index finger. Response mappings were counterbalanced across participants.

---

[1]The HAL frequency was derived using a corpus of words from Usenet, which includes all newsgroups using English dialog.

The number of letters in the prime and target was equal for each trial. Repetition primes were characterized by the same item appearing in lowercase font as a prime and uppercase font as a target. Form primes were characterized as differing by one letter position from the target. All letter positions were changed an equal number of times. Unrelated primes shared a maximum of one letter in the same position as the target. The lexicality of the prime was selected to give no indication of the lexicality of the target item. For repetition prime trials, word primes always preceded word targets (e.g., rice-RICE) and non-word primes always preceded non-word targets (e.g., deat-DEAT). For form prime trials, non-word primes preceded word targets (e.g., ruce-RICE) and word primes preceded non-word targets (e.g., dean-DEAT). For unrelated prime trials, half of the targets were preceded by non-word primes and half were preceded by word primes (e.g., lunt-RICE or epic-RICE; tond-DEAT or milk-DEAT). All unrelated non-word primes were orthographically legal. Non-word form primes were created by replacing consonants with other consonants and vowels with other vowels. The prime lexicality was chosen to replicate the Castles et al. (1999) experimental design to facilitate cross-study comparisons.

For each target, three primes (repetition, form, and unrelated) were created. Initially, 12 lists (6 for adults, 6 for children/adolescents) were created with different combinations of the 3 prime types, so that every participant viewed each target once but, across the sample, every target was preceded by every prime type. Each list was then pseudorandomized with the constraint that no more than 6 examples of a particular response type were presented sequentially. Two pseudorandomized versions were generated for each list, yielding a total of 24 stimulus lists (12 for adults, 12 for children/adolescents) (see Supplementary Material for a list of stimuli).

### Stimulus materials: neighbor knowledge test

The neighbor knowledge test served as our in-house vocabulary test and was used to measure the children's and adolescents' knowledge of the neighbors of the high N stimuli. Stimuli were white letter strings displayed in Courier font on a black background. Stimuli subtended 0.573 visual degrees vertically and up to 1.637 visual degrees horizontally. The possible targets were the 605 unique neighbors of the word targets from the lexical decision task. The foils were 195 orthographically legal non-words.

Lexical properties of the target stimuli are displayed in **Table 3**. The targets and foils were randomly divided into 5 lists of 160 items. 20–30% ($M = 24.38\%$) of each list consisted of non-word foils. Each list was pseudorandomized, with the constraint that no more than 6 examples of a particular response type were presented sequentially.

Each trial began with a centered row of Xs presented for 800 ms, matched in length to the number of letters to the target/foil (e.g., XXXX for a four-letter item). Then, a centered target or foil was presented in lowercase font until the participant made a lexical decision. Response mappings were kept constant from the priming experiment.

Adults did not take the Neighbor Knowledge Test because we assumed that the adults knew most of the neighbors. The neighbors were fairly frequent [Hyperspace Analog of Language (HAL) mean: 26639.6, range: 16–1060831], and the adults were of high ability (estimated IQ = 126, all but one adult participant were students at Washington University in St. Louis).

### PROCEDURE

Participants were tested individually in a dark, quiet, and windowless room. Stimulus presentation and response collection was controlled by PsyScope X (Carnegie Mellon University, Build 53) scripts running on an Apple OS X computer. Stimuli were displayed on a Trintron PC monitor. Participants' heads were held in place by a chin rest positioned 70 cm from the display monitor. Child and adolescent participants completed the practice trials, the lexical decision task, and the neighbor knowledge test. Adolescent participants also completed psychometric testing.

Table 3 | Lexical properties of neighbor knowledge test stimuli.

|  | Length | Freq. HAL | Log freq. HAL | Number of orthographic neighbors | Number of syllables |
|---|---|---|---|---|---|
| Words | 4.51 (0.5) | 26640 (85148) | 8.01 (2.2) | 10.09 (4.2) range: 1–24 | 1.10 (0.3) |
| Non-words | 4.34 (0.5) |  |  | 8.04 (2.7) range: 0–22 |  |

Properties of the target stimuli expressed as mean (standard deviation). Abbreviations and conventions as in **Table 2**.

Table 2 | Lexical properties of target stimuli.

|  |  | Length (no. of letters) | No. of syllables | Freq. HAL | Log freq. HAL | 3rd grade freq. | Range of neighbors | No. of neighbors | HAL freq. of neighbors |
|---|---|---|---|---|---|---|---|---|---|
| Words | Low N | 4.76 (0.4) | 1.13 (0.3) | 50796 (163900) | 8.75 (2.0) | 102.83 (299.6) | 0 | 0.00 (0.0) | |
| | Matched N | 4.73 (0.4) | 1.09 (0.3) | 45884 (156600) | 8.88 (1.8) | | 8–9 | 8.59 (0.5) | 7.96 (0.8) |
| | High N | 4.66 (0.5) | 1.06 (0.2) | 36436 (71635) | 9.08 (1.9) | 67.23 (177.5) | 10–19 | 13.06 (2.8) | 7.97 (0.8) |
| Non-words | Low N | 4.74 (0.4) | | | | | 1 | 1.00 (0.0) | |
| | Matched N | 4.73 (0.4) | | | | | 8–9 | 8.59 (0.5) | |
| | High N | 4.66 (0.5) | | | | | 10–19 | 12.70 (2.6) | |

Properties of the target stimuli expressed as mean (standard deviation). Frequency (freq.) measures were identified from the Hyperspace Analog to Language (HAL) estimates in the e-Lexicon database. Coltheart's N was used to identify neighbors.

Adult participants completed the practice trials and the lexical decision task.

Participants were instructed to determine whether each stimulus was a real word or a "made-up" word. Participants were told to respond as quickly and accurately as possible, and not to worry if they made an occasional mistake.

Additional instructions were given to children and adolescents to reduce the high false alarm rate seen during pilot testing. They were told that all of the words were fairly easy words, which they might have read before in books and whose meaning they knew. Additionally, if something looked like a word, but they hadn't read it before or did not know what it meant, it probably was a made-up word.

### Practice trials

After receiving instructions, child and adolescent participants viewed 10 flashcards, with different example stimuli, and were asked to determine whether each stimulus was a word or a non-word. The experimenter gave feedback, pointing out that some of the non-words looked or sounded like real words. All participants were given 10 practice trials on the computer (with the same procedure as the lexical decision task trials).

### Lexical decision task

Adults completed 420 experimental trials, with five breaks. Children and adolescents completed 280 experimental trials, with four breaks. The difference in the number of trials was due to adults also seeing the medium N stimuli. The task lasted approximately 20–30 min.

### Neighbor knowledge test

Child and adolescent participants completed 160 experimental trials, with 1 break. The task lasted approximately 10 min. This task was always presented after the lexical decision task.

Participants were instructed that accuracy was more important than speed on this task. To reduce guessing, they were told to only answer "word" if they were sure that they had read the item before and knew what it meant.

### Psychometric testing

Adolescent participants completed the Vocabulary and Matrix Reasoning subsets of the Wechsler Abbreviated Scale of Intelligence (WASI, Wechsler, 1999) and Letter-Word ID, Word Attack, and Reading Fluency subtests of the Woodcock Johnson (WJ) Tests of Achievement (WJ III-R COG; Woodcock et al., 2001). Although adults and children did not complete this testing, some of their IQ and reading ability measures (calculated using the same assessments) were available from prior studies. We believe that the subset is representative of the entire sample, as there was no systematic variation on the experimental task between the participants for whom scores were and were not available.

## RESULTS

### NEIGHBOR KNOWLEDGE TEST

For child and adolescent participants, the effective N of the high N word stimuli was estimated from the neighbor knowledge test.

Each child was not tested on every neighbor; estimates of each child's effective N were generated using the sample of neighbors on which he/she was tested. First, each neighbor word was weighted by the number of word targets (henceforth, "points") from the lexical decision task for which it was a neighbor (e.g., "cases" was worth 5 points because it was a neighbor of 5 high N targets including "cages" and "bases"). A weighted estimate of the proportion of neighbors of the high N targets that each child or adolescent knew was computed to estimate their effective Ns (see Equation 1). First, we summed the points for each hit (i.e., each real word that the child identified as such). This sum was called the number of points earned. Then, we summed together the total possible points (# possible points). We multiplied the number of possible points by the false alarm rate to estimate the number of points the child earned through random guessing. We then subtracted this product from the number of points earned to calculate the number of points the child earned by knowing the vocabulary words, rather than by randomly guessing. We then divided this amount by the number of possible points to calculate the proportion of points the child earned by knowing the vocabulary words. This proportion was multiplied by the average N of the high N target words (13.06) to calculate the average number of neighbors that each child or adolescent knew. A weighted estimate was used because there was a great range in the number of targets (1–5) for which a given item was a neighbor. This weighted calculation allowed us to give more credit when known words were neighbors of multiple targets.

$$\frac{(\text{\# points earned}) - \left[(\text{false alarm rate}) * (\text{\# possible points})\right]}{(\text{\# possible points})}$$
(1)

On average, children knew 9.38 neighbors ($SD = 1.03$) of each target word and adolescents knew 9.98 neighbors ($SD = 1.20$) out of 13.06. Although this difference is small, the correlation with age was significant ($r = 0.36$, $p = 0.01$). Neighbor knowledge test scores strongly correlated with WASI raw vocabulary scores ($r = 0.62$, $p < 0.01$), but not WASI matrix reasoning raw scores ($r = 0.30$, $p = 0.07$)[2], suggesting that the neighbor knowledge test tapped into an aspect of children's general vocabulary knowledge.

The neighbor knowledge test confirmed that target words were fairly high N for the children and adolescents. Furthermore, the "matched N" list (where $N = 8.59$) shown to the adults closely approximated, or slightly underestimated, the effective N for the children (i.e., 9.38). It is preferable for the matched N condition to slightly underestimate the average effective N for the children, because it is over-correcting for most children and very closely matching the effective N for the youngest children [the average effective N of the three youngest children (mean age = 8.95 years) was 8.65]. The matched N condition was therefore used in subsequent analyses as a control for differences in effective N across development by determining whether similar results were seen for

_____

[2]These correlations are based on the data from all of the adolescents and 13 children. The scores from the 13 children included in the correlation analyses were obtained within a year of participation in the present study.

adults with the matched N list and children and adolescents with the high N list.

## LEXICAL DECISION TASK

The adults, adolescents, and children were 93, 90, and 82% correct on all trials respectively.

We conducted mixed linear analyses using the lme4 package (Bates et al., 2010). The analysis methodology replicated that of Andrews and Lo (2012). First, we filtered the responses to examine only correct responses to word targets. Then we calculated the mean and standard deviation RT for each participant. Outlier RTs more than two standard deviations from a participant's mean were removed from the analysis (see **Table 4**). The negative inverse RTs were calculated as visual inspection showed that this best approximated a normal distribution and this transformation was used in similar studies (Andrews and Lo, 2012). The analyses treated participants and targets as crossed random effects. We assessed the effects of target neighborhood size and prime type with two orthogonal normalized contrasts comparing (a) average priming (mean of repetition and form primes as compared to unrelated primes) and (b) form and repetition primes. A generalized matrix inversion was then conducted on the contrast weights to yield interpretable main effects. To facilitate comparison with previous evidence of form priming, a second set of models tested generalized matrix inverted normalized contrasts that separately compared the form and repetitions primes with the unrelated primes. Higher order interactions of these contrasts with neighborhood size were included as fixed effects. Since the *t*-values obtained using linear mixed effects models are not conventionally associated with degrees of freedom, Markov-chain Monte Carlo simulations with 10,000 simulations were used to obtain the associated *p*-values.

We were interested in testing the three-way interaction between target orthographic neighborhood size (categorical high/low), participant age (continuous), and prime type (using the contrasts described above). The most straightforward support of our main hypothesis would be a significant interaction between age, neighborhood size, and the contrast between form and unrelated primes. This finding would suggest that the extent of form priming (as compared against the unrelated baseline, the typical calculation) to high N words changes with age. However,

our hypothesis would also be supported if there were a significant interaction between age, neighborhood size, and the contrast between form and repetition primes. Repetition and form primes only differ by one letter. The significant interaction would suggest that the way participants respond to partially and fully matching primes preceding high N targets changes with age. We also regressed out factors which can affect reaction time: frequency, length, number of syllables, bigram mean, RT on the preceding trial, and accuracy on the preceding trial (see **Table 5**

**Table 5 | The coefficients and their significances in the model using the high N targets in adults.**

|  | *B* | Std. Error | *t* | *pMCMC* |
|---|---|---|---|---|
| Intercept | −1.623 | 0.0189 | −85.62 | 0.0001 |
| Log freq. | −0.0218 | 0.0033 | −6.62 | 0.0001 |
| Prev. RT | 0.1225 | 0.0084 | 14.54 | 0.0001 |
| Bigram mean | −0.0001 | 0.0001 | −1.42 | 0.1502 |
| Length | 0.0179 | 0.0145 | 1.24 | 0.1996 |
| No. of syllables | 0.0395 | 0.0222 | 1.78 | 0.0684 |
| Prev. accuracy | 0.0431 | 0.0088 | 4.90 | 0.0001 |
| Age | −0.0409 | 0.0040 | −10.17 | 0.0001 |
| N | −0.0287 | 0.0124 | −2.31 | 0.0160 |
| Age*N | −0.0010 | 0.0011 | 0.87 | 0.4120 |
| **UNRELATED/REPETITION & FORM CONTRAST; FORM/REPETITION CONTRAST** | | | | |
| U/R&F | −0.0759 | 0.0055 | −13.90 | 0.0001 |
| F/R | −0.0984 | 0.0062 | −15.77 | 0.0001 |
| Age* U/R&F | −0.0017 | 0.0012 | −1.35 | 0.1866 |
| Age* F/R | −0.0081 | 0.0014 | −5.81 | 0.0001 |
| N* U/R&F | −0.0333 | 0.0109 | 3.05 | 0.0018 |
| N* F/R | −0.0414 | 0.0125 | −3.31 | 0.0008 |
| Age*N* U/R&F | −0.0010 | 0.0024 | −0.42 | 0.6706 |
| Age*N*F/R | −0.0074 | 0.0028 | −2.65 | 0.0112 |
| **UNRELATED/FORM CONTRAST; UNRELATED/REPETITION CONTRAST** | | | | |
| U/F | −0.0267 | 0.0063 | −4.22 | 0.0001 |
| U/R | −0.1251 | 0.0063 | −20.02 | 0.0001 |
| Age*U/F | 0.0024 | 0.0014 | 1.70 | 0.0952 |
| Age*U/R | −0.0057 | 0.0014 | −4.09 | 0.0001 |
| N*U/F | 0.0540 | 0.0127 | 4.26 | 0.0001 |
| N*U/R | 0.0127 | 0.0125 | 1.01 | 0.3096 |
| Age*N*U/F | 0.0027 | 0.0028 | 0.95 | 0.3548 |
| Age*N*U/R | −0.0047 | 0.0028 | −1.69 | 0.0924 |
| **RANDOM EFFECTS** | | | | |
|  | **Variance** | **Stdev.** | | |
| Target | 0.0042 | 0.0650 | | |
| Ordering by participant | <0.0001 | 0.0003 | | |
| Participant | 0.0226 | 0.1504 | | |

*pMCMC is the p-value obtained using Markov-Chain Monte Carlo simulations. N is an abbreviation for orthographic neighborhood size. U, F, and R are abbreviations for unrelated, form, and repetition priming respectively. Therefore, U/F represents the contrast between the unrelated and repetition priming conditions and U/R & F represents the contrast between the unrelated condition and both the repetition and form priming conditions, etc. All continuous variables are centered. Previous accuracy is a categorical variable, with a correct response being the baseline. Orthographic neighborhood size is a categorical variable with 2 levels. A contrast code was used to compare orthographic neighborhood size.*

**Table 4 | Reaction time on lexical decision task.**

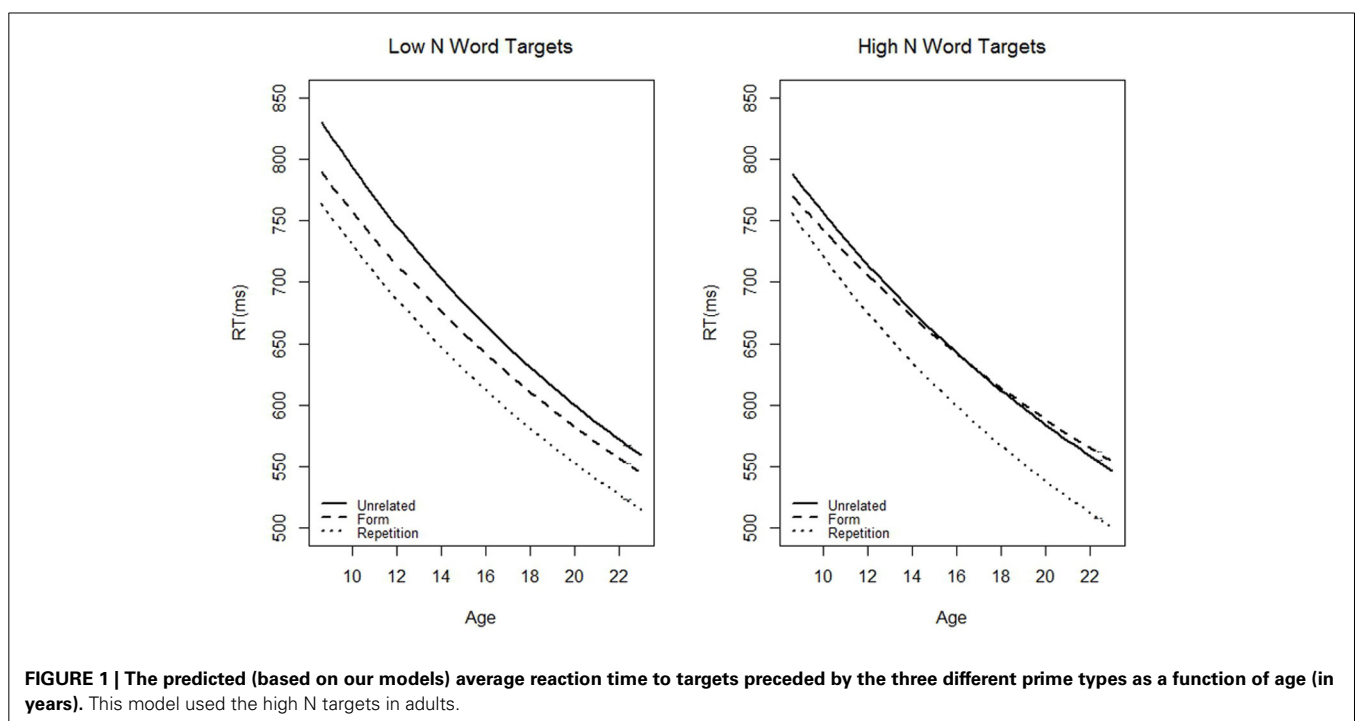|  |  | **Children** | **Adolescents** | **Adults** |
|---|---|---|---|---|
| Low N | Repetition | 799 (281) | 649 (157) | 551 (109) |
|  | Form | 836 (273) | 677 (160) | 582 (91) |
|  | Unrelated | 857 (244) | 700 (165) | 594 (91) |
| Matched N | Repetition | – | – | 543 (97) |
|  | Form | – | – | 583 (97) |
|  | Unrelated | – | – | 584 (89) |
| High N | Repetition | 799 (258) | 626 (153) | 535 (94) |
|  | Form | 822 (259) | 672 (166) | 582 (93) |
|  | Unrelated | 836 (268) | 665 (144) | 577 (89) |

*The trim reaction time (ms) to correct word targets displayed as mean (standard deviation).*

and **Figure 1**). As one can see from the figure, RT decreased with age. For Low N targets, the effects of the three prime conditions were relatively constant across age: repetition primes were more beneficial than form primes which were more beneficial than unrelated primes. However, for high N targets, the effects of the three priming conditions varied with age. Although repetition primes were always the most beneficial, the benefit derived from repetition primes (as compared to unrelated primes) increased with age. In contrast, the benefit derived from form primes (as compared to unrelated primes) decreased with age. In fact, for the oldest participants, form primes had a slight inhibitory effect.

Log frequency, orthographic neighborhood size, accuracy on the previous trial, and age were negatively correlated with RT, whereas RT on the previous trial was positively correlated with RT. Consistent with our key hypothesis, the three-way interaction between age, neighborhood size, and the contrast between form and repetition priming was significant, $t = -2.65$, $pMCMC = 0.01$. In the low N condition, form and repetition priming decreased slightly with age; in the high N condition, form priming greatly decreased with age whereas repetition priming increased with age (**Figure 1**). This interaction can be further unpacked by examining the predicted values. For low N targets, the amount of benefit derived from both repetition and form primes preceding low N targets decreased by about 20 ms between the ages of 9 and 22. After controlling for confounding variables, a hypothetical 9 year old (corresponding to the average age of the three youngest participants) would yield a 63.79 ms repetition priming effect and a 38.58 ms form priming effect; whereas a hypothetical 22 year old would yield a 44.25 ms repetition priming effect and a 15.04 ms form priming effect. A different pattern of results emerges for high N targets. For children, priming is

more beneficial in the low than high N condition. A hypothetical 9 year old would yield a 32.26 ms repetition priming effect and a 19.49 ms form priming effect. In adults, however, repetition priming is equally beneficial in the high N condition (45.54 ms). Furthermore, although form primes benefited children in the high N condition, they actually inhibited adults (−6.3 ms).

We repeated the analysis using the matched N words for the adults. If the models using the high and matched N words were similar, it would suggest that development differences in priming effects are not solely due to vocabulary acquisition, as the effective N is controlled for in the matched N model. Target-specific random effects were excluded from matched N model since the adults and children/adolescents saw different items. The pattern of results seen in the matched N and high N models were very similar (see **Table 6**). All significant effects replicated, save that the three-way interaction between age, neighborhood size, and the contrast between form and repetition priming was a trend $t = -1.78$, $pMCMC = 0.08$. We re-ran the analyses with a slightly different method of cleaning outliers; replacing outliers with a boundary value rather than replacing them (fence method). Using this method of data cleaning, the three-way interaction between age, neighborhood size, and the contrast between form and repetition priming was significant, $t = -2.33$, $pMCMC = 0.02$. Although the interaction was only significant using one of the methods of cleaning outliers, it is important to remember that we over-corrected for effective N in this analysis. Therefore, we were able to find marginally significant effects even when the stimuli the adults saw had fewer neighbors than the children's effective N. Presumably, a closer matching of N would yield significant results. The coefficients for length, number of syllables, and bigram frequency were significant in the matched



**FIGURE 1 | The predicted (based on our models) average reaction time to targets preceded by the three different prime types as a function of age (in years).** This model used the high N targets in adults.

**Table 6 | The coefficients and their significances in the model using the matched N targets in adults.**

| | Trim | | | | Fence | | | |
|---|---|---|---|---|---|---|---|---|
| | **B** | **Std. Error** | **t** | **pMCMC** | **B** | **Std. Error** | **t** | **pMCMC** |
| Intercept | −1.626 | 0.0184 | −88.33 | 0.0001 | −1.605 | 0.0184 | −87.17 | 0.0001 |
| Log Freq | −0.0230 | 0.0015 | −15.53 | 0.0001 | −0.0264 | 0.0016 | −16.87 | 0.0001 |
| Prev. RT | 0.1172 | 0.0087 | 13.55 | 0.0001 | 0.1178 | 0.0088 | 13.37 | 0.0001 |
| Bigram mean | −0.0001 | < 0.0001 | −2.36 | 0.0210 | 0.0001 | < 0.0001 | −2.39 | 0.0160 |
| Length | 0.0153 | 0.0063 | 2.41 | 0.0140 | 0.0203 | 0.0067 | 3.01 | 0.0026 |
| No. of Syllables | 0.0458 | 0.0095 | 4.84 | 0.0001 | 0.0523 | 0.0099 | 5.27 | 0.0001 |
| Prev. Accuracy | 0.0356 | 0.0089 | 3.98 | 0.0001 | 0.0505 | 0.0093 | 5.42 | 0.0001 |
| Age | −0.0406 | 0.0041 | −9.95 | 0.0001 | −0.0406 | 0.0041 | −9.96 | 0.0001 |
| N | −0.0190 | 0.0054 | −3.53 | 0.0004 | −0.0156 | 0.0057 | −2.74 | 0.0060 |
| Age*N | 0.0010 | 0.0012 | 0.81 | 0.4114 | 0.0021 | 0.0013 | 1.67 | 0.0918 |
| **UNRELATED/REPETITION & FORM CONTRAST; FORM/REPETITION CONTRAST** | | | | | | | | |
| U/R&F | −0.0771 | 0.0057 | −13.62 | 0.0001 | −0.0817 | 0.0060 | −13.71 | 0.0001 |
| F/R | −0.0926 | 0.0065 | −14.30 | 0.0001 | −0.0973 | 0.0068 | −14.23 | 0.0001 |
| Age* U/R&F | −0.0018 | 0.0013 | −1.39 | 0.1594 | −0.0028 | 0.0013 | −2.06 | 0.0380 |
| Age* F/R | −0.0073 | 0.0014 | −5.02 | 0.0001 | −0.0083 | 0.0015 | −5.39 | 0.0001 |
| N* U/R&F | 0.0260 | 0.0114 | 2.29 | 0.0214 | 0.0396 | 0.0119 | 3.32 | 0.0008 |
| N* F/R | −0.0295 | 0.0130 | −2.27 | 0.0230 | −0.0319 | 0.0137 | −2.33 | 0.0180 |
| Age*N* U/R&F | −0.0015 | 0.0025 | −0.58 | 0.5430 | 0.0011 | 0.0027 | 0.43 | 0.6694 |
| Age*N*F/R | −0.0051 | 0.0029 | −1.78 | 0.0754 | −0.0071 | 0.0031 | −2.33 | 0.0214 |
| **UNRELATED/FORM CONTRAST; UNRELATED/REPETITION CONTRAST** | | | | | | | | |
| U/F | −0.0308 | 0.0066 | −4.69 | 0.0001 | −0.0331 | 0.0069 | −4.79 | 0.0001 |
| U/R | −0.1234 | 0.0065 | −19.05 | 0.0001 | −0.1304 | 0.0068 | −19.06 | 0.0001 |
| Age*U/F | 0.0019 | 0.0015 | 1.28 | 0.1946 | 0.0014 | 0.0015 | 0.89 | 0.3690 |
| Age*U/R | −0.0054 | 0.0014 | −3.72 | 0.0004 | −0.0069 | 0.0015 | −4.49 | 0.0001 |
| N*U/F | 0.0407 | 0.0131 | 3.10 | 0.0014 | 0.0556 | 0.0148 | 4.02 | 0.0001 |
| N*U/R | 0.0112 | 0.0130 | 0.86 | 0.3868 | 0.0237 | 0.0137 | 1.73 | 0.0838 |
| Age*N*U/F | 0.0011 | 0.0029 | 0.37 | 0.7296 | 0.0047 | 0.0031 | 1.52 | 0.1304 |
| Age*N*U/R | −0.0040 | 0.0029 | −1.39 | 0.1646 | −0.0024 | 0.0031 | −0.79 | 0.4270 |
| **RANDOM EFFECTS** | | | | | | | | |
| | **Variance** | **Stdev.** | | | **Variance** | **Stdev.** | | |
| Ordering by participant | < 0.0001 | 0.0003 | | | < 0.0001 | 0.0004 | | |
| Participant | 0.0233 | 0.1526 | | | 0.0232 | 0.1524 | | |

*Abbreviations and conventions as in **Table 5**. Trim refers to the data cleaning method in which outliers are removed, whereas fence refers to the data cleaning method in which outliers are replaced with a boundary value.*

N model although they were not in the high N model, possibly due to the exclusion of target-specific random effects in the current model. The matched and high N model similarity can be discerned by comparing **Figures 1**, **2**. Inspection of the predicted values reveals that even when the effective neighborhood size was matched, a hypothetical 22 year old adult showed more repetition priming (44.77 ms) and less form priming (−0.76 ms, again revealing slight inhibition) than the hypothetical 9 year old child did (repetition priming: 34.52 ms; form priming: 16.88 ms).

Next, we tested whether the developmental trajectory was better explained by age or by the neighbor knowledge test. We restricted our test to children and adolescents because adults did not take the neighbor knowledge test. We used a linear mixed analysis, but instead of using Age as a factor in the three-way interaction, we used (Age + Neighbor Knowledge). This analysis was appropriate because the correlation between age and neighbor

knowledge ($r = 0.36$) is well below accepted cutoffs for collinearity. Since the three way interaction was of main interest, we only ran the first pair of contrasts (unrelated/repetition&form; repetition/form). The results are displayed in **Table 7** and **Figure 3**. In the interest of space, only factors involved in the three-way interaction are reported. The interaction between age, N, and the form/repetition priming contrast was significant, $t = -2.50$, $pMCMC = 0.01$. Furthermore, the interaction between Neighbor Knowledge, N, and the form/repetition priming contrast was non-significant, $t = 0.59$, $pMCMC = 0.57$. This analysis suggests that age, and not vocabulary, drives developmental differences in priming.

## DISCUSSION
Previous studies have shown that children are facilitated by both repetition and form primes preceding both low and high N targets. In contrast, adults do not show facilitation when form
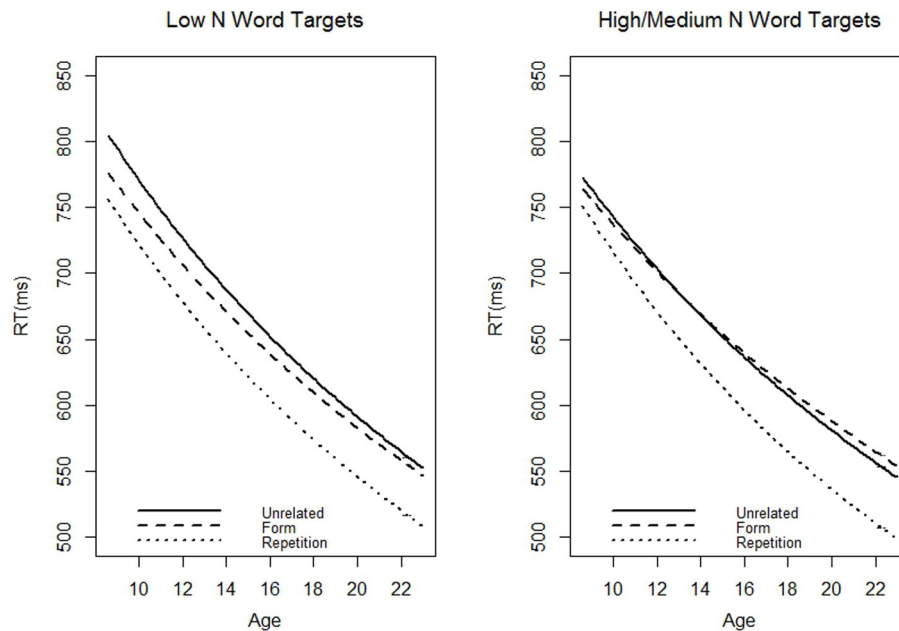
**FIGURE 2 | The predicted (based on our models) average reaction time to targets preceded by the three different prime types as a function of age (in years).** The trim method was used to clean outliers. This model used the medium N targets in adults. Note that the graph for the Low N words is slightly different than **Figure 1** because the item-specific random effects are not included in this model.

primes precede high N targets (Castles et al., 1999). However, it was unclear whether increases in written vocabulary size underlie these developmental changes. This study sought to replicate previous findings and test the hypothesis regarding vocabulary across a broader range of ages than had been previously studied. Our study replicated previous findings in that children were facilitated by repetition and form primes, but adults were facilitated in three conditions (high and low N repetition priming; low N form priming) but inhibited by form primes preceding high N targets. When we examined whether written vocabulary growth could explain this developmental differences, we found that it could not. Our models predicted developmental differences when controlling for effective N (using a matched N stimulus set). They also indicated that vocabulary size, measured using the neighbor knowledge test, could not predict priming effects.

Treating age as a continuous variable also allowed us to identify a previously unreported trend: the benefit derived from repetition primes preceding high N targets slightly increased over the course of development. Although previous developmental studies have not shown changes in repetition priming with age (Castles et al., 1999), there is evidence that more skilled adults (as measured by faster RTs in a lexical decision task) showed more repetition priming than low skilled adults (Kliegl et al., 2010). Since our adults responded much faster than our children, our results nicely dovetail with these findings.

Since written vocabulary does not seem to be related to developmental differences in priming, another mechanism must be at play. Andrews and Hersch (2010) identified a candidate mechanism: lexical precision. In an adult study, they found that spelling skill, but not written vocabulary size, was able to predict individual differences in masked form priming. Poor spellers were facilitated by form primes preceding high N words, whereas good spellers were slightly inhibited. Since spelling ability is a measure of orthographic precision, these results suggest that it is differences in lexical precision, rather than the number of neighbors known (i.e., written vocabulary size), which determine form priming effects. Although these effects were reported with adult participants, it is possible that a similar mechanism underlies developmental differences in priming. Children may show more facilitation due to form primes preceding high N targets because their orthographic representations are less precise.

Precise representations are fully specified so that a written word can fully determine the lexical representation to be activated, and this lexical representation can be quickly activated with minimal activation of its neighbors. Let us quickly summarize how an increase in the precision of lexical entries accounts for both our expected and rather surprising findings, before discussing the mechanism by which the lexical entry achieves this precision. The first finding is that adults derive equal benefit from repetition primes preceding both low and high N targets, whereas children derive more benefit in the low N condition. When a person with high quality lexical representations (presumably an adult) sees a repetition prime, the prime will quickly and correctly activate its corresponding lexical entry and nothing else. The correct activation of its lexical entry will make response time to the target faster. Therefore, adults will display equal repetition priming to low and high N targets. When a person with lower quality lexical representations (presumably a child) sees a repetition prime, it will activate its corresponding lexical representation and the lexical representations of its neighbors (if any).

**Table 7 | The coefficients and their significances in the model that tested the predictive power of effective N.**

|                          | B       | Std. Error | t      | pMCMC  |
|--------------------------|---------|------------|--------|--------|
| Intercept                | −1.661  | 0.0440     | −37.76 | 0.0001 |
| Age                      | −0.0475 | 0.0098     | −4.86  | 0.0001 |
| Vocab                    | −0.0088 | 0.0224     | −0.39  | 0.6426 |
| N                        | −0.0428 | 0.0164     | −2.60  | 0.0082 |
| U/R&F                    | −0.0804 | 0.0124     | −6.46  | 0.0001 |
| F/R                      | −0.1029 | 0.0142     | −7.23  | 0.0001 |
| Age*N                    | −0.0032 | 0.0026     | −1.20  | 0.2342 |
| Vocab*N                  | 0.0158  | 0.0060     | 2.62   | 0.0092 |
| Age* U/R&F               | −0.0026 | 0.0028     | −0.94  | 0.3528 |
| Age*F/R                  | −0.0094 | 0.0032     | −2.92  | 0.0046 |
| Vocab* U/R&F             | 0.0018  | 0.0064     | 0.29   | 0.7778 |
| Vocab*F/R                | 0.0040  | 0.0074     | 0.54   | 0.6016 |
| N* U/R&F                 | 0.0300  | 0.0250     | 1.20   | 0.2316 |
| N*F/R                    | −0.0846 | 0.0285     | −2.96  | 0.0032 |
| Age* N*U/R&F             | 0.0017  | 0.0056     | −0.29  | 0.7774 |
| Age*N*F/R                | −0.0162 | 0.0065     | −2.50  | 0.0102 |
| Vocab* N*U/R&F           | 0.0049  | 0.0129     | 0.38   | 0.6946 |
| Vocab*N*F/R              | 0.0088  | 0.0149     | 0.59   | 0.5732 |

|                          | Variance  | Stdev. |
|--------------------------|-----------|--------|
| Target                   | 0.0044    | 0.0662 |
| Ordering by participant  | <0.0001   | 0.0004 |
| Participant              | 0.0262    | 0.1617 |

*This model only included children and adolescents. Although not shown in the table, the length, frequency, bigram mean, and the number of syllables in the target were controlled for in the model, as was the participants' accuracy and RT on the preceding trial. Here, "vocab" refers to the effective N as calculated by the neighbor knowledge test. Abbreviations and conventions as in **Table 5**.*

If the neighbors are slightly activated, they may weakly inhibit the correct lexical representation. Therefore, if the prime has no neighbors, the repetition prime will be more beneficial than if the prime has many neighbors. Therefore, children will display more repetition priming to low than high N targets.

This mechanism can also explain why children are facilitated by form primes preceding high N targets but adults are not. *Target preactivation* occurs because the prime and target share many of the same letters, so the letter level activates the target. This facilitation can be counteracted by *target neighbor suppression* due to lateral inhibition between orthographic neighbors at the word level. Thus, a form prime will activate all of its neighbors via the *target preactivation effect*. If the target has many neighbors, as in the high N condition, the target word and many of its neighbors will be activated. As people with high quality lexical representations are assumed to have more lateral inhibition, the target word would be strongly inhibited by its neighbors and the *target neighbor suppression* would override all facilitation from the *target preactivation effect*. In contrast, people with low quality lexical representations have less lateral inhibition, so the *target preactivation effect* would remain stronger than the *target neighbor suppression*. Note that in both cases, vocabulary is equated: people with low and high quality lexical representations know the same number of neighbors of a given target word. But, the lateral inhibition from a given neighbor is stronger in people with high quality lexical representations. Of course, the above argument is purely speculative as we, unfortunately, did not acquire measures of spelling ability (i.e., lexical precision). Nonetheless, the present results do not support the written vocabulary hypothesis. The strongest alternative explanation is that changes in lexical precision underlie
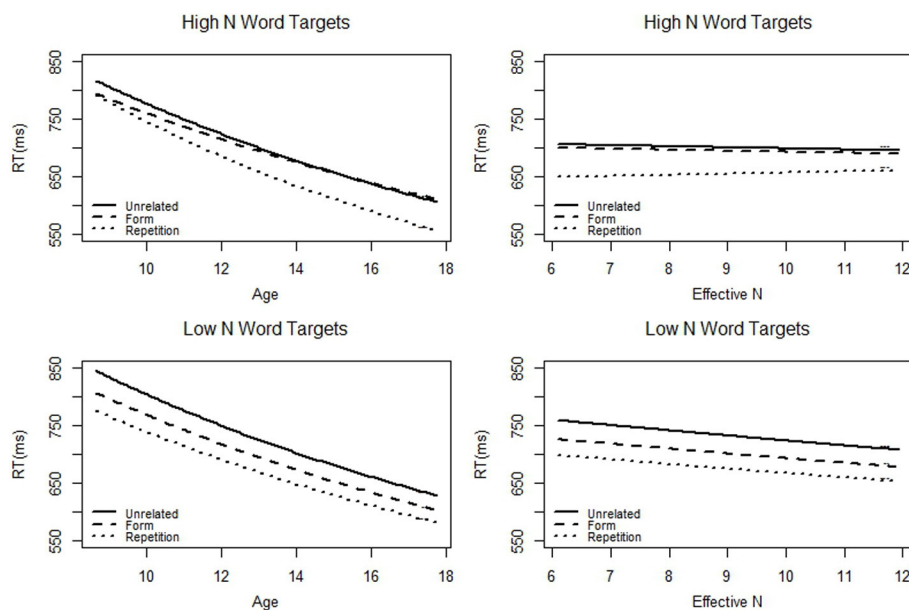


**FIGURE 3 | The predicted (based on our models) average reaction time to targets preceded by the three different prime types as a function of age (in years) or effective N (in number of neighbors known) for just children and adolescents.** Effective N was calculated using the neighbor knowledge test.

developmental differences in form priming. Reading experience cannot explain the differential priming effects in adults with varying spelling abilities, because Andrews and Hersch (2010) found an effect of spelling ability while controlling for reading experience. However, the strength of reading experience as a predictive variable could be moderated by age, specifically, it may wane over development. It is possible that reading experience could be responsible for the results found in this study for children. Alternatively, writing experience, where children have to not only recognize, but also produce, the correct spelling could underlie these developmental changes. Future studies which directly correlate spelling ability and priming across the developmental spectrum are needed.

Before concluding, we acknowledge additional limitations of the present study. We restricted our target word stimuli to higher frequency, shorter words. It is unknown whether neighborhood effects on the development of form priming would persist across different word types. Second, to allow for a close comparison to previous developmental studies, we approximated as closely as possible the experimental timing used by Castles et al. (1999). However, adult form priming appears sensitive to subtle variations in experimental timing (Ferrand and Grainger, 1994). It is unknown if children display more adult-like patterns at longer prime durations. An additional concern is that the lexical decision task elicited large developmental differences in response time. Prior studies have demonstrated that apparent developmental differences in letter processing are reaction time dependent (Lachmann and van Leeuwen, 2008). However, after accounting for RT in a mixed linear effects model, age was still a significant predictor ($t = -3.21$, $pMCMC < 0.01$). In addition, it is unlikely that the developmental differences in RT reflect a difference in speed/accuracy trade-offs across age, as the children were both slower and more inaccurate than the adults.

Our results suggest that age-related factors beyond written vocabulary size underlie the developmental differences in high N form priming. Future studies may benefit from using designs that more closely match children's effective N and examining other individual differences (e.g., spelling ability) to pinpoint specific mechanisms that lead to developmental changes in the precision of lexical representations. Understanding why children are differentially affected by orthographic neighborhood size is crucial to understanding how children learn to distinguish between words with similar spellings, and why some children are not able to do so even after adequate instruction.

## AUTHORS' NOTE

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fpsyg.2014.00667/abstract

## REFERENCES

Andrews, S., and Hersch, J. (2010). Lexical precision in skilled readers: individual differences in masked neighbor priming. *J. Exp. Psychol. Gen.* 139, 299–318. doi: 10.1037/a0018366

Andrews, S., and Lo, S. (2012). Not all skilled readers have cracked to code: individual differences in masked form priming. *J. Exp. Psychol. Learn. Mem. Cogn.* 38, 152–163. doi: 10.1037/a0024953

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., et al. (2007). The English lexicon project. *Behav. Res. Methods* 39, 445–459. doi: 10.3758/BF03193014

Bates, D. M., Maechler, M., and Dai, B. (2010). *Lmer4: Linear Mixed-Effect Models Using S4 Classes. R Package Version 0.999375-34. [Software].* Vienna: R Foundation for Statistical Computing.

Castles, A., Davis, C., Cavalot, P., and Forster, K. I. (2007). Tracking the acquisition of orthographic skills in developing readers: masked priming effects. *J. Exp. Child Psychol.* 97, 165–182. doi: 10.1016/j.jecp.2007.01.006

Castles, A., Davis, C., and Letcher, T. (1999). Neighborhood effects on masked form priming in developing readers. *Lang. Cogn. Process.* 14, 201–224. doi: 10.1080/016909699386347

Castles, A. E., Davis, C., and Forster, K. I. (2003). "Word recognition development in children: insight from masked priming," in *Masked Priming: The State of the Art,* eds S. Kinoshita and S. J. Lupker (New York, NY: Psychology Press), 345–361.

Ferrand, L., and Grainger, J. (1994). Effects of orthography are independent of phonology in masked form priming. *Q. J. Exp. Psychol. A* 47, 365–382. doi: 10.1080/14640749408401116

Forster, K. I. (1987). "Form-priming with masked primes: the best match hypothesis," in *Attention and Performance XII,* ed M. Coltheart (Hillsdale, NJ: Erlbaum), 127–146.

Forster, K. I. (1989). "Basic issues in lexical processing," in *Lexical Representation and Process,* ed W. Marslen-Wilson (Cambridge, MA: MIT Press), 75–107.

Forster, K. I., and Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *J. Exp. Psychol. Learn. Mem. Cogn.* 10, 680–698. doi: 10.1037/0278-7393.10.4.680

Forster, K. I., and Davis, C. (1991). The density constraint on form-priming in the naming task: interference effects from a masked prime. *J. Mem. Lang.* 30, 1–25. doi: 10.1016/0749-596X(91)90008-8

Forster, K. I., Davis, C., Schoknecht, C., and Carter, R. (1987). Masked priming with graphemically related forms: repetition or partial activation? *Q. J. Exp. Psychol.* 39, 211–251. doi: 10.1080/14640748708401785

Forster, K. I., and Taft, M. (1994). Bodies, antibodies, and neighborhood-density effects in masked form priming. *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 844–863. doi: 10.1037/0278-7393.20.4.844

Kliegl, R., Masson, M. E. J., and Richter, E. M. (2010). A linear mixed model analysis of masked repetition priming. *Vis. Cogn.* 18, 655–681. doi: 10.1080/13506280902986058

Lachmann, T., and van Leeuwen, C. (2008). Different letter-processing strategies in diagnostic subgroups of developmental dyslexia. *Cogn. Neuropsychol.* 2, 730–744. doi: 10.1080/02643290802309514

Lund, K., Burgess, C., and Atchley, R. A. (1995). "Semantic and associative priming in high-dimensional semantic space," in *Proceedings of the 17th Annual Meeting of the Cognitive Science Society,* eds J. D. Moore and J. F. Lehman (Pittsburgh, PA: Erlbaum), 660–665.

Rumelhart, D. E., and McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: part II. The contextual enhancement

effect and some tests and extensions of the model. *Psychol. Rev.* 89, 60–94. doi: 10.1037/0033-295X.89.1.60

Schmidt, T. (2013). Invisible stimuli, implicit thresholds: why "not-seen" judgments cannot be interpreted in isolation. arXiv:1306.0756v3 [q-bio.NC], 1–18.

Segui, J., and Grainger, J. (1990). Priming word recognition with ortho-graphic neighbors: effects of relative prime-target frequency. *J. Exp. Psychol. Hum. Percept. Perform.* 16, 65–76. doi: 10.1037/0096-1523. 16.1.65

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence.* San Antonio, TX: The Psychological Corporation.

Woodcock, R. W., McGrew, K. S., and Mather, N. (2001). *Woodcock-Johnson III.* Itasca, IL: Riverside Publishing.

Zeno, S. M., Ivens, S. H., Millard, R. T., and Duvvuri, R. (1995). *The Educator's Word Frequency Guide.* Brewster, NJ: Touchstone Applied Science Associates, Inc.

**Conflict of Interest Statement:** The authors declare that the research was con-ducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

# The write way to spell: printing vs. typing effects on orthographic learning

## Gene Ouellette[1]* and Talisa Tims[2]

[1] Department of Psychology, Mount Allison University, Sackville, NB, Canada
[2] Department of Psychology, Dalhousie University, Halifax, NS, Canada

Prior research has shown superior orthographic learning resulting from spelling practice relative to repeated reading. One mechanism proposed to underlie this advantage of spelling in establishing detailed orthographic representations in memory is the motoric component of the manual movements evoked in printing or writing. This study investigated this contention directly by testing the effects of typing vs. printing on the orthographic learning achieved through spelling practice, and further evaluated whether practice modality interacts with pre-existing individual characteristics. Forty students in grade 2 (mean age 7 years 5 months) were introduced to 10 novel non-words. Some of the students practiced spelling the items by printing, while the others practiced spelling them on a keyboard. Participants were tested for recognition and spelling of these items 1 and 7 days later. Results revealed high rates of orthographic learning with no main effects of practice modality, testing time, or post-test modality. Hierarchical regression analyses revealed an interaction between typing proficiency and practice modality, such that pre-existing keyboarding skills constrained or facilitated learning within the typing-practice group. A similar interaction was not found between printing skills and learning within the printing group. Results are discussed with reference to both prominent reading theory and educational applications.

**Keywords: orthographic learning, lexical representations, self-teaching, spelling, reading, literacy, printing, typing**

## INTRODUCTION

When children enter into the often arduous task of mastering early literacy skills, they begin applying their knowledge of the alphabet by mapping speech sounds onto letters. Such letter-sound associations underlie early literacy as children sound out words in learning to read, and conversely analyze the sounds in words to create a spelling attempt. But as children progress on the pathway to literacy, they become better able to recognize words fluently with far less apparent effort and to spell words correctly by conventional standards. There thus appears to be a transition in literacy acquisition from a reliance on more laborious phonologically based sounding out strategies to the use of memory representations for longer letter patterns and entire words (Ehri, 2005). These memory representations are referred to as orthographic representations and the process of storing such representations as orthographic learning. There is now considerable evidence that orthographic representations are stored as a result of print exposure during decoding practice, resulting in a growing corpus of representations to be used in subsequent reading and writing activity (e.g., Share, 2004; Castles and Nation, 2006). More recently, spelling practice has been found to result in superior orthographic learning, relative to print exposure through reading alone (Conrad, 2008; Ouellette, 2010), although the reason for this has not been established. The present study evaluates the role of one component of the spelling process often hypothesized to underlie this advantage, i.e., the manual movements

involved in printing, while also evaluating the effectiveness of computer keyboarding for learning new orthographic representations. In the process, we further consider whether any modality effects interact with individual characteristics to support or constrain orthographic learning.

This research draws on two presently distinct bodies of literature: one dealing with orthographic learning and the other with effects of printing vs. keyboarding in establishing lexical representations in long term memory. Indeed one goal of this research is to bring these two areas of study together, in evaluating the role of the modality used in spelling practice and how this may interact with individual learner characteristics, when it comes to learning new word representations. To the best of our knowledge this has yet to be specifically tested within an experimental orthographic learning paradigm, yet is especially important when one considers the prominence prescribed to orthographic learning in developmental literacy theory, the possible involvement of motor commands/patterns in establishing lexical representations, the increasing use of computers within the home and classroom, and the diversity seen across learners.

## ORTHOGRAPHIC LEARNING AND SPELLING

As just outlined, orthographic learning allows a student to progress from less efficient sounding-out strategies to the use of word-specific memory representations. Much developmental theory explains fluent reading and accurate spelling as hinging

upon such representations. Ehri (2005) has provided a now well cited descriptive theory for instance, that depicts the beginning reader/speller as one who progresses through phases of proficiency related to their developing alphabetic and phonological knowledge. Through experience with print, longer and longer letter strings become stored in memory. Children in the final "consolidated alphabetic phase" are able to read fluently and to spell accurately, by relying upon these stored orthographic representations.

According to Ehri (2005), orthographic learning comes about through experience with printed language. The importance of sounding out words or phonological recoding in orthographic learning is further detailed in Share's (1995, 2004) self-teaching hypothesis. Share proposes that it is the process of applying letter-sound knowledge in decoding printed text that allows the reader to store longer and more detailed representations for encountered words. It is through decoding that children in essence, teach themselves word-specific representations, which are then available for future encounters with these now learned words. As in Perfetti and Hart's (2002) lexical quality hypothesis, proficient reading and spelling are seen to rely upon such highly refined lexical representations. While Share, along with Perfetti and Hart, posits an item-based theory rather than one of developmental phases, the focus on orthographic learning is the same. There have been a number of recent studies in support of Share's hypothesis, showing that elementary school aged children learn word specific orthographic representations from reading both contextual passages and isolated words (e.g., Nation et al., 2007; Ouellette and Fraser, 2009; Wang et al., 2011). Much of this research has employed an orthographic learning paradigm modeled after Share's (1995, 1999) earlier work. The basic paradigm involves exposing children to ambiguously spelled non-words. The use of non-words controls for effects of previous exposure, and the spellings are ambiguous in the sense that the pronunciation could be matched to more than one possible spelling (e.g., *yait* which could conceivably also be spelled as *yate*). Following a series of practice trials in which these non-words are read, participants are tested for spelling and/or recall with a forced choice task where one of the choices is a homophone foil (i.e., the alternate yet plausible spelling). If a participant persists with a phonologically based approach, they would be as likely to spell or identify the homophone as they would the practiced item. Success on these post-tests is thus seen as reflecting orthographic learning, as accurate identification or spelling reflects a newly stored memory representation.

Following the lead of Shahar-Yames and Share (2008), Ouellette (2010) modified the orthographic learning paradigm just described to replace the repeated readings for some students with spelling practice. English-speaking students in grade 2 were randomly assigned to either a traditional orthographic learning condition involving reading or to one where the reading was replaced with repeated spelling to dictation. The auditory and visual exposures to the non-words were carefully controlled across conditions, yet Ouellette found that the children in the spelling practice group outperformed the other students on post-tests administered 1 and 7 days after the practice session. Similar results have been reported for Hebrew speaking grade 3 students

(Shahar-Yames and Share). It thus appears that spelling practice provides a superior milieu for orthographic learning relative to print exposure garnered through reading alone. This contention is further supported by Conrad (2008) who directly compared reading and spelling practice and the transfer between the two skills with grade 2 students. Employing a list of real words with shared orthographic rime units, Conrad reported that representations learned through one skill transferred to the other, as the students were better able to spell words they had practiced reading, and to read words they had practiced spelling. Importantly, the transfer was greater from spelling practice to reading (than from reading practice to spelling). There was also transfer within each skill to untrained words, but this was again greater for spelling than for reading.

The research just reviewed points to superior orthographic learning through spelling compared to reading practice, further establishing the relevance and importance of spelling practice in establishing lexical representations for use in subsequent literacy tasks. What remains uncertain is the mechanism behind this effectiveness of spelling practice. In comparing the exercise of spelling to that of reading, one salient difference is the motoric component of the manual movement involved in writing out words. There has long been a notion that there is something special about the manual movements involved in printing or writing by hand that aids in memory encoding and/or retrieval, suggesting a possible motoric component to lexical representations (Masterson and Apel, 2006). Indeed, multi-sensory teaching approaches are very much based on this premise (Hulme, 1983; Hulme and Bradley, 1984).

## PRINTING BY HAND vs. KEYBOARDING IN SPELLING AND LITERACY LEARNING

The possible role of a motoric component in establishing representations for literacy has been directly tested in the past in a small number of studies that have compared the effects of printing vs. keyboarding on specific learning outcomes. Research with children just learning the alphabet for example, has shown superior letter learning following printing practice relative to keyboarding (Longcamp et al., 2005, 2008). This has led Longcamp and colleagues to propose that memory representations of letters incorporate visual and motor information across a complex neural network (Longcamp et al., 2005, 2008). Indeed, similar brain regions, specifically Broca's and areas of bilateral inferior parietal lobes, have been implicated in both printing by hand and visual letter recognition (Longcamp et al., 2005). It remains uncertain whether a similar role of motoric knowledge exists for longer, more refined lexical representations as there have been few studies that have focused on modality effects for learning more complex orthographic representations; we next turn our attention to this limited extant literature.

The seminal study directly comparing effects of printing and keyboarding on learning to spell and read was reported by Cunningham and Stanovich (1990), as motivated by the earlier work of Hulme and Bradley (1984). Cunningham and Stanovich gave grade 1 students practice spelling 30 words over 4 consecutive days. The words were randomly split into two lists of 15 words, and each participant practiced each list on 2 days of

the week (1 list Monday and Wednesday, the other Tuesday and Thursday). Post-testing on reading and spelling was completed on the Friday. Spelling modality was manipulated within subjects, as five words on each list were spelled by printing, 5 by typing on a keyboard, and 5 by arranging letter tiles. Post-tests revealed superior spelling accuracy for words practiced through printing, although there was no effect of practice modality on reading accuracy. It should be noted that the words used in this study varied in terms of sound-letter consistency; most could be read by sounding out and blending and at least some of the words could be spelled accurately by sounding out rather than relying on orthographic representations (e.g., man, help). The participants were also only in grade 1, at an age when phonological strategies may be more appropriate and spelling skills unstable (Masterson and Apel, 2006). Further, the use of real words raises concerns about previous exposure and pre-existing orthographic knowledge. Thus it is not clear as to whether these oft cited results can be confidently interpreted with respect to modality effects on orthographic learning.

To address the issue of previous exposure and pre-existing orthographic representations, Vaughn et al. (1992) replicated Cunningham and Stanovich's (1990) study with first graders, but this time the children were pre-tested on the word lists and any known words were discarded. In all, 21 of the original 30 stimulus words had to be replaced as they were spelled correctly at pre-test by at least one student. In contrast to the findings from the original study, Vaughn et al. found no significant differences in learning across the spelling modality conditions, leading them to conclude that printing by hand was not a superior milieu for learning to read and spell. Adding further support to this conclusion, Vaughn and colleagues completed another replication, adding individualized feedback to increase learning, and reported the same null results (Vaughn et al., 1993).

Although the research conducted by Vaughn et al. (1992, 1993) controlled for pre-existing word knowledge, the methodology still suffered from the same limitations raised earlier for Cunningham and Stanovich's (1990) original study. In particular, the variability in word consistency makes the results difficult to interpret with respect to the important developmental skill of orthographic learning. Further, all of these studies have taught quite a large corpus of words, allowing only two practice trials per word, to young grade 1 students. As result, learning rates were quite low across studies. In addition, it is important to note that in the methodology detailed by Vaughn et al., students copied the spellings rather than deriving them from memory. This is important as the benefits of spelling practice have been proposed to be related to the process of analyzing a word and retrieving information from memory in generating the spelling (Ouellette and Sénéchal, 2008; Sénéchal et al., 2012); copying may not provide for the same deep level of processing and this could have also contributed to the low learning rates and null results reported in this research. When these concerns are taken together, the present literature cannot establish whether there are modality effects on orthographic learning beyond single letter learning, and hence the mechanisms that underlie the effectiveness of spelling practice in learning word forms remain elusive.

## THE PRESENT STUDY AND THE ROLE OF INDIVIDUAL DIFFERENCES

In considering the limited research comparing the effects of printing to keyboarding in learning longer orthographic representations for spelling and reading, the results are clearly equivocal. Together this literature paints an unclear picture and most importantly, methodological concerns prevent the interpretation of results with respect to the role of spelling modality in orthographic learning, an issue of significant theoretical and practical significance. The present study aims to address this issue by directly comparing the effects of spelling practice through printing and keyboarding, within a carefully designed orthographic learning paradigm as described earlier and adapted to include spelling practice (as per Shahar-Yames and Share, 2008; Ouellette, 2010). The use of ambiguously spelled non-words allows for the specific evaluation of orthographic learning while also controlling for previous exposure, and the present study involves a sample of students in Grade 2, a grade level where orthographic learning would be especially relevant in making the transition to more fluent reading and accurate spelling (Conrad, 2008; Ouellette, 2010).

The present study also incorporates a number of other important methodological improvements over the extant literature. Primarily, the word set is restricted to 10 items, and each is practiced four times, allowing more opportunity for orthographic learning to occur than what has been reported in the past. Further, the spelling practice implemented requires the participant to spell to dictation after being exposed to the correct spellings rather than just copying the items, thus providing for a more analytic process and potentially deeper level of processing. We also incorporate a counterbalanced design with respect to the modality used in post-testing. Although research with older students has shown performance on spelling assessment not to be affected by the modality used (printing or keyboarding) in the administration of the test (Masterson and Apel, 2006), it is important in research that post-test methodology not resemble one training condition more so than another. Therefore, half the non-words practiced are assessed at post-test in the same modality as practiced, while the others are assessed in the opposite modality.

The present research also includes a pre-test battery to assess baseline levels of printing, typing, reading and spelling proficiency, allowing for the evaluation of the effects of pre-existing skills in these areas on orthographic learning. Of particular interest in the present study are possible interactions between individual differences in terms of pre-existing skills and the modality used to practice the new spellings. In other words, do particular students benefit more from practice in one modality over the other (or conversely, are some hindered within one modality more than the other)? One area that may be hypothesized to interact with the practice modality used for spelling is printing and typing ability. In the research reviewed earlier that compared printing with typing, children's baseline skills for printing and typing were not assessed. Yet it may be reasonable to hypothesize that success with learning through practice in either modality may depend at least in part upon the skills that children bring with them into a study, especially considering that

printing and typing skills appear to develop independently and there is considerable variability in these skills across children (Berninger et al., 2006).

While typing proficiency has yet to be examined with respect to its impact on early spelling practice, some have proposed slow or laborious printing to limit written composition and spelling by tapping cognitive resources and straining working memory; indeed, printing fluency has been shown to be correlated with spelling in the early grades (Kim et al., 2014), although whether this impacts orthographic learning specifically is not certain. In a study evaluating modality effects on spelling for students with spelling disabilities, Berninger et al. (1998), in accord with the studies of Vaughn et al. (1992, 1993), reported an overall null result in comparing effects of printing and keyboarding in spelling instruction. Interestingly, these researchers also assessed printing skills to evaluate a possible interaction between printing proficiency and practice modality, yet did not find any such interaction within their data. Berninger et al. did not assess keyboarding skills however, and it is not certain if their results apply to a general population of early learners. Further, many of the concerns surrounding word consistency and familiarity raised previously apply to the Berninger et al. research as well. Accordingly, it remains uncertain if pre-existing printing and typing skills do indeed interact with practice modality when it comes to orthographic learning.

The present study has been designed to address a prominent gap in current research. The information garnered here stands to add to current theory and to inform teaching practice. The topic of study is of special relevance given the advancement of computer technology and applications into the home and classroom (see Blok et al., 2002) and the prominence of the self-teaching hypothesis (Share, 2004).

## METHODS

### PARTICIPANTS

Forty-four Grade 2 students from an elementary school in a small Canadian town participated. Three children were absent on the day of the first post-test and were therefore excluded from the final sample. One student was identified as both a univariate ($z$-scores $>3.0$) and multivariate (through scatterplots) outlier on a number of pre-test measures and was also excluded from the final sample. Thus, a total of 40 children (18 males and 22 females) with a mean age of 7.42 years ($SD = 0.26$) were included in the analysis reported here. Of these children, 27.5% had a parent with a post-graduate degree, 32.5% had a parent with an undergraduate university degree, 17.5% had a parent who had attended college, 20% had a parent whose highest level of education was high school, and 2.5% had a parent who had not completed high school. All participants were English speaking with no history of speech, language, or learning difficulties.

### MATERIALS: INITIAL ASSESSMENT

#### Word reading and decoding

The Test of Word Reading Efficiency (TOWRE; Torgesen et al., 1999) was administered as a measure of reading skills. This is a timed reading test in which participants have 45 s to read a list of words and receive a score based on how many words are read

correctly. The test is repeated using a list of non-words. Many forms of reliability are reported, all of which are at or above 0.90.

#### Spelling

The Woodcock Johnson III (WJ-III; Woodcock et al., 2001) spelling subtest was administered. Children were asked to spell letters and words that increased in difficulty. Testing continued until six consecutive errors were made or until the participant reached item 59. Many forms of reliability are reported for this test, with a median of 0.90.

#### Baseline printing and typing skills

To obtain baselines of printing and typing proficiency, non-standardized tests based on previous research were administered (Berninger et al., 2006; Masterson and Apel, 2006; Kim et al., 2014). Children were asked to copy the passage "Are you amazed at how much you have learned so far? Just how high to build your speed is the next question," by typing and by printing. This passage is often used to assess typing as it contains nearly all letters on the keyboard. Children were also asked to produce both capital and lowercase letters of the alphabet in order by typing and by printing (e.g., Aa, Bb, Cc). In all tasks children were given 60 s to complete the test. These tasks were scored by counting the number of correctly produced characters to achieve a characters-per-minute score to reflect automaticity and proficiency in these areas. For the printing tasks, the letters had to be identifiable out of context and only reversals that could not be confused with other letters were accepted as correct. Inter-rater reliability was excellent (0.97).

### MATERIALS: TRAINING STIMULI

Participants were trained on 10 non-words used in previous research (see Bowey and Miller, 2007; Ouellette, 2010). These 10 non-words are ambiguous such that there is more than one possible spelling using a phonetic approach (see the Appendix).

### PROCEDURE

Children were first administered the pre-tests to obtain information about their skills prior to the study. This was done in a quiet, empty room by a trained research assistant. Children were administered the tests in one individual session, in the following order: typing passage, typing alphabet, TOWRE: Words, TOWRE: Non-words, WJ-III: Spelling, printing passage, printing alphabet. Half of the participants had the reverse order of the typing and printing tasks (i.e., the two printing tests at the start of the session and the typing tasks at the end).

Following completion of all individual assessments, each child received a training session in which they practiced spelling the 10 non-words in their assigned practice modality (printing or typing). Modalities were randomly assigned within each classroom, such that half the children from each class were assigned to each of the two conditions. The 10 non-words were typed on index cards and presented one at a time at the start of the practice session for the child to read aloud (visual exposure). Each card was in view for 5 s and any errors were corrected with a model to repeat. Once all the words had been read, the practice trials began. The index cards were shuffled and the child was once again shown a card and asked to read the non-word aloud. The card was removed from

view after 5 s; following a 5 s pause, the child was asked to spell the same non-word in a dictation (i.e., following a pronunciation by the researcher). Children in the printing condition spelled the non-word with a pencil on a blank index card; those in the typing condition used a standard PC keyboard. Typed spellings were displayed on the monitor within Microsoft Word, with the font set at 24 point (Arial) to make the character size approximately equivalent to those produced in the printing group. If the item was spelled correctly, the child was asked to read it aloud once more and then the spelling was immediately removed from view (the card flipped over for the printers and the computer screen cleared for the typers). If the child's spelling was incorrect, the original card was shown for the child to read. Regardless, the child was then asked to spell the word a second time. Once more, they read their spelling (if correct) or the original stimulus card (if incorrect), and the spellings were removed from view immediately. In all, children saw, read, heard, spelled, read, heard, spelled, and read each item on each trial. This procedure was followed until the entire deck of index cards was completed twice. It may be most accurate to describe this practice as spelling *plus* reading rather than as just spelling. Separating spelling from reading would jeopardize ecological validity (Conrad, 2008; Shahar-Yames and Share, 2008), and thus the spelling practice here deliberately incorporated reading as would be the natural occurrence.

All children were individually tested both one and 7 days later with a multiple-choice identification test and a spelling to dictation test. The multiple-choice test involved 10 items, one for each non-word, which included four different choices. The choices included the target and a (pseudo)homophone, as well as two other choices that were visually similar and/or contained the same letters but in a different order. The child was instructed to circle the correct spelling of the target word for each of the ten items. In the spelling test, the researcher simply dictated the target non-words for the child to spell. No feedback was provided in either task.

The design was fully crossed with respect to post-test spelling modality. This means that all participants were tested for spelling on half the words in their trained modality and on half the words in the other modality. The words were thus split into two lists, with vowel patterns matched across lists (see Appendix). Additionally, post-test modality was counter-balanced across lists, such that children in each practice group printed List A and typed List B, while others typed List A and printed List B. The design was fully counterbalanced.

## Results

A Principle Components Analysis was run with Direct Oblim Rotation on the multiple measures of Printing (Alphabet and Passage) and Typing (Alphabet and Passage) as well as the two TOWRE subtests (words and non-words) to explore possible data reduction by combining these into three composite scores (printing, typing, reading). However, the passage measures were unable to load on one factor and instead split between the three. The analysis was rerun without the Printing and Typing Passage tests (as the semantic and syntactic complexity of the phrase used was thought to have influenced performance) and this resulted in

three factors with simple structure accounting for 97% of the variance. All loadings were >0.95. Therefore, only the Alphabet tests were used in the following analyses as indices of printing and typing skill, and the two TOWRE tests were combined to create a reading composite.

Descriptive statistics for the initial assessment of printing, typing, and literacy skills are provided in **Table 1** along with decoding accuracy for the first exposure to the training stimuli (from the start of the training session). A multivariate analysis of variance indicated no significant differences between the two groups on any measure (all $Fs < 1.01$; $ps$ ranged from 0.32 to 0.90). Thus, the children in each group had comparable skills prior to the practice session.

The first objective of this research was to investigate the effect of practice modality (printing vs. typing) on orthographic learning. **Table 2** presents the proportions of practiced words identified correctly during the recognition post-tests, as a function of practice modality and post-test time (1 and 7 days). Accuracy rates were high but not at ceiling levels, hovering around 80% across groups and test dates. To investigate whether performance on the recognition tasks differed between the spelling practice groups or testing dates, a 2 (Training group: printing vs. typing) × 2 (Time: day 1 vs. day 7 post-test) repeated-measures analysis of variance (ANOVA) was conducted. There was no significant effect for group, $F_{(1, 38)} \leq 1.00$, $p = 0.86$ or for time, $F_{(1, 38)} = 1.29$, $p = 0.26$. The interaction between time and training group was also not found to be significant, $F_{(1, 38)} \leq 1.00$, $p = 0.62$.

The second and more stringent post-test of orthographic learning required participants to spell the practiced non-words in a dictation. **Table 3** presents the means and standard deviations

**Table 1 | Initial assessment performance as a function of practice group.**

| | | Practice group | |
| | | Printing (*n* = 19) | Typing (*n* = 21) |
| | Max. | Mean (*SD*) | Mean (*SD*) |
|---|---|---|---|
| Reading composite | 167 | 79.74 (26.39) | 78.71 (24.93) |
| Spelling | 59 | 25.26 (4.74) | 24.24 (3.65) |
| Printing baseline | 52 | 24.74 (9.72) | 23.38 (8.57) |
| Typing baseline | 52 | 16.89 (6.38) | 18.43 (7.26) |
| Stimuli decoding | 10 | 7.95 (2.12) | 7.24 (2.85) |

Note. Max., maximum score possible.

**Table 2 | Proportions of target non-words selected on recognition post-tests.**

| | Practice group | |
| | Printing (*n* = 19) | Typing (*n* = 21) |
| | Mean (*SD*) | Mean (*SD*) |
|---|---|---|
| Recognition Day 1 | 0.82 (0.17) | 0.80 (0.18) |
| Recognition Day 7 | 0.78 (0.18) | 0.78 (0.19) |

of the proportion of non-words spelled correctly by each practice group, across post-test days and post-test modality. Recall that within each group half of the items were post-tested via printing, the other half through typing. Again, accuracy rates appear consistent across groups and time, as well as across post-test modalities. A 2 (Training group: printing vs. typing) × 2 (Time: day 1 vs. day 7 post-test) × 2 (Post-test modality: printing vs. typing) repeated-measures ANOVA was conducted to evaluate this pattern of results. As suggested by the data presented, there was no significant main effect for group, $F_{(1, 38)} \leq 1.00$, $p = 0.99$, time, $F_{(1, 38)} \leq 1.00$, $p = 0.71$, or for modality used in the post-test, $F_{(1, 38)} \leq 1.00$, $p = 0.46$. There were also no significant interaction effects. Thus, groups responded similarly across post-test modality and days, with both groups showing impressive orthographic learning.

The next research question concerns the possible role of individual differences in literacy, printing, and typing skills on orthographic learning. In particular, it is of interest to explore whether skills in any of these areas interacted with the practice modality, which would suggest one modality may be more preferable over the other for certain students. This was addressed with multiple regression analyses in which individual data from the pre-tested areas served as predictor variables and performance on the post-tested recognition and spelling tests served as criterion variables. Given the null results reported above, data was collapsed across test dates and also across post-test modalities for the spelling tests. Preliminary analysis revealed that only pre-tested reading and spelling levels directly predicted the overall orthographic learning outcomes, and hence these literacy skills were entered in the first step of the models. Practice group was dummy- coded, and interaction terms were created by multiplying the dummy coded variable with each of the assessed areas. These interaction terms were tested individually in the last step of the regression models. All models are presented in **Table 4**.

In the first models presented in **Table 4**, the criterion variable was performance on the recognition tasks. Entered in step 1, children's pre-existing literacy skills accounted for a sizeable 51.6% of the variance in post-test recognition scores. Adding the practice group coding in step 2 did not account for any additional variance, consistent with the ANOVA results. Adding interaction terms separately at step 3 did not add any explanatory power to the model, except in the case of the term involving typing skills:

the addition of a term modeling the interaction between typing ability and group assignment accounted for an additional 5.5% of the variance in orthographic recognition, bringing the total variance accounted for to an impressive 57.1%.

The bottom half of **Table 4** shows the regression results with total post-tested spelling performance as the criterion variable. Pre-existing literacy skills accounted for a significant 56.5% of variance in spelling post-test performance. Entering practice group assignment in step 2 did not account for any additional variance, once again consistent with the ANOVA results. The only interaction term to make a significant contribution to the model was again found to be one incorporating pre-tested typing proficiency: the typing interaction term accounted for an additional 8.0% of unique variance in spelling post-tests, bringing the total variance explained to 64.8%. The pattern of results behind this significant interaction term is depicted clearly in **Figure 1**. From these scatterplots, it is apparent that typing skills facilitated and/or constrained learning but only within the typing practice group. For comparison purposes, the lower panels in the Figure show how a similar influence of printing skills was not observed within the printing practice group[1].

## DISCUSSION

The present study evaluated the influence of manual printing on establishing orthographic representations in memory, by comparing practice modality effects on the orthographic learning that occurs through spelling practice. To the best of our knowledge we are the first to address this research question by employing a carefully devised orthographic learning paradigm in which grade 2 students practiced spelling novel non-words either by printing or by typing. The non-words, as used in previous research, had ambiguous spellings and thus success in learning these new forms

---

[1]The interaction reported for the recognition task reflected a similar pattern of results.

**Table 3 | Proportions of target non-words spelled correctly on post-tests.**

| | Practice Group | |
| | Printing (n = 19) | Typing (n = 21) |
| | Mean (SD) | Mean (SD) |
|---|---|---|
| **SPELLING DAY 1** | | |
| Assessed via printing | 0.63 (0.27) | 0.65 (0.28) |
| Assessed via typing | 0.69 (0.26) | 0.62 (0.29) |
| **SPELLING DAY 7** | | |
| Assessed via printing | 0.60 (0.25) | 0.69 (0.24) |
| Assessed via typing | 0.69 (0.29) | 0.66 (0.29) |

**Table 4 | Regression analysis predicting performance on multiple-choice and spelling post-tests.**

| | $R^2$ | $\Delta R^2$ | $\Delta F$ |
|---|---|---|---|
| **CRITERION: RECOGNITION TASK** | | | |
| 1. Literacy (Reading and spelling) | 0.516 | 0.516 | 19.66*** |
| 2. Practice group | 0.516 | 0.000 | < 1.00 |
| 3. Reading × group interaction | 0.517 | 0.001 | < 1.00 |
| 3. Spelling × group interaction | 0.516 | 0.000 | < 1.00 |
| 3. Printing × group interaction | 0.517 | 0.001 | < 1.00 |
| 3. Typing × group interaction | 0.571 | 0.055 | 4.51* |
| **CRITERION: SPELLING TASK** | | | |
| 1. Literacy (Reading and Spelling) | 0.565 | 0.565 | 24.01*** |
| 2. Practice group | 0.568 | 0.003 | < 1.00 |
| 3. Reading × group interaction | 0.574 | 0.006 | < 1.00 |
| 3. Spelling × group interaction | 0.576 | 0.008 | < 1.00 |
| 3. Printing × group interaction | 0.569 | 0.001 | < 1.00 |
| 3. Typing × group interaction | 0.648 | 0.080 | 7.96** |

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.
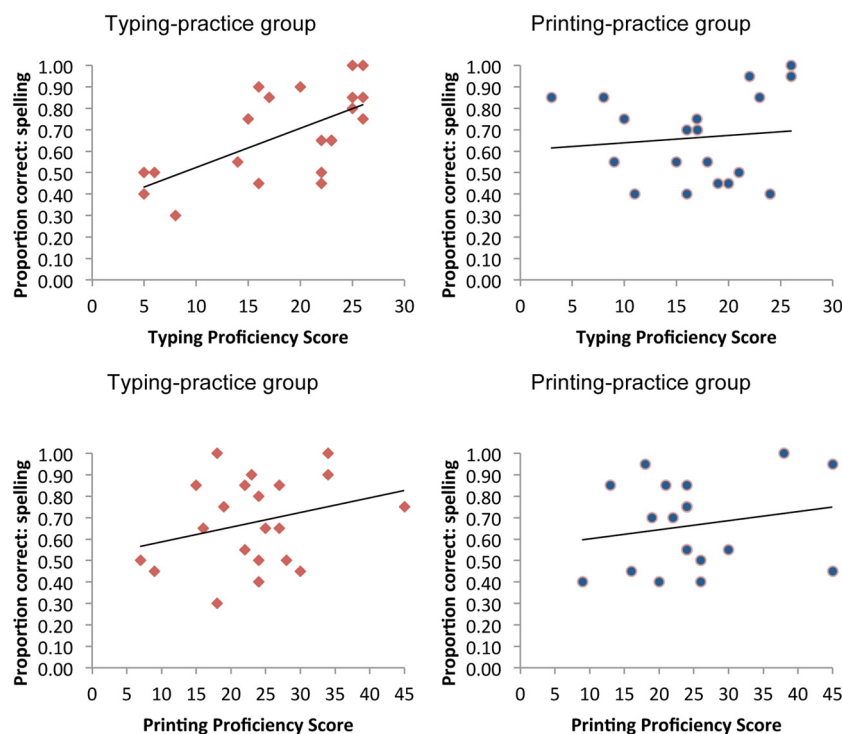
**FIGURE 1 | Scatterplots of the relations between pre-tested areas and performance on the spelling post-tests, as a function of practice group.** The upper panels depict the nature of the interaction between typing skills and practice group; the lower panels show the absence of any interaction between printing skills and practice group.

is seen as a clean metric of orthographic learning. This methodology then, makes it possible to specifically isolate the effects of printing practice vs. typing practice on the learning of new representations. The results indicated that spelling practice via printing and typing led to comparable amounts of orthographic learning, as measured by both visual recognition and spelling post-tests. The only pre-existing participant characteristic that interacted with practice modality in influencing orthographic learning was found to be typing skills; printing skills did not interact with practice modality in a similar fashion.

To measure orthographic learning, the present research examined performance on a recognition task and a spelling task. While not at ceiling levels, performance was strong across these tasks, both 1 and 7 days following the practice session and greater than what has typically been reported in the past (Cunningham and Stanovich, 1990; Vaughn et al., 1992, 1993). Thus, it appears that the methodology employed here was successful in bringing about orthographic learning, adding validity to the reported findings. The present research is the first to compare the effects of printing and typing utilizing an orthographic learning paradigm with a constrained set of ambiguously spelled non-words; previous studies have used a larger corpus of real words varying in consistency as well as younger participants and a procedure that included copying rather than devising spellings from memory. All of these factors may well have contributed to insufficient learning and the conflicting results of past research.

While the current results (of successful orthographic learning) support the use of spelling as a self-teaching mechanism

(Share, 1995, 2004; Ouellette, 2010), they do not support the hypothesis that spelling's effectiveness is linked to the manual movements involved in printing (Hulme, 1983). Given the methodological care of the present study, there is reason to have confidence that the lack of between-group differences in orthographic learning reported here is a valid and important finding in itself and makes an important contribution to both theory and teaching practice. That is, the null findings for any between-group differences suggest that printing and typing bring about equivalent levels of orthographic learning at this phase of literacy acquisition, confirming the earlier (null) findings of Vaughn and colleagues (1992, 1993) and Berninger et al. (1998) but with a more rigorous experimental design that specifically targeted orthographic learning. This may ease concerns of using keyboards within literacy curricula, while also clarifying the role of motoric knowledge and manual printing motions in learning; while there is evidence to suggest these may be important in initial alphabet learning where visual shape and motoric information appear connected (Longcamp et al., 2005, 2008), the current findings add to the literature showing no such connection for larger more detailed orthographic representations. The present results, in accord with Masterson and Apel (2006), suggest that lexical representations utilized in spelling are modality-free in terms of stored detail.

## INDIVIDUAL DIFFERENCES IN ORTHOGRAPHIC LEARNING
The present research design importantly allowed for an evaluation of the effects of pre-existing individual differences

on orthographic learning, as we obtained measures of literacy, printing, and typing proficiency at the onset of the study. Regression analyses indicated a primary role of pre-existing reading and spelling skills in orthographic learning. This is consistent with other research that has found skills such as decoding and orthographic knowledge to be significant predictors of orthographic learning (Castles and Nation, 2006; Ouellette and Fraser, 2009). Pre-existing literacy skills very much facilitate and/or constrain the acquisition of new orthographic representations regardless of the practice modality employed, highlighting the stability of early individual differences in literacy and the importance of early identification and intervention efforts.

An important novel finding of the present study was the significant interaction between typing skills and practice modality, such that pre-existing typing skills constrained and/or facilitated success in learning new words through typing practice; the same effect was not found for pre-existing printing skills on learning new words through printing. In other words, within the typing group only, orthographic learning was facilitated or constrained by pre-existing typing skills, even after controlling for pre-existing reading and spelling levels. While keyboarding skills have previously been found to interact with overall writing quality for older students in terms of content and style (e.g., Russell, 1999), this is the first study to show such an interaction with learning new orthographic representations. What makes this finding all the more interesting is that a comparable relation was not found between pre-existing printing skills and performance within the printing-practice group, a finding similar to that reported for spelling disabled students by Berninger et al. (1998), but somewhat surprising given past correlations between printing fluency and spelling (Kim et al., 2014). In the present study, weaker printing skills did not appear to constrain learning within the printing group and stronger printing skills did not facilitate learning. Thus the contention that weaker hand writers may benefit more from keyboarding (Russell, 1999; Blok et al., 2002) may not be empirically supported when it comes to learning new orthographic representations.

These results raise a pertinent question: why would orthographic learning be influenced by typing proficiency but not by printing skill? Printing fluency has been proposed to potentially interact with spelling in so far as laborious printing would tap cognitive resources and strain working memory (Kim et al., 2014), yet in the present study, slower printing did not appear to negatively impact the learning of new spellings. In contrast, laborious typing did have such a negative impact. This would suggest that there is something unique to typing that may affect cognitive and attentional resources more so than printing fluency does. At an elementary grade level, keyboarding can be more difficult than printing for some. Especially for children unfamiliar with keyboards, the other letters may serve as visual distracters and the child may expend more cognitive energy-and time- on visual scanning to find the right key. Anecdotally, our slower typers would often subvocalize while searching for the letter key (i.e., repeat the letter) which may have created even more interference within phonological

working memory as well (for remembering the subsequent sounds). Due to the visual and phonological processes evoked in non-fluent typing, lack of typing proficiency may cause even more interference with attentional and cognitive resources than would weak printing, in turn detrimentally impacting spelling. It reasons then, that as children become more proficient with typing, they gain considerable speed, extraneous letters/sounds become less interfering and orthographic learning benefits. It is evident within the present results that the students varied considerably in their familiarity and comfort with the keyboard and this impacted learning; this concern can be traced back to when microcomputers were first introduced into classrooms (e.g., Varnhagen and Gerber, 1984). What is more surprising is that this concern is finding empirical support today.

The present study adds to the growing literature showing strong orthographic learning resulting from spelling practice (Conrad, 2008; Shahar-Yames and Share, 2008; Ouellette, 2010). The question remains, beyond individual modality differences, what explains the strong orthographic learning that occurs through spelling practice? Ouellette and Sénéchal (2008) have suggested that the benefit of spelling lies in its highly analytical nature that forces the child to consider each and every sound in a word. In producing the spelling, the child then must focus on each and every letter in their production. The result is that children attend to both the phonology and orthography of the word in more detail than they would need to during reading. Consequently, orthographic learning through spelling may result in representations that are more complete than would be created through reading (Conrad, 2008). As discussed by Perfetti and Hart (2002), while reading may proceed with partial representations, accurate spelling cannot. The analytic nature of spelling also promotes student engagement which can further benefit learning (Ouellette et al., 2013).

## LIMITATIONS, APPLICATIONS AND DIRECTIONS FOR FUTURE RESEARCH

The present study provides insight into the role of printing and typing in orthographic learning. A grade 2 sample was chosen as this represents a time where the transition to orthographic learning should be of particular relevance. However, it remains unclear whether these results are applicable at different grade levels. The methodology employed here lends itself well to future research with students at different grade levels, to trace the developmental progression of modality influences in spelling practice. In addition, while the modest size of the present sample is comparable to previous orthographic learning studies and sufficient for the number of steps in the regression models, replications with larger samples and with students of differing learning profiles will further advance knowledge in this area. Further, while the number of words per cell in our statistical analyses are modest, this is consistent with (actually greater than in) previous research that has employed an orthographic learning paradigm (e.g., Nation et al., 2007). Still, future research may wish to expand the non-word set and increase the orthographic complexity of stimuli used within this paradigm. Finally, it may be of interest in future research to explore word level evaluation

of printing and typing skills. We found the complex sentence transcription task to be too difficult or abstract for this age group, but perhaps a task at the word level would add valuable insight into these developing skills; there may be as of yet unexplored lexical influence over printing not evident in printing isolated letters as is typically done in testing printing (see Kim et al., 2014). Likewise, future research may wish to qualitatively evaluate printing and hand-writing, to test for any possible role of quality over automaticity when it comes to literacy learning.

In summary, the current study assessed the effect of typing and printing on the orthographic learning garnered through spelling practice by grade 2 students. Results revealed no significant differences in learning between participants who practiced spelling the novel non-words by printing and those who practiced the non-words by typing. A hierarchical regression did reveal a significant role for pre-existing literacy skills, as well as an interaction between typing skill and practice modality. The present research is the first to employ an orthographic learning paradigm to compare the effects of typing vs. printing in literacy acquisition. The results do not support the hypothesis that the manual movements involved in printing make it a more effective learning modality, but instead highlight the importance of individual differences in learning and suggest that literacy draws upon modality-free lexical representations.

## ACKNOWLEDGMENTS

## REFERENCES

Berninger, V. W., Abbott, R., Rogan, L., Reed, E., Abbott, S., Brooks, A., et al. (1998). Teaching spelling to children with specific learning disabilities: the mind's ear and eye beat computer or pencil. *Learn. Disabil. Q.* 21, 106–122. doi: 10.2307/1511340

Berninger, V. W., Abbott, R. D., Jones, J., Wolf, B. J., Gould, L., Anderson-Youngstrom, M., et al. (2006). Early development of language by hand: composing, reading, listening, and speaking connections; three letter-writing modes; and fast mapping in spelling. *Dev. Neuropsychol.* 29, 61–92. doi: 10.1207/s15326942dn2901_5

Blok, H., Oostdam, R., Otter, M. E., and Overmaat, M. (2002). Computer-assisted instruction in support of beginning reading instruction: a review. *Rev. Educ. Res.* 72, 101–130. doi: 10.3102/00346543072001101

Bowey, J., and Miller, R. (2007). Correlates of orthographic learning in third-grade children's silent reading. *J. Res. Read.* 30, 115–128. doi: 10.1111/j.1467-9817.2007.00335.x

Castles, A., and Nation, K. (2006). "How does orthographic learning happen?" in *From Inkmarks to Ideas: Current Issues in Lexical Processing*, ed S. Andrews (Hove: Psychology Press), 151–179.

Conrad, N. J. (2008). From reading to spelling and spelling to reading: transfer goes both ways. *J. Educ. Psychol.* 100, 869–878. doi: 10.1037/a0012544

Cunningham, A. E., and Stanovich, K. E. (1990). Early spelling acquisition: writing beats the computer. *J. Educ. Psychol.* 82, 159–162. doi: 10.1037/0022-0663.82.1.159

Ehri, L. C. (2005). Learning to read words: theory, findings and issues. *Sci. Stud. Read.* 9, 167–188. doi: 10.1207/s1532799xssr0902_4

Hulme, C. (1983). *Multi-Sensory Teaching and Specific Reading Retardation*. London: Routledge and Kegan Paul.

Hulme, C., and Bradley, L. (1984). "An experimental study of multi-sensory teaching with normal and retarded readers," in *Dyslexia: Global Issues*, eds

R. Malatesha and H. Whitaker (The Hague: Martinus Nijhoff), 431–443. doi: 10.1007/978-94-009-6929-2_23

Kim, Y.-S., Al Otaiba, S., Puranik, C., Folsom, J. S., and Gruelich, L. (2014). The contributions of vocabulary and letter writing automaticity to word reading and spelling for kindergartners. *Read. Writ.* 27, 237–253. doi: 10.1007/s11145-013-9440-9

Longcamp, M., Boucard, C., Gilhodes, J.-C., Anton, J.-L., Roth, M., Nazarian, B., et al. (2008). Learning through hand-or typewriting influences visual recognition of new graphic shapes: behavioral and functional imaging evidence. *J. Cogn. Neurosci.* 20, 802–815. doi: 10.1162/jocn.2008.20504

Longcamp, M., Zerbato-Poudou, M.-T., and Velay, J.-L. (2005). The influence of writing practice on letter recognition in preschool children: a comparison between handwriting and typing. *Acta Psychol.* 119, 67–79. doi: 10.1016/j.actpsy.2004.10.019

Masterson, J. J., and Apel, K. (2006). Effect of modality on spelling words varying in linguistic demands. *Dev. Neuropsychol.* 29, 261–277. doi: 10.1207/s15326942dn2901_13

Nation, K., Angell, P., and Castles, A. (2007). Orthographic learning via self-teaching in children learning to read english: effects of exposure, durability, and context. *J. Exp. Child Psychol.* 96, 71–84. doi: 10.1016/j.jecp.2006.06.004

Ouellette, G. (2010). Orthographic learning in learning to spell: the roles of semantics and type of practice. *J. Exp. Child Psychol.* 107, 50–58. doi: 10.1016/j.jecp.2010.04.009

Ouellette, G., and Fraser, J. R. (2009). What exactly is a *yait* anyway: the role of semantics in orthographic learning. *J. Exp. Child Psychol.* 104, 239–251. doi: 10.1016/j.jecp.2009.05.001

Ouellette, G., and Sénéchal, M. (2008). Pathways to literacy: a study of invented spelling and its role in learning to read. *Child Dev.* 79, 899–913. doi: 10.1111/j.1467-8624.2008.01166.x

Ouellette, G., Sénéchal, M., and Haley, A. (2013). Guiding children's invented spellings: a gateway into literacy learning. *J. Exp. Educ.* 81, 261–279. doi: 10.1080/00220973.2012.699903

Perfetti, C., and Hart, L. (2002). "The lexical quality hypothesis," in *Precursors of Functional Literacy,* Vol. 11, eds L. Verhoeven, C. Elbrow, and P. Reitsma (Amsterdam: John Benjamins), 67–86.

Russell, M. (1999). Testing writing on computers: a follow-up study comparing performance on computer and on paper. *Educ. Policy Anal. Arch.* 7, 1–47.

Sénéchal, M., Ouellette, G., Pagan, S., and Lever, R. (2012). The role of invented spelling on learning to read in low-phoneme awareness kindergartners: a randomized controlled trial study. *Read. Writ.* 25, 917–934. doi: 10.1007/s11145-011-9310-2

Shahar-Yames, D., and Share, D. L. (2008). Spelling as a self-teaching mechanism in orthographic learning. *J. Res. Read.* 31, 22–39. doi: 10.1111/j.1467-9817.2007.00359.x

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

Share, D. L. (1999). Phonological recoding and orthographic learning: a direct test of the self-teaching hypothesis. *J. Exp. Child Psychol.* 72, 95–129. doi: 10.1016/0010-0277(94)00645-2

Share, D. L. (2004). Orthographic learning at a glance: on the time course and developmental onset of self-teaching. *J. Exp. Child Psychol.* 87, 267–298. doi: 10.1016/j.jecp.2004.01.001

Vaughn, S., Schumm, J. S., and Gordon, J. (1992). Early spelling acquisition: does writing really beat the computer? *Learn. Disabil. Q.* 15, 223–228. doi: 10.2307/1510245

Vaughn, S., Schumm, J. S., and Gordon, J. (1993). Which motoric condition is most effective for teaching spelling to students with and without learning disabilities? *J. Learn. Disabil.* 26, 191–198. doi: 10.1177/002221949302600306

Torgesen, J., Wagner, R., and Rashotte, C. (1999). *TOWRE: Test of Word Reading Efficiency.* Austin, TX: Pro-Ed, Inc.

Varnhagen, S., and Gerber, M. (1984). Use of microcomputers for spelling assessment: reasons to be cautious. *Learn. Disabil. Q.* 7, 266–270. doi: 10.2307/1510484

Wang, H.-C., Castles, A., Nickels, L., and Nation, K. (2011). Context effects on orthographic learning of regular and irregular words. *J. Exp. Child Psychol.* 109, 39–57. doi: 10.1016/j.jecp.2010.11.005

Woodcock, R. W., McGrew, K. S., and Mather, N. (2001). *Woodcock-Johnson III.* Itasca, IL: Riverside Publishing.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## APPENDIX

Target non-words and their alternatives.

| TARGET | HOMOPHONE |
| --- | --- |
| **WORD LIST A** | |
| YAIT | YATE |
| LURT | LERT |
| VEAK | VEEK |
| WOTE | WOAT |
| ROOP | RUPE |
| **WORD LIST B** | |
| CALE | CAIL |
| VERN | VURN |
| ZEET | ZEAT |
| POAN | PONE |
| JUME | JOOM |

# Inflectional and derivational morphological spelling abilities of children with Specific Language Impairment

**Sarah Critten[1], Vincent Connelly[2]\*, Julie E. Dockrell[3] and Kirsty Walter[3]**

[1] Department of Psychology, Coventry University, Coventry, UK
[2] Psychology, Oxford Brookes University, Oxford, UK
[3] Department of Psychology and Human Development, Institute of Education, University of London, London, UK

Children with Specific Language Impairment (SLI) are known to have difficulties with spelling but the factors that underpin these difficulties, are a matter of debate. The present study investigated the impact of oral language and literacy on the bound morpheme spelling abilities of children with SLI. Thirty-three children with SLI (9–10 years) and two control groups, one matched for chronological age (CA) and one for language and spelling age (LA) (aged 6–8 years) were given dictated spelling tasks of 24 words containing inflectional morphemes and 18 words containing derivational morphemes. There were no significant differences between the SLI group and their LA matches in accuracy or error patterns for inflectional morphemes. By contrast when spelling derivational morphemes the SLI group was less accurate and made proportionately more omissions and phonologically implausible errors than both control groups. Spelling accuracy was associated with phonological awareness and reading; reading performance significantly predicted the ability to spell both inflectional and derivational morphemes. The particular difficulties experienced by the children with SLI for derivational morphemes are considered in relation to reading and oral language.

**Keywords: spelling, SLI, morphemes, inflectional, derivational, reading, language, writing**

## 1. INTRODUCTION

Children with specific language impairment (SLI) experience problems with the acquisition and processing of oral language. They often have difficulties with the semantic, syntactic and phonological aspects of language and there is some debate around how children with SLI process morphemic affixes. In particular, there is some evidence that children with SLI have more difficulty with inflections suffixes (Montgomery and Leonard, 1998; Marshall and Van Der Lely, 2007; Oetting and Hadley, 2009).

Children with SLI also often have associated literacy difficulties in reading (Botting et al., 2006) and the production of written text (Dockrell et al., 2007). However, the specific difficulties that the children experience with spelling and the cognitive processes responsible for these difficulties remain a matter of debate (Silliman et al., 2006). The current study compares the performance of children with SLI and matched peers in spelling inflectional and derivational morphemes. The extent to which language or literacy skills underpin spelling performance of these different types of affixes is examined.

The ability to coordinate phonemes, orthographic features of written words and the morphological analysis of both base words and bound morphemes underlies the development of spelling (Nagy et al., 2006). Developing an accurate orthographic lexicon to support conventional spelling is an extended process and all three word forms (phonological, orthographic, morphological) are involved from the initial stages of learning to spell (Bahr et al., 2012).

Inflectional and derivational affixes are bound morphemes which play an important role when constructing meaningful text. Inflectional morphemes are suffixes which provide grammatical information about the base words they are bound to through marking, for example, agreement or tense. By contrast derivational morphemes may occur at the beginning (prefixes) or end of a word (suffixes) and produce semantic changes by transforming the grammatical form of a word. Any difficulty in spelling these bound morphemes will impact on the grammatical and semantic accuracy and the complexity of texts produced, and may help partly explain the writing difficulties of children with SLI (Dockrell and Connelly, 2013).

The difficulties experienced by children with SLI in spelling are well established and analysis of single word errors has typically shown a disproportionately high level of phonological spelling errors in the children's texts (e.g., Bishop and Clarkson, 2003). However, to date, the majority of studies have suggested that the spelling performance of children with SLI is commensurate with younger spelling or language matched peers for general spelling ability (Mackie and Dockrell, 2004; Cordewener et al., 2012). In particular, it has been shown that children with SLI consistently spell the root morphemes of inflected and derived words consistently and that this is closely tied and predicted by their general spelling ability and showed no difference when compared to a spelling matched typically developing group of children (Deacon et al., 2014). Therefore, while knowledge of the spelling of a word root can be helpful when dealing with the spelling of inflected

or derived forms (Goodwin et al., 2013) the difficulties of children with SLI are thought to lie more with particular aspects of morphology such as suffixes rather than the use of morphemes in general (Oetting and Hadley, 2009).

A number of studies that have focused on the spelling of inflectional morphemes have suggested that children with SLI have a particular weakness with producing regular past tense verbs ending in -ed and regular plural nouns ending in -s (Windsor et al., 2000; Mackie and Dockrell, 2004; Silliman et al., 2006; Larkin et al., 2013) and often omit them entirely. Two results are of particular importance from the Larkin et al. study (2013). Firstly there were trends that quantitative differences in accuracy might exist not only between the children with SLI and their chronological age matches, but also their younger spelling matches as well. However, the sample was too small to detect significant differences ($N = 15$). Secondly the error patterns differed within the group with SLI, suggesting that participants with weaker phonological skills experienced particular difficulties. Differential spelling difficulties in children with SLI may, thus, reflect the consequence of different vulnerabilities within the language system.

In contrast, the examination of the spelling of derivational morphemes has been comparatively neglected. Silliman et al. (2006) found trends suggesting that children with SLI may be poorer when spelling words comprising phonological and orthographic shifts from the base to the derived form compared to spelling matches again, reporting omission errors. However, the identification of qualitative differences may be obscured by the small sample size ($N = 8$). Given the importance of morphological skills in reading and writing (Green et al., 2003) and the efficacy of morphological instruction, especially for struggling writers (McCutchen et al., 2014) there is a need to further examine the difficulties that children with SLI experience with the spelling of derivational morphemes. Therefore, the first aim of the present study was to examine accuracy and error type when spelling inflectional and derivational morphemes for children with SLI compared to age and language/spelling matches.

The extent to which patterns of performance in spelling morphemes are related to performance on language and literacy measures, is also considered. It is theorized that language abilities should influence spelling via semantic-orthographic connections (Nation and Snowling, 1998a) and oral language skills are associated with spelling ability in typically developing children (e.g., Ouellette and Sénéchal, 2008). However, there have been varied results in determining which language skills may be linked to spelling in children with SLI as neither vocabulary (Dockrell and Connelly, 2013), nor narrative comprehension (McCarthy et al., 2012) significantly predicted spelling ability.

The consistent relationship found between phonological skills and spelling is relatively uncontested. As such it is an important skill to consider when examining links between oral language and spelling in children who are reported to have poor phonological skills (e.g., Bishop and Snowling, 2004; Fraser et al., 2010). Difficulties with phonologically based tasks have been associated with spelling problems in children with SLI (e.g., Bishop and Clarkson, 2003). Both rhyme and nonword reading have been found to predict general spelling ability (Dockrell and Connelly,

2013) and inflectional morpheme spelling (Larkin et al., 2013) respectively, in children with SLI.

However, English is a morphophonemic orthography and therefore spelling also involves an understanding and awareness of the linguistic relationship between sound and meaning. Morphological awareness (particularly of affixes) develops as children learn to recognize the regularities of bound morphemes across many words and has been consistently shown to make a unique contribution to spelling development in typically developing children (e.g., Nagy et al., 2006). Furthermore, inflectional and derivational morphological awareness may contribute to inflectional and derivational morpheme spelling respectively (Apel et al., 2012). To date the relationship between morphological awareness and spelling has not been examined in children with SLI. Children with SLI have poorer levels of morphological awareness in comparison to their same aged peers (Smith-Lock, 1995) and tend to omit inflected forms when speaking (Montgomery and Leonard, 1998; Marshall and Van Der Lely, 2007) and it is predicted that these oral language difficulties with morphology will impact on their spelling performance. Therefore, the second aim of the present study was to examine oral language ability, phonological awareness and morphological awareness in relation to both inflectional and derivational morpheme spelling in children with SLI. Building on the work of Larkin et al. (2013) the current study will examine the relationships between oral language and bound morpheme spelling specifically. The present study will also extend the Larkin et al. (2013) study by considering both derivational and inflectional morphemes.

Although the difficulties experienced with language by children with SLI may lead to consequent spelling problems there is also a close developmental relationship between reading and spelling (e.g., Zutell and Rasinski, 1989; Swanson et al., 2003) and children with SLI struggle with learning to read (Botting et al., 2006). It is, therefore, possible that reading may be an important moderator of spelling in children with SLI. Indeed recent studies have suggested that it is reading skills not oral language that predicts spelling in children with SLI (e.g., McCarthy et al., 2012; Mackie et al., 2013). However, it was general spelling ability that was examined in these studies rather than bound morpheme spelling specifically and the failure to examine specific spelling skills may have masked the influence of specific dimensions of the oral language system.

Error analyses can be used to highlight transitions in the relationship between phonological and morphological knowledge when learning to spell (Nunes et al., 1997; Critten et al., 2007). For example, initially when spelling complex words such as "filled" young children may omit the inflectional morpheme, e.g., "fil" as initial sounds are the first to be noted while awareness of the final sounds and middle sounds of words develops later (Ehri, 2005). When there is some awareness of the final sounds a phonologically implausible letter string may be supplied for the morpheme using incorrect phoneme-grapheme correspondences, e.g., filt where -t represents -ed. However, once more advanced phonological knowledge starts to develop then children may over-apply phoneme-grapheme correspondences to spell all aspects of a word including the morpheme, e.g., fild where -d represents -ed. It is only when children make correspondences

between morphological units in the oral language and their specific spellings and realize that not all words are spelled as they sound that correct application of units such as -*ed* can be observed, e.g., *filed*. Therefore, error analyses can highlight the underlying role of different aspects of oral language in bound morpheme spelling and suggest developmentally how children are progressing in their understanding of bound morphemes in the orthography. Error analysis will be used to investigate qualitative differences in spelling performance.

In this present study, children with SLI were compared to two control groups; one matched for chronological age (CA group) and one younger group matched for language and spelling (LA group). Since children with SLI show consistent spelling of word roots tied to general spelling ability (Deacon et al., 2014) then a spelling matched group should allow us to control for this factor (Goodwin et al., 2013) while concentrating on the suffix issue that the literature points out as a particular difficulty for children with SLI. The three groups were given dictated spelling tasks containing bound morphemes; inflectional morphemes of regular past tense verbs and regular plural nouns and derivational morphemes with phonological and orthographic shifts as indicated by previous findings (e.g., Silliman et al., 2006). An error-coding scheme was employed to focus on the bound morpheme spelling errors in reference to a typical developmental sequence of spelling errors (Critten et al., 2007) and the scheme used by Larkin et al. (2013) for children with SLI.

Finally a detailed assessment of language and literacy skills was conducted. Given the absence of any predictive effects derived for spelling from receptive vocabulary (Dockrell and Connelly, 2013) or narrative comprehension (McCarthy et al., 2012) oral language was measured by an expressive task of sentence generation. Phonological awareness was measured by both rhyme (given the findings of Dockrell and Connelly, 2013) and elision abilities. Morphological awareness was measured in relation to both inflectional and derivational awareness to build on the findings with typical children (Apel et al., 2012). Word reading was also examined (McCarthy et al., 2012; Mackie et al., 2013).

Our first objective was to examine both the accuracy and any errors in the children's spelling of inflectional and derivational morphemes in order to establish any differences between the children with SLI and their matched peers. We predicted that the children with SLI would perform significantly lower than the CA matches but commensurate with the LA matches. By contrast we reasoned, given the indicative data from Larkin et al. (2013) and Silliman et al. (2006) that accurate spelling of inflectional and derivational morphemes for the children with SLI would be poorer than both CA and LA matches and more omission errors would be made. The second objective was to examine which, if any, of our oral language dimensions were associated with inflectional and derivational morpheme spelling. We predicted that both oral morphological and phonological awareness would account for significant amounts of variance but that these associations would be moderated by reading ability.

## 2. METHODS

### 2.1. PARTICIPANTS
Ninety-nine children in three matched groups: (a) 33 children identified with SLI; (22 = males, 11 = females), mean age =

9:10 years, $SD = 3.57$ months (range = 11 months). Children of this age were chosen as the spelling of younger children with SLI may be difficult to interpret due to floor effects when required to carry out a complex spelling task involving morphology. (b) 33 children matched for chronological age (CA) and gender, mean age = 9:10 years, $SD = 2.94$ months (range = 10 months) and (c) 33 children matched for gender, language (formulated sentences) and single word spelling abilities (LA), mean age = 8;1 years, $SD = 6.25$ months (range = 7 months). All children had English as their first language and were predominantly of white, British ethnicity. The level of Social Economic Status (SES) was controlled for across schools by checking that the percentage of children receiving free schools meals (a strong indicator of SES in the UK) was in the average range.

To recruit the SLI sample, children were identified across five counties in southern England. Professionals were asked to nominate children who had specific language impairments who participated in a screening process using the four core sub-tests of the Clinical Evaluation of Language Fundamentals, 4th edition (CELF-4 UK, Semel et al., 2006): concepts and following directions, recalling sentences, formulated sentences, word classes receptive and expressive. For a diagnosis of SLI, children had to achieve a standard score of 75 or below (2 SDs below the mean). The matrices test from the British Ability Scales, 2nd Edition (BAS II: Elliott et al., 1997) established non-verbal abilities within the average range. As **Table 1** shows all participants met the criteria for SLI, with a significant difference between their CELF-4 test score and their BAS II matrices test: $t_{(64)} = 15.39$, $p < 0.001$, $r = 0.89$. Additional measures examined phonological awareness, morphological awareness and reading and are detailed in **Table 2**.

The two groups of comparison children attended the same primary schools as those diagnosed with SLI, and were selected by teachers on the basis of average attainment on curriculum assessments and no additional learning needs. The CA comparison children were confirmed as having language ability and non-verbal ability within the normal range using the same CELF-4 UK core tests and the BAS II matrices and were matched in age to the children with SLI within 3 months and did not differ overall in age.

The LA comparison children also had scores on language and non-verbal ability within the average range and were matched with the children with SLI using their raw score on the formulated sentences task from the CELF-4 UK. The LA comparison children were also matched to the SLI group using their raw score on the single word spelling task from the BAS II. Despite the fact that the CA group was chosen purely for their age they scored significantly higher than the other two groups for non-verbal ability although the SLI and LA groups did not differ.

### 2.2. MEASURES
#### 2.2.1. General language ability
Clinical Evaluation of Language Fundamentals (CELF-4 UK, Semel et al., 2006). The CELF provides core sub-tests of receptive and expressive language abilities. This produces a Total Language Score that can be utilized for the identification of language impairment. Children from the SLI and CA groups were screened for language ability using the four core sub-tests

**Table 1 | Means, (standard deviations), f score, df, *p*-value, effect size and Bonferroni *post-hoc* results (where applicable) for screening measures per group: SLI, CA, LA.**

| Measure | SLI | CA | LA | F | df | p | Partial $\eta^2$ | Bonferroni *post-hoc* |
|---|---|---|---|---|---|---|---|---|
| Core language standard score (CELF) | 68.4 (5.5) | 102.9 (11.3) | 93.6 (7.9) | | | | | |
| Non-verbal abilities: matrices ability score (BAS) | 96.1 (6.5) | 104.9 (9.7) | 98.9 (7.8) | | | | | |
| **MATCHING VARIABLES** | | | | | | | | |
| Age in years/months (SD in months) | 9/10 (3.6) | 9/10 (2.9) | 8/1 (6.2) | 244.3 | 2,96 | <0.001 | 0.84 | SLI = CA > LA |
| Formulated sentences raw score (CELF) | 31.4 (4.2) | 47.5 (4.4) | 31.2 (4.2) | 155.8 | 2,96 | <0.001 | 0.77 | SLI = LA, SLI < CA LA < CA |
| Spelling raw score (BAS) | 16.3 (4.3) | 22.3 (5.2) | 16.6 (4.8) | 16.1 | 2,96 | <0.001 | 0.25 | SLI = LA, SLI < CA LA < CA |
| Spelling ability score (BAS) | 85.0 (16.5) | 121.1 (16.3) | 94.4 (13) | 48.5 | 2,96 | <0.001 | 0.50 | SLI < LA < CA |

**Table 2 | Means, standard deviations, f score, df, *p*-value, effect size and Bonferroni *post-hoc* results for language and literacy measures per group: SLI, CA, LA.**

| Measure | SLI | CA | LA | F | df | P | Partial $\eta^2$ | Bonferroni *post-hoc* |
|---|---|---|---|---|---|---|---|---|
| Inflectional morphological awareness raw score | 10.9 (1.4) | 12.8 (0.36) | 12.1 (0.97) | 30.4 | 2,96 | <0.001 | 0.39 | SLI < LA < CA |
| Inflectional morphological awareness z score | −0.79 (1.1) | 0.71 (0.29) | 0.09 (0.76) | 30.4 | 2,96 | <0.001 | 0.39 | SLI < LA < CA |
| Derivational morphological awareness raw score | 5.1 (1.1) | 5.8 (0.48) | 5.3 (0.95) | 4.9 | 2,96 | 0.009 | 0.09 | SLI < CA SLI = LA CA = LA |
| Derivational morphological awareness z score | −0.3 (1.2) | 0.41 (0.52) | −0.11 (1.0) | 4.9 | 2,96 | 0.009 | 0.09 | SLI < CA SLI = LA CA = LA |
| Phonological elision raw score (CTOPP) | 10.8 (4.3) | 17.2 (2.9) | 14.2 (4.2) | 22.9 | 2,96 | <0.001 | 0.32 | SLI < LA < CA |
| Phonological rhyme raw score (PhAB) | 12.3 (4.5) | 18.4 (3.1) | 17.1 (3.4) | 24.6 | 2,96 | <0.001 | 0.34 | SLI < CA SLI < LA CA = LA |
| Phonological awareness z score (elision z + rhyme z) | −1.5 (1.58) | 1.2 (1.0) | 0.29 (1.3) | 36.5 | 2,96 | <0.001 | 0.43 | SLI < LA < CA |
| Single word reading raw score (YARC) | 31.6 (11.0) | 49.2 (5.8) | 38.7 (8.0) | 35.0 | 2,96 | <0.001 | 0.42 | SLI < LA < CA |
| Single word reading z score (YARC) | −0.73 (0.9) | 0.84 (0.52) | −0.09 (0.72) | 35.0 | 2,96 | <0.001 | 0.42 | SLI < LA < CA |

for 9–16 years: (a) *Concepts and following directions*; children are shown pictures and asked to identify items and/or point to them in a prescribed order according to a verbal instruction, (b) *Recalling sentences*; children are asked to imitate orally presented sentences, (c) *Formulated sentences*; children are shown a picture of a scene and asked to verbalize a sentence that both describes the picture and includes a target word, (d) *Word classes*; children are verbally presented with four words and asked to first identify the two words that go together (receptive component) and then to explain why they go together (expressive component). The children from the LA group were given the four core sub-tests for 5–8 years where the *word classes* task is replaced by *word structure*; children are shown pictures and asked to describe them using a verbal prompt designed to elucidate understanding of word class and morphology. Reliability for the core sub-tests for 9–10 years, 0.94 and for 5–8 years, 0.95–0.96.

### 2.2.2. Non-verbal ability
The British Ability Scale II (BAS II) Matrices subtest (Elliott et al., 1997). Children are presented with a set of patterns presented in a four or six part grid where one part of the grid is incomplete and children are required to select the missing piece from six possible responses; reliability 0.85, validity with the WISC-III performance scale 0.47

### 2.2.3. Spelling
The British Ability Scale II (BAS II) Spelling subtest (Elliott et al., 1997). Children are verbally presented with a series of phonetically regular and irregular monosyllabic and bisyllabic words. The

words are first presented in isolation, then within the context of a sentence and finally in isolation and asked to respond by writing the word: reliability 0.91.

### 2.2.4. Phonological awareness
Complete Test Of Phonological Processing (CTOPP; Wagner et al., 1999) and Phonological Assessment Battery (PhAB; Frederickson et al., 1997). (1) Children were tested on the elision task from the CTOPP which requires identification and segmentation of the different phonological units within words, reliability, 0.80; validity with the Woodcock Reading Mastery Test—R (Word Attack and Word Identification sub-tests) 0.49–0.84 and (2) A test of rhyme from the PhAB where children chose two words that rhyme out of a choice of three (one irrelevant word and two that rhyme); reliability ≥0.80; validity with the Neale Analysis of Reading Ability (NARA; Neale et al., 1997) reading accuracy 0.24–0.56.

### 2.2.5. Inflectional and derivational morphological awareness
A test of morphological awareness was created from selected items on the CELF-4 UK (Semel et al., 2006): only items assessing awareness of inflectional morphemes ($N = 13$) and derivational morphemes ($N = 6$) were used in the current study. An example of an inflectional item is to show children a picture of a horse and say "Here is one horse," then another picture with two horses is pointed to: "Here are two . . ." and the child has to supply the word with the correct inflected morpheme of -s. An example of a derivational item is to show a picture of a teacher and say: "This man teaches. He is called a . . ." and the child has to supply the word with the correct derived morpheme of -er.

### 2.2.6. Reading

York Assessment of Reading Comprehension (YARC) Passage reading (Snowling et al., 2009). Children were given the Single Word Reading Task (SWRT) comprising 60 words presented on a card and asked to read them aloud, reliability, 0.85.

### 2.2.7. Experimental morphological spelling tasks

A list of 42 words was developed and presented as two 21 word spelling tests, delivered in a randomized order. The majority of words were derived from previous studies conducted with children aged 7–11 years and so were considered appropriate for the ages of the sample in this study. Written word frequency analyses had been completed by the original researchers for inflectional words ending in -ed (Nunes et al., 1997) and -s (Kemp and Bryant, 2003) and derivational words including phonological, orthographic and phonological and orthographic shifts (Mossing et al., 2009; Wiggins et al., 2010) to establish comparable levels within the morpheme types. Furthermore, for the present study written word frequency was also checked using the UK derived Children's Printed Word Database (Masterson et al., 2003). This demonstrated that the frequency of the inflectional words ranged from 3 to 1652 and that the derivational words were generally less frequent, as would be expected, ranging from 3 to 533. See in supplemental materials for the complete word list and written word frequency scores.

### 2.2.8. Inflectional morphemes

Derived and adapted from Nunes et al. (1997) and Kemp and Bryant (2003). There were 24 words containing inflectional morphemes; 12 regular past tense verbs containing -ed, e.g., *filled* and 12 regular plural nouns, e.g., *trees*.

### 2.2.9. Derivational morphemes

Derived and adapted from Silliman et al. (2006), Mossing et al. (2009) and Wiggins et al. (2010). There were 18 words containing derivational morphemes: six where there was a phonological shift from the base word to the derived form, e.g., *different*, six where there was an orthographic shift from the base word to the derived form, e.g., *attention* and six where there were both phonological and orthographic shifts, e.g., *student*.

### 2.3. PROCEDURE

All children were assessed individually in a quiet room at school. Ethical approval for the study had been gained in line with guidelines from the British Psychological Society (BPS) through the university ethics committee and informed consent from schools, parents and children was provided prior to any testing. During the screening process the CELF core tests, BAS matrices and BAS spelling were administered in two testing sessions. The two morphological spelling tasks and the phonological awareness and morphological awareness tasks were delivered over two further testing sessions. Children were allowed to terminate the sessions if they wished. However, no child terminated the sessions since the organization of data collection into different sessions resulted in manageable time periods of testing for the children.

All standardized tests were administered according to the procedures in the manual. For the morphological spelling tasks, each word was verbally presented in isolation, in the context of a sentence and then in isolation again and children were asked to write out the word.

### 2.4. CLASSIFICATION OF SPELLING ERRORS WITHIN THE MORPHOLOGICAL SPELLING TASKS

The focus was only on the spelling of the inflectional or derivational morpheme within each word, i.e., the spelling of the base was not analyzed further. Morphemes which were incorrectly spelled were categorized into one of the following mutually exclusive error types (Larkin et al., 2013) (1) Omission where the morpheme was not attempted at all, e.g., *fill* as an error attempt of *filled*, or *atten* as an error attempt of *attention* (2) Phonologically implausible where the morpheme was attempted (incorrectly) but the phoneme-grapheme correspondences did not produce a correct pronunciation, e.g., *fillt* where -t is not a phonologically plausible version of -ed or *attensed* where -sed is not a phonologically plausible version of -tion, (3) Phonologically plausible where the morpheme was again incorrectly spelled but the phoneme-grapheme correspondences did produce a correct pronunciation of the target morpheme, e.g., *filld*, where -d is a phonologically plausible attempt for -ed, or *attenshun* where -shun is a phonologically plausible attempt for -tion. The spelling errors were coded by two of authors of this paper and achieved an inter-rater reliability of 100%.

## 3. RESULTS

The results are presented in three sections. Section 1 examines group differences in children's spelling performance according to morpheme type. Section 2 examines associations between to inflectional and derivational morphological spelling abilities and language and literacy measures. Finally Section 3 examines predictors of children's inflectional and derivational morphological spelling ability using hierarchical regressions.

### 3.1. GROUP DIFFERENCES IN MORPHOLOGICAL SPELLING ABILITY

Means (SD) of group performance for each morpheme type are presented in **Table 3**. A Mixed ANOVA with group (between subjects factor) and morpheme type; inflectional and derivational (within subjects factor) was conducted for the number of words (base + morpheme) spelled correctly. Non-verbal ability and chronological age were added as co-variates although neither were significant [non-verbal ability $F_{(1, 94)} = 0.09$, $p = ns$; chronological age $F_{(1, 94)} = 0.01$, $p = ns$]. There was a main effect of group $F_{(2, 94)} = 30.62$, $p < 0.001$, $\eta p^2 = 0.39$ and morpheme type $F_{(2, 94)} = 9.69$, $p = 0.002$, $\eta p^2 = 0.09$ confirming that more words containing inflectional morphemes were correctly spelled and also a significant interaction between group and morpheme type $F_{(2, 94)} = 8.17$, $p = 0.001$, $\eta p^2 = 0.15$. Subsequent multivariate ANOVAs confirmed that there were group differences for both the number of words containing inflectional morphemes correctly spelled $F_{(2, 94)} = 31.21$, $p < 0.001$, $\eta p^2 = 0.39$ and the number of words containing derivational morphemes correctly spelled $F_{(2, 94)} = 34.26$, $p < 0.001$, $\eta p^2 = 0.42$. Bonferroni *post-hoc* analyses revealed that for both word types, the SLI and LA groups did not differ but were significantly less accurate than the CA group ($p < 0.001$).

**Table 3 | Means and standard deviations for the number of words and morphemes spelled correctly and number and proportions of error types (omission, phonologically implausible, non-phonologically plausible) according to Group (SLI, CA, LA) and morpheme type (Inflectional, Derivational).**

| Morpheme type | SLI | CA | LA |
|---|---|---|---|
| **INFLECTIONAL (/24)** | | | |
| Correct words | 9.2 (5.9) | 18.6 (3.7) | 11.4 (5.3) |
| Correct morphemes | 16.8 (5.0) | 22.0 (3.0) | 17.9 (3.9) |
| Omission: number | 1.6 (2.3) | 0.1 (0.4) | 0.5 (0.9) |
| Omission: proportion | 0.2 (0.2) | 0.2 (0.4) | 0.1 (0.1) |
| Phonologically implausible: number | 0.3 (0.6) | 0.0 (0.1) | 0.1 (0.6) |
| Phonologically implausible: proportion | 0.0 (0.1) | 0.0 (0.0) | 0.0 (0.0) |
| Phonologically plausible: number | 5.4 (4.2) | 1.3 (2.2) | 5.1 (3.3) |
| Phonologically plausible: proportion | 0.8 (0.2) | 0.8 (0.3) | 0.9 (0.2) |
| **DERIVATIONAL (/18)** | | | |
| Correct words | 3.1 (3.0) | 9.8 (4.2) | 4.3 (3.1) |
| Correct morphemes | 5.9 (3.9) | 12.9 (2.9) | 8.5 (4.0) |
| Omission: number | 1.1 (1.8) | 0.0 (0.0) | 0.2 (0.8) |
| Omission: proportion | 0.1 (0.1) | 0.0 (0.0) | 0.0 (0.0) |
| Phonologically implausible: number | 5.5 (3.4) | 0.7 (1.2) | 3.0 (3.39) |
| Phonologically implausible: proportion | 0.4 (0.2) | 0.1 (.1) | 0.3 (0.2) |
| Phonologically plausible: number | 5.4 (2.4) | 4.0 (2.1) | 6.4 (3.0) |
| Phonologically plausible: proportion | 0.5 (0.2) | 0.9 (0.1) | 0.7 (0.3) |

This analysis was then repeated using the same factors and co-variates but this time with the accuracy scores for the spelling of the morphemes alone, again neither co-variate was found to be significant; non-verbal ability $F_{(1, 94)} = 1.85$, $p =$ ns and chronological age $F_{(1, 94)} = 0.08$, $p =$ ns. As before there was a main effect of group $F_{(2, 94)} = 27.33$, $p < 0.001$, $\eta p^2 = 0.37$ and morpheme type $F_{(1, 94)} = 19.21$, $p < 0.001$, $\eta p^2 = 0.17$ and a significant interaction between group and morpheme type $F_{(2, 94)} = 6.43$, $p < 0.001$, $\eta p^2 = 0.12$. Subsequent multivariate ANOVAs confirmed that there were group differences for both the number of inflectional morphemes correctly spelled $F_{(2, 96)} = 14.98$, $p < 0.001$, $\eta p^2 = 0.24$ and the number of derivational morphemes correctly spelled $F_{(2, 96)} = 31.78$, $p < 0.001$, $\eta p^2 = 0.40$. In both cases the effect sizes were large. Bonferroni *post-hoc* analyses revealed that for the inflectional morphemes the SLI and LA groups did not differ but were significantly less accurate than the CA group ($p < 0.001$). In contrast the three groups differed in their performance on derivational morphemes where the SLI group was less accurate than both the CA ($p < 0.001$), and LA ($p < 0.001$), groups and the LA group was poorer than the CA group ($p < 0.001$).

To examine this further the proportions of error type were compared between the groups (between subjects factor). **Table 3** presents the number and proportions of the types of errors made by the three different groups. Some children (CA group $N = 19$, SLI group $N = 2$) made no inflectional morpheme spelling errors and were therefore excluded from the analysis. Overall all three groups tended to make phonologically plausible errors when spelling inflectional morphemes and the number of omissions and phonologically implausible attempts were negligible. Thus, there were no group differences for omission errors: $F_{(2, 75)} =$

2.39, $p =$ ns, phonologically implausible errors: $F_{(2, 75)} = 0.72$, $p =$ ns or phonologically plausible errors: $F_{(2, 75)} = 2.65$, $p =$ ns.

For derivational morphemes, all children made at least one spelling error and therefore no child was excluded from the analysis. Group differences were apparent when exploring error type for omission errors: $F_{(2, 96)} = 9.32$, $p < 0.001$, $\eta p^2 = 0.16$, phonologically implausible errors: $F_{(2, 96)} = 19.08$, $p < 0.001$, $\eta p^2 = 0.28$ and phonologically plausible errors: $F_{(2, 96)} = 19.07$, $p < 0.001$, $\eta p^2 = 0.33$. The largest effect was evident for phonologically plausible errors whereas the difference for omission errors was negligible. Bonferroni *post-hoc* analyses revealed that for the phonologically implausible errors the SLI group made proportionately more than both the CA ($p < 0.001$), and LA ($p < 0.001$), groups and the LA group made more of both type of error than the CA group ($p < 0.001$). In contrast for the phonologically plausible errors, the SLI made proportionately fewer compared to both the CA ($p < 0.001$), and LA ($p < 0.001$), groups and the LA group made fewer than the CA group ($p < 0.001$).

### 3.2. THE RELATIONSHIPS BETWEEN INFLECTIONAL AND DERIVATIONAL MORPHOLOGICAL SPELLING, LANGUAGE AND READING MEASURES

Group correlations (partialling out non-verbal ability) were conducted for inflectional morpheme spelling ability (number of inflectional morphemes spelled correctly), derivational morpheme spelling ability (number of derivational morphemes spelled correctly), oral language ability (CELF Formulated sentences), phonological ability (combined z scores from the CTOPP elision and PhAB rhyme tasks), inflectional morphological awareness, derivational morphological awareness and word reading ability (YARC single word reading test) and are shown in the supplemental materials link for **Table 4**. To control for Type I errors, a Bonferroni correction was computed at 6/0.05 = 0.008.

For both the SLI and LA groups, there were significant relationships between inflectional and derivational morphological spelling ability and phonological and word reading abilities. However, for the CA group while the relationships with phonological and reading abilities remained for inflectional morphological spelling, for derivational spelling the relationship with phonological awareness was no longer significant but rather derivational morphological spelling ability was significantly related to derivational morphological awareness. Notably there were no significant relationships between oral language ability and morphological spelling in any group while phonological and reading abilities were correlated for all groups. Furthermore, reading related to derivational morphological awareness but only for the CA group and oral language ability related to phonological ability but only for the LA group.

### 3.3. PREDICTORS OF INFLECTIONAL AND DERIVATIONAL MORPHOLOGICAL SPELLING

Hierarchical regressions (See **Table 5**) were used to examine the predictors of inflectional and derivational morphological spelling ability. Analyses were collapsed across the groups to provide sufficient power to address this question. The first regression analyses examined predictors of inflectional morphological

**Table 4 | Correlations between various measures according to group (SLI, CA, LA) controlling for non-verbal ability.**

| Group | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| **SLI** | | | | | | |
| 1. Inflectional morpheme spelling ability raw score | 0.71** | 0.14 | 0.62** | 0.07 | 0.03 | 0.84** |
| 2. Derivational morpheme spelling ability raw score | | 0.02 | 0.61** | 0.23 | 0.19 | 0.83** |
| 3. Oral language ability raw score (CELF Formulated sentences) | | | 0.11 | 0.14 | 0.16 | 0.05 |
| 4. Phonological ability z score (CTOPP elision + PhAB Rhyme) | | | | 0.40* | 0.02 | 0.69** |
| 5. Inflectional morphological awareness raw score | | | | | 0.12 | 0.14 |
| 6. Derivational morphological awareness raw score | | | | | | 0.08 |
| 7. Word reading raw score (YARC) | | | | | | – |
| **CA** | | | | | | |
| 1. Inflectional morpheme spelling ability raw score | 0.66** | 0.15 | 0.47** | 0.21 | 0.30 | 0.40* |
| 2. Derivational morpheme spelling ability raw score | | 0.01 | 0.34 | 0.33 | 0.40* | 0.51** |
| 3. Oral language ability raw score (CELF: Formulated sentences) | | | 0.03 | 0.03 | 0.13 | 0.00 |
| 4. Phonological ability z score (CTOPP elision + PhAB Rhyme) | | | | 0.08 | 0.42* | 0.50** |
| 5. Inflectional morphological awareness raw score | | | | | 0.02 | 0.20 |
| 6. Derivational morphological awareness raw score | | | | | | 0.48** |
| 7. Word reading raw score (YARC) | | | | | | – |
| **LA** | | | | | | |
| 1. Inflectional morpheme spelling ability raw score | 0.46** | 0.0 | 0.46** | 0.07 | 0.05 | 0.60** |
| 2. Derivational morpheme spelling ability raw score | | 0.17 | 0.51** | 0.03 | 0.14 | 0.68** |
| 3. Oral language ability raw score (CELF Formulated sentences) | | | 0.37* | 0.03 | 0.15 | 0.02 |
| 4. Phonological ability z score (CTOPP elision + PhAB Rhyme) | | | | 0.03 | 0.23 | 0.65** |
| 5. Inflectional morphological awareness raw score | | | | | 0.03 | 0.07 |
| 6. Derivational morphological awareness raw score | | | | | | 0.03 |
| 7. Word reading raw score (YARC) | | | | | | – |

**Significant at 0.008 (Bonferroni adjusted for multiple comparisons).*
*Significant at 0.05.*

**Table 5 | Summary of the final model of hierarchical regressions analysis when predicting inflectional and derivational morphological spelling ability.**

| Variable | B | SE B | Beta | T | p |
|---|---|---|---|---|---|
| **PREDICTING INFLECTIONAL MORPHOLOGICAL SPELLING** | | | | | |
| Non-verbal ability (BAS matrices) | −0.086 | 0.032 | −0.185 | −2.67 | 0.009 |
| Chronological age | 0.031 | 0.032 | 0.074 | 0.95 | ns |
| Oral language ability raw score (CELF: Formulated sentences) | −0.009 | 0.050 | −0.017 | −0.176 | ns |
| Inflectional morphological awareness raw score (CELF: Word Structure) | 0.286 | 0.267 | 0.079 | 1.07 | ns |
| Phonological ability z score (CTOPP elision + PhAB Rhyme) | 0.455 | 0.284 | 0.173 | 1.60 | ns |
| Word reading raw score (YARC) | 0.267 | 0.044 | 0.655 | 6.13 | <0.001 |
| **PREDICTING DERIVATIONAL MORPHOLOGICAL SPELLING** | | | | | |
| Non-verbal ability (BAS matrices) | −0.030 | 0.031 | −0.062 | −0.972 | ns |
| Chronological age | 0.032 | 0.030 | 0.074 | 1.07 | ns |
| Oral language ability raw score (CELF: Formulated sentences) | 0.034 | 0.044 | 0.064 | 0.779 | ns |
| Derivational morphological awareness raw score (CELF: Word Structure) | 0.128 | 0.316 | 0.025 | 0.404 | ns |
| Phonological ability z score (CTOPP elision + PhAB Rhyme) | 0.384 | 0.268 | 0.143 | 1.43 | ns |
| Word reading raw score (YARC) | 0.283 | 0.040 | 0.680 | 7.04 | <0.001 |

spelling ability. Non-verbal ability and chronological age was entered in the first step, followed by oral language ability, inflectional morphological awareness, phonological awareness, and word reading in the second step. The model from the first step did not prove significant [$F_{(2, 95)} = 16$, $p =$ ns, Adjusted $R$-square = 0.02]. However, once the variables in the second step were added a significant model did emerge [$F_{(6, 91)} = 26.79$, $p < 0.001$, Adjusted $R$-square = 0.62, $r$-square change = 0.64] and demonstrated that non-verbal ability and word reading were the only significant predictors of inflectional morphological spelling ability and that chronological age, oral language ability, inflectional morphological awareness and phonological

awareness, did not significantly contribute to explaining the variance.

A second regression analyses examined predictors of derivational morphological spelling ability. Variables were entered in the same steps as the inflectional morphology regression although derivational morphological awareness was entered in place of inflectional awareness. The model from the first step did not prove significant [$F_{(2, 95)} = 1.78$, $p = ns$, Adjusted $R$-square = 0.02]. However, once the variables in the second step were added a significant model did emerge [$F_{(6, 91)} = 34.19$, $p < 0.001$, Adjusted $R$-square = 0.67, $r$-square change = 0.66] and demonstrated that word reading was the only significant predictor of derivational morphological spelling ability and that non-verbal ability, chronological age, oral language ability, inflectional morphological awareness and phonological awareness, did not significantly contribute to explaining the variance.

## 4. DISCUSSION

### 4.1. INFLECTIONAL MORPHEME SPELLING

Previous studies of inflectional morpheme spelling indicated that children with SLI might be poorer at spelling regular past tense and plural morphemes and these inflections would be frequently omitted in comparison with both CA and LA matched peers. However, it was found that the children with SLI were as proficient at spelling inflectional morphemes as their language and spelling ability matched peers but both these groups of children were poorer at spelling inflectional morphemes than their chronological age matched peers. These results demonstrate that performance in spelling ability is more predictive of how accurately these suffixes are spelt rather than morphological awareness. Other studies that have examined both general spelling ability (Cordewener et al., 2012) and the spelling of word roots in children with SLI (Deacon et al., 2014) have reported similar results.

When children failed to spell the inflection accurately the pattern of errors across the groups were broadly similar. There was a predominance of phonologically plausible errors and a very small proportion of phonologically implausible error types. Therefore, most children were employing the developmentally sophisticated strategy of using phoneme-grapheme correspondences (as evidenced by error category) when attempting to spell these morphemes. The minority of omission errors for the SLI group was surprising given previous research (Larkin et al., 2013). However, our larger sample of children with SLI was slightly older than the Larkin et al. (2013) sample and might be showing a benefit of longer experience at school.

All groups showed relationships between inflectional morpheme spelling and phonological awareness. However, no group showed a relationship between inflectional morpheme spelling and either measure of morphological awareness or our measure of expressive oral language. Finally for all groups, there were strong relationships between inflectional morpheme spelling and reading. Thus, it is apparent that in the current cohort inflectional morpheme spelling was associated with the quality of the underlying orthographic and phonological representations that are most often associated with spelling and reading skills. Thus, although the children with SLI seem delayed in their spelling of

inflectional morphemes compared to chronological age matches, their spelling ability is underpinned by the same factors as their language and spelling matches. This further confirms previous findings with children with SLI (Dockrell et al., 2007, 2009; Dockrell and Connelly, 2013).

Contrary to other work with typically developing children (e.g., Apel et al., 2012) we found no relationship between inflectional morphological awareness and inflectional spelling in any of the groups sampled despite the fact that inflectional awareness seems quite well developed overall. Therefore, while inflectional morphological awareness could potentially still be contributing to the children's general knowledge of English spelling, phonological and orthographic knowledge are likely forming the representational basis for inflectional morpheme spelling rather than awareness of inflectional morphemes specifically.

### 4.2. DERIVATIONAL MORPHEME SPELLING

The SLI group were less accurate when spelling derivational morphemes compared to both control groups, despite being matched for language and spelling with the LA group, and they also made proportionately more phonologically implausible errors. This study confirms previous research that suggested the SLI group might struggle when spelling words containing phonological and orthographic shifts from the base to derived forms Silliman et al. (2006). Children in the control groups were generally making errors in a phonologically plausible manner. In contrast the SLI group were unable to apply phoneme-grapheme correspondences plausibly when attempting to spell the morpheme, e.g., -sed for -tion in attention and -ets for -ity in majority.

However, despite these differences in accuracy and error type the SLI group and their LA matches showed similar links between derivational morpheme spelling, phonological awareness and word reading. However, the poorer phonological and reading skills of the SLI group did not allow them to match the performance of the LA group for these more challenging derivations. It could be hypothesized that children with SLI are displaying a difficulty with the semantic links between language and spelling in relation to these derivational morphemes. However, the fact that they achieved parity on the derivational morphological awareness task with the LA group might rule that out. Instead it might be more plausible to suggest that the lower phonological and reading abilities the children with SLI are being more strongly highlighted when the difficulty of the bound morpheme spelling demands increase, showing a specific impairment in the underlying representations of these derivational morphemes.

The older CA group showed a different pattern of relationships whereby successful derivational morpheme spelling was related to derivational morphological awareness and not phonological awareness. They were showing a close link between a complex language task and their spelling ability. The reading skills of the CA group also showed an association with derivational morphological awareness unlike the children with SLI and the LA group so that derivational morphological awareness may be reliant on an appropriate level of reading.

The regressions provided consistent findings. Out of the four key predictors tested, word reading was the only significant predictor when spelling inflectional and derivational morphemes.

The predominance of word reading confirms findings from studies of general spelling ability (McCarthy et al., 2012) that it is the strength of underlying orthographic representations rather than dimensions of oral language that may primarily determine spelling attainment. This further demonstrates the close developmental relationship between single word reading and spelling (Zutell and Rasinski, 1989; Swanson et al., 2003).

Inflectional and derivational morphological awareness were not predictive of overall sample performance as had been suggested by some studies of typically developing children (Nagy et al., 2006). However, Nunes and Bryant (2009) argue that explicit understanding and awareness of morphemes may not be crucial for correctly spelling all morphemes and that it can often be achieved by word specific knowledge and in appropriate instances, by the application of phoneme-grapheme correspondences. Therefore, like expressive oral language ability, morphological awareness may have more of an impact later in development. At this point, bound morpheme spelling for the SLI and LA groups is determined by orthographic representations and most likely their connections to phonological awareness rather than morphological awareness. It is also likely that children will revert to phonological strategies if there is any uncertainty when spelling derivational morphemes as these are more challenging for all children in this age range, not just the SLI sample.

### 4.3. LIMITATIONS AND FUTURE DIRECTIONS

Language is a complex skill and the current study used one measure of expressive language to evaluate performance in this area. While there were strong theoretical and empirical reasons to use the expressive language task it could be that this test was not sensitive enough to tap into connections between oral language and morphological spelling. Furthermore, it could be argued that morphological skills also reflect receptive and expressive vocabulary given the hypothesized semantic link to derived morphemes in particular. Thus, consideration of the breadth and depth of children's vocabulary levels at different developmental phases in relation to spelling would further our understanding of the relationships between morphological spelling ability and semantic representations.

Similarly consideration should also be given to the way that inflectional and derivational morphological awareness is measured. We have already outlined the rationale for the tasks that were used, however there is some concern about the possible ceiling effects in both tasks and small effect sizes for the group differences and therefore future tasks could utilize words/bound morphemes that link directly to those included in the spelling tasks. Given the semantic aspect of bound morphemes (derivational morphemes in particular) it would also be interesting to compare spelling of the words in isolation and within sentences to examine contextual influences.

Another possible limitation of this study was the focus on the spelling performance of the bound morphemes specifically, rather than an examination of the ability to spell the root or base word in comparison to the inflected and derived forms. This is particularly pertinent for interpreting our derivational shift word findings as recent work examining typically developing children shows that accuracy when reading derived forms is determined by accuracy when reading the root words (Goodwin et al., 2013). However, other recent research has also shown, as we have, that the spelling of root words by children with SLI are consistent across both root and derived forms and are no worse that spelling matched children (Deacon et al., 2014). Nonetheless, further study examining the frequency of the roots and derived forms and the degree of phonological orthographic and semantic opacity would be very useful.

In conclusion, the current study has demonstrated the importance of reading skills in the spelling performance of typically developing children and those with SLI. Further we have shown that inflectional and derivational spelling may provide a window into the spelling difficulties experienced by pupils with SLI. Further research should examine these conclusions with children at different phases of spelling development, more elaborate measures of morphological awareness and a consideration of the relationship between the root words and the derived forms.

### SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fpsyg.2014.00948/abstract

### REFERENCES

Apel, K., Wilson-Fowler, E. B., Brimo, D., and Perrin, N. A. (2012). Metalinguistic contributions to reading and spelling in second and third grade students. *Read. Writ.* 25, 1283–1305. doi: 10.1007/s11145-011-9317-8

Bahr, R. H., Silliman, E. R., Berninger, V. W., and Dow, M. (2012). Linguistic pattern analysis of misspellings of typically developing writers in Grades 1-9. *J. Speech Lang. Hear. Res.* 55, 1587–1599. doi: 10.1044/1092-4388

Bishop, D. V. M., and Clarkson, B. (2003). Written language as a window into residual language deficits: a study of children with persistent and residual speech and language impairments. *Cortex* 39, 215–237. doi: 10.1016/S0010-9452(08)70106-0

Bishop, D. V. M., and Snowling, M. J. (2004). Developmental dyslexia and Specific Language Impairment: same or different? *Psychol. Bull.* 130, 858–886. doi: 10.1037/0033-2909.130.6.858

Botting, N., Simkin, Z., and Conti-Ramsden, G. (2006). Associated reading skills in children with a history of Specific Language Impairment (SLI). *Read. Writ.* 19, 77–98. doi: 10.1007/s11145-005-4322-4

Cordewener, K. A. H., Bosman, A. M. T., and Verhoeven, L. (2012). Specific language impairment affects the early spelling process quantitatively but not qualitatively. *Res. Dev. Disabil.* 33, 1041–1047. doi: 10.1016/j.ridd.2012.01.011

Critten, S., Pine, K. J., and Steffler, D. (2007). Spelling development in young children: a case of Representational-Redescription? *J. Educ. Psychol.* 99, 207–220. doi: 10.1037/0022-0663.99.1.207

Deacon, S. H., Cleave, P. L., Baylis, J., Fraser, J., Ingram, E., and Perlmutter, S. (2014). The representation of roots in the spelling of children with Specific Language Impairment. *J. Learn. Disabil.* 47, 13–21. doi: 10.1177/0022219413509965.

Dockrell, J. E., and Connelly, V. (2013). The role of oral language in underpinning the text generation difficulties in children with specific language impairment. *J. Res. Read.* doi: 10.1111/j.1467-9817.2012.01550.x. [Epub ahead of print].

Dockrell, J. E., Lindsay, G., and Connelly, V. (2009). The impact of Specific Language Impairment on adolescents' written text. *Except. Child.* 75, 427–446. doi: 10.1177/001440290907500403

Dockrell, J. E., Lindsay, G., Connelly, V., and Mackie, C. (2007). Constraints in the production of written text in children with specific language impairments. *Except. Child.* 73, 147–164. doi: 10.1177/001440290707300202

Ehri, L. C. (2005). "Development of sight word reading: phases and findings," in *The Science of Reading: A Handbook*, eds M. J. Snowling, and C. Hulme (Oxford: Blackwell Publishing), 135–154.

Elliott, C. D., Murray, D. J., and Pearson, L. S. (1997). *British Ability Scales II: Matrices*. Windsor: NFER-Nelson.

Fraser, J., Goswami, U., and Conti-Ramsden, G. (2010). Dyslexia and specific language impairment: the role of phonology and auditory processing. *Sci. Stud. Read.* 14, 8–29. doi: 10.1080/10888430903242068

Frederickson, N., Frith, U., and Reason, R. (1997). *The phonological assessment battery. Standardisation Edition*. Slough: NFER-Nelson.

Goodwin, A. P., Gilbert, J. K., and Cho, S. (2013). Morphological contributions to adolescent word reading: an item response approach. *Read. Res. Q.* 48, 39–60. doi: 10.0002/rrq.037

Green, L., McCutchen, D., Schwiebert, C., Quinlan, T., Eva-Wood, A., and Juelis, J. (2003). Morphological development in children's writing. *J. Educ. Psychol.* 95, 752–761. doi: 10.1037/0022-0663.95.4.752

Kemp, N., and Bryant, P. (2003). Do beez buzz? Rule-based and frequency-based knowledge in learning to spell plural -s. *Child Dev.* 74, 63–74. doi: 10.1111/1467-8624.00521

Larkin, R. F., Williams, G. J., and Blaggan, S. (2013). Delay or deficit? Spelling processes in children with specific language impairment. *J. Commun. Disord.* 46, 401–412. doi: 10.1016/j.jcomdis.2013.07.003

Mackie, C., and Dockrell, J. E. (2004). The nature of written language deficits in children with SLI. *J. Speech Lang. Hear. Res.* 47, 1469–1483. doi: 10.1044/1092-4388(2004/109)

Mackie, C. J., Dockrell, J., and Lindsay, G. (2013). An evaluation of the written texts of children with SLI: the contributions of oral language, reading and phonological short-term memory. *Read. Writ.* 26, 865–888. doi: 10.1007/s11145-012-9396-1

Marshall, C. R., and Van Der Lely, H. J. K. (2007). The impact of phonological complexity on past tense inflection in children with Grammatical-SLI. *Adv. Speech Lang. Pathol.* 9, 191–203. doi: 10.1080/14417040701261509

Masterson, J., Stuart, M., Dixon, M., and Lovejoy, S. (2003). *Children's Printed Word Database*. ESRC research project. Available online at: http://www.essex.ac.uk/psychology/cpwd/

McCarthy, J. H., Hogan, T. P., and Catts, H. W. (2012). Is weak oral language associated with poor spelling in school-age children with specific language impairment, dyslexia or both? *Clin. Linguist. Phon.* 26, 791–805. doi: 10.3109/02699206.2012.702185

McCutchen, D., Stull, S., Herrera, B. L., Lotas, S., and Evans, S. (2014). Putting words to work: effects of morphological instruction on children's writing. *J. Learn. Disabil.* 47, 86–97. doi: 10.1177/0022219413509969

Montgomery, J. W., and Leonard, L. B. (1998). Real-time inflectional processing by children with Specific Language Impairment: effects of phonetic substance. *J. Speech Lang. Hear. Res.* 41, 1432–1443. doi: 10.1044/jslhr.4106.1432

Mossing, C. R., Diehl, S. F., Bahr, R. H., and Silliman, E. R. (2009). *The Morphological Complexity of Spelling, Ages 8 to 15 Years*. Unpublished master's thesis. University of South Florida.

Nagy, W., Berninger, V. W., and Abbott, R. D. (2006). Contributions of morphology beyond phonology to literacy outcomes of upper elementary and middle-school students. *J. Educ. Psychol.* 98, 134–147. doi: 10.1037/0022-0663.98.1.134

Nation, K., and Snowling, M. J. (1998). Semantic processing and the development of word-recognition skills: evidence from children with reading comprehension difficulties. *J. Mem. Lang.* 39, 85–101. doi: 10.1006/jmla.1998.2564

Neale, M. D., Christophers, U., and Whetton, C. (1997). *Neale Analysis of Reading Ability—Revised British Edition*. Windsor: NFER-Nelson.

Nunes, T., Bindman, M., and Bryant, P. (1997). Morphological spelling strategies: developmental stages and processes. *Dev. Psychol.* 33, 637–649. doi: 10.1037/0012-1649.33.4.637

Nunes, T., and Bryant, P. (2009). *Children's Reading and Spelling: Beyond the First Steps*. Singapore: Wiley-Blackwell.

Oetting, J. B., and Hadley, P. (2009). "Morphosyntax in child language disorders," in *Handbook of Child Language Disorders*, ed R. G. Schwartz (New York, NY; Hove: Psychology Press), 341–364.

Ouellette, G. P., and Sénéchal, M. (2008). A window into early literacy: exploring the cognitive and linguistic underpinnings of invented spelling. *Sci. Stud. Read.* 12, 195–219. doi: 10.1080/10888430801917324

Semel, E., Wiig, E. H., and Secord, W. A. (2006). *Clinical Evaluation of Language Fundamentals, 4th Edn; UK Standardisation (CELF-4 UK)*. San Antonio, TX: The Psychological Corporation/A Harcourt Assessment Company.

Silliman, E. R., Bahr, R. H., and Peters, M. L. (2006). Spelling patterns in preadolescents with atypical language skills: phonological, morphological, and orthographic factors. *Dev. Neuropsychol.* 29, 93–123. doi: 10.1207/s15326942dn2901_6

Smith-Lock, K. M. (1995). Morphological usage and awareness in children with and without specific language impairment. *Ann. Dyslexia* 45, 161–185. doi: 10.1007/BF02648217

Snowling, M. J., Stothard, S. E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., et al. (2009). *YARC York Assessment of Reading for Comprehension: Passage Reading*. London: GL Assessment.

Swanson, H. L., Trainin, G., Necoechea, D. M., and Hammill, D. D. (2003). Rapid naming, phonological awareness, and reading: a meta-analysis of the correlation evidence. *Rev. Educ. Res.* 73, 407–440. doi: 10.3102/00346543073004407

Wagner, R. K., Torgeson, J. K., and Rashotte, C. A. (1999). *The Comprehensive Test of Phonological Processing*. Austin, TX: PsychCorp.

Wiggins, K. I., Diehl, S. F., Bahr, R. H., and Silliman, E. R. (2010). *Spelling Errors in Children with Autism*. Unpublished master's thesis. University of South Florida.

Windsor, J., Scott, C. M., and Street, C. K. (2000). Verb and noun morphology the spoken and written language of children with language learning disabilities. *J. Speech Lang. Hear. Res.* 43, 1322–1336. doi: 10.1044/jslhr.4306.1322

Zutell, J., and Rasinski, T. (1989). Reading and spelling connections in third and fifth grade students. *Read. Psychol.* 10, 137–155. doi: 10.1080/0270271890100203

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

# Educational attainment in poor comprehenders

Jessie Ricketts[1]*, Rachael Sperring[1] and Kate Nation[2]

[1] Institute of Education, University of Reading, Reading, UK
[2] Department of Experimental Psychology, University of Oxford, Oxford, UK

To date, only one study has investigated educational attainment in poor (reading) comprehenders, providing evidence of poor performance on national UK school tests at age 11 years relative to peers (Cain and Oakhill, 2006). In the present study, we adopted a longitudinal approach, tracking attainment on such tests from 11 years to the end of compulsory schooling in the UK (age 16 years). We aimed to investigate the proposal that educational weaknesses (defined as poor performance on national assessments) might become more pronounced over time, as the curriculum places increasing demands on reading comprehension. Participants comprised 15 poor comprehenders and 15 controls; groups were matched for chronological age, nonverbal reasoning ability and decoding skill. Children were identified at age 9 years using standardized measures of nonverbal reasoning, decoding and reading comprehension. These measures, along with a measure of oral vocabulary knowledge, were repeated at age 11 years. Data on educational attainment were collected from all participants ($n = 30$) at age 11 and from a subgroup ($n = 21$) at 16 years. Compared to controls, educational attainment in poor comprehenders was lower at ages 11 and 16 years, an effect that was significant at 11 years. When poor comprehenders were compared to national performance levels, they showed significantly lower performance at both time points. Low educational attainment was not evident for all poor comprehenders. Nonetheless, our findings point to a link between reading comprehension difficulties in mid to late childhood and poor educational outcomes at ages 11 and 16 years. At these ages, pupils in the UK are making key transitions: they move from primary to secondary schools at 11, and out of compulsory schooling at 16.

**Keywords: poor comprehenders, educational attainment, reading comprehension, specific reading comprehension impairment, oral vocabulary**

## INTRODUCTION

In the early stages of learning to read, children must learn to map letters onto sounds so that they can decode and recognize words. However, the ultimate goal of reading is to understand the messages conveyed by text; simply being able to read words and texts accurately is not sufficient for comprehension to occur. A substantial number of children (∼8% in UK studies; Clarke et al., 2010) show reading comprehension impairments despite age-appropriate word recognition skills; these children are typically referred to as "poor comprehenders" or "children with specific reading comprehension impairments." Research conducted in Italy, the UK and the US has made good progress with understanding the cognitive and linguistic profiles that characterize poor comprehenders in mid to late childhood (e.g., poor oral language, poor inferential skills; for reviews, see Nation, 2005; Floyd et al., 2006; Cain and Oakhill, 2007; Carretti et al., 2009) but we know very little about the progress that such children make in adolescence, and at school. We conducted a longitudinal study tracking reading, vocabulary, and educational attainment in poor comprehenders over the course of eight years: from age 9 to 16 years. Educational attainment was indexed through performance on national UK school assessments at the end of primary school (11 years) and at the end of compulsory

education (16 years)[1]. Given that poor comprehenders struggle to learn from what they read (Cain et al., 2004; Ricketts et al., 2008), and that acquiring knowledge through the process of reading becomes an increasingly important learning strategy as children move through the school system, it seems likely that poor comprehenders will be at a disadvantage at school. Despite the likely educational consequences of the reading comprehension difficulties experienced by poor comprehenders, their difficulties may be masked by good reading accuracy in the classroom (Nation and Angell, 2006; Hulme and Snowling, 2011), and only one study to date has investigated educational attainment in this group (Cain and Oakhill, 2006).

Research with poor comprehenders has shed light on the factors, beyond word recognition, that support successful reading comprehension, particularly focussing on oral language (e.g., Catts et al., 2006; Nation et al., 2010), discourse level processes such as inference generation and comprehension monitoring (e.g., Oakhill and Cain, 2012) and executive functions such as working memory (e.g., Carretti et al., 2009). Longitudinal data and intervention studies provide particularly convincing evidence for

---

[1] Note that since this study was conducted the age at which compulsory schooling ends in the UK has been raised from 16 to 17 years.

causal relationships. However, there is a dearth of longitudinal and intervention research with poor comprehenders. Nonetheless, existing longitudinal studies indicate that poor oral language can be observed in poor comprehenders before their reading comprehension difficulties are identified, suggesting that oral language weaknesses precede (and therefore may cause) their reading comprehension difficulties. In a US study, Catts et al. (2006) selected 57 poor comprehenders in eighth Grade (14 years) and looked retrospectively at their oral language skills in Kindergarten, second Grade and fourth Grade (age 6, 8, and 10 years, respectively). Poor comprehenders performed more poorly than typically developing readers on a language comprehension composite at each time point. In the UK, Nation et al. (2010) conducted a prospective longitudinal study, assessing oral language and reading in 242 children for the first time at age 5 years and following children over time until poor comprehenders ($n = 15$) could be reliably identified at age 8 years. Again, weaknesses in oral language comprehension were detected earlier in time, when children had experienced very little reading instruction. In the only randomized controlled trial conducted with poor comprehenders to date, Clarke et al. (2010) showed significant improvements in reading comprehension scores following an oral language intervention program, concluding that oral language weaknesses play a causal role in determining the reading comprehension difficulties that are experienced by poor comprehenders (aged 8–9 years). At present however, we know very little about poor comprehenders later in development, as they transition to secondary school and beyond.

The idea that oral language skills such as vocabulary and grammar provide a foundation for successful reading comprehension is embodied by the Simple View of Reading (Gough and Tunmer, 1986; Tunmer and Chapman, 2012), a key theoretical framework that has been used to conceptualize reading development and reading difficulties. On this view, word recognition and oral language comprehension are separable variables that underpin reading comprehension, and both are necessary for successful reading (reading for meaning). Substantial support for the assumptions of the Simple View derive from a wide range of empirical approaches, including longitudinal research with typically developing children (e.g., Oakhill et al., 2003; Muter et al., 2004), the study of children with specific reading difficulties (e.g., Catts et al., 2006), behavioral genetics (e.g., Harlaar et al., 2010) and factor analysis (e.g., Tunmer and Chapman, 2012). Despite its wide use in reading research, the Simple View is not without its critics. Notable are arguments that the word recognition and oral language comprehension components of the Simple View are poorly specified, that they are not entirely independent, and that reading comprehension involves more than just these components (e.g., Kirby and Savage, 2008; Ouellette and Beers, 2010; Tunmer and Chapman, 2012; Ricketts et al., 2013).

The relationship between oral language and reading is reciprocal, with reading activities providing important opportunities for growth in aspects of oral language such as vocabulary knowledge (e.g., Nagy et al., 1985). Importantly, the extent to which children learn new words while reading will depend on their reading proficiency (e.g., Ricketts et al., 2011). Poor comprehenders show particular difficulty learning and retaining the meanings of novel words from context (Cain et al., 2003, 2004; Ricketts et al., 2008), suggesting that slowed growth in vocabulary (Matthew effects) is a possibility in this group. As mentioned above, few studies have tracked development in poor comprehenders (for a summary of existing studies, see Elwér et al., 2013). Nonetheless, the longitudinal work of Cain and Oakhill (2011) lends support to the hypothesis that poor comprehenders show Matthew effects for vocabulary. Matthew effects refer to the widening of gaps between low and high achievers over time (Stanovich, 1986). Cain and Oakhill (2011) assessed reading and receptive vocabulary in 17 poor comprehenders and 14 good comprehenders at ages 8 and 11 years. Using Scarborough and Parker's (2003) ANOVA approach for detecting Matthew effects, Cain and Oakhill (2011) demonstrated slowed receptive vocabulary growth in poor relative to good comprehenders. In contrast, differences between groups were relatively constant over time for reading comprehension, indicating persistent reading comprehension impairments in the poor comprehenders (see also Cain and Oakhill, 2006).

Reading for meaning provides not only important opportunities for the acquisition of vocabulary and other aspects of language, but also for learning more generally. As mentioned above, it is likely that reading comprehension impairments will be associated with poor educational outcomes and yet only one UK-based study has explored educational attainment in poor comprehenders. In the UK, children complete national School Assessment Tests (SAT-UK) tests at 11 years, just before they transition from primary to secondary school. Currently, SAT-UK tests focus on English and maths curriculum subjects, but in the past science was also examined. Cain and Oakhill (2006) reported data from SAT-UK tests for 16 poor comprehenders and 17 good comprehenders who had been identified 3 years earlier (age 8 years) from UK primary schools. Cain and Oakhill (2011) found that group means for poor and good comprehender groups were in line with government targets (a level 4). However, the good comprehender group obtained a significantly higher mean score than the poor comprehender group on English, maths, and science SAT-UK tests. Thus, Cain and Oakhill's (2011) study indicates that, on average, poor comprehenders attain at an age-appropriate level at age 11 years. However, they are at a disadvantage in comparison to peers without a history of reading comprehension difficulty.

The primary aim of the present study was to investigate educational attainment in poor comprehenders. To this aim, we collected longitudinal data over a period of 8 years, identifying poor comprehenders and age-matched controls without reading comprehension difficulties at age 9 years, and recording their performance in national UK school assessments at the end of primary school (SAT-UK tests at 11 years) and at the end of compulsory education (16 years). At 16 years, pupils in the UK sit General Certificate of Secondary Education (GCSE) tests and equivalents; the present study investigates GCSE attainment in poor comprehenders for the first time (for studies on GCSE performance of children with a history of primary language impairment, see Snowling et al., 2001; Dockrell et al., 2011). Both SAT-UK and GCSEs are described in more detail later in this paper. Based on Cain and Oakhill (2006), we anticipated that as a group, poor comprehenders' SAT-UK attainment would be in line with national

norms but that poor comprehenders would perform more poorly than controls.

We sought to build on Cain and Oakhill's (2006) study in two ways. First, Cain and Oakhill (2011) did not report individual scores on SAT-UK tests. Given the heterogeneous nature of poor comprehender groups (Nation et al., 2002; Cain and Oakhill, 2006; Floyd et al., 2006), we sought to examine individual profiles to ascertain whether there are poor comprehenders who are attaining below national expectations as they transition from the primary school curriculum to its more demanding secondary counterpart. Second, we collected data on national assessments at the end of compulsory schooling in the UK to investigate longer term educational outcomes for children who had been identified as poor comprehenders in middle childhood. We anticipated that as the curriculum places greater demands on reading comprehension, group differences in attainment might become more pronounced and that later in the educational system poor comprehenders might show evidence of falling behind government targets.

Measures of reading comprehension and expressive oral vocabulary were administered at ages 9 and 11 years. Therefore, in addition to exploring educational progress, we sought to replicate studies showing that the reading comprehension difficulties experienced by poor comprehenders are persistent over time (Cain and Oakhill, 2006, 2011) and to investigate oral vocabulary development in this group. Given evidence for poor vocabulary learning (Cain et al., 2003, 2004; Nation et al., 2007; Ricketts et al., 2008) and slowed receptive vocabulary development in poor comprehenders (Cain and Oakhill, 2011), we expected to see Matthew effects for vocabulary.

To our knowledge, the present study is the first of its kind, tracking development in poor comprehenders over a particularly long timeframe: from identification at age 9 years to adolescence (16 years). By considering reading and vocabulary at 9 years (Time 1) and 11 years (Time 2), and attainment as measured by UK national school assessments at 11 years (Time 2) and 16 years (Time 3), we sought to address the following key research questions:

1. Do poor comprehenders show low educational attainment at ages 11 and 16 years compared to controls (typically developing readers) matched for age, nonverbal reasoning, and decoding (nonword reading)?
2. Do poor comprehenders show poor educational attainment at ages 11 and 16 years as compared to the attainment of pupils nationally in the UK?
3. Are the reading comprehension difficulties experienced by poor comprehenders stable over time?
4. Do poor comprehenders show Matthew effects for vocabulary (i.e., slowed growth)?

## MATERIALS AND METHODS
### PARTICIPANTS
Participants were 15 poor comprehenders and 15 controls drawn from a sample of 81 children who were attending mainstream schools that serve socially mixed catchment areas in the UK. None of the larger sample of 81 children spoke English as an additional language or had any recognized special educational need. Participants for each group were selected according to the following

criteria. Poor comprehenders obtained reading comprehension standard scores of at least one standard deviation below the test mean ($\leq$85) and controls' scores were well into the average range or above (>95). Groups were matched for chronological age, nonverbal reasoning ability and decoding (nonword reading) skill, with all children performing within the average range (or above) on nonverbal reasoning and decoding tasks. Groups were also matched for gender, with 11 girls and 4 boys in each group. Details of all measures are included below and performance of both groups is summarized in **Table 1**. Ethical approval for the study was obtained from the University of Oxford (Time 1 and Time 2) and University of Reading (Time 3) Research Ethics Committees.

## MATERIALS AND PROCEDURE
Poor comprehenders and controls were identified at Time 1 using the standardized measures of nonverbal reasoning, decoding, and reading comprehension outlined below. These measures, along with a measure of oral vocabulary knowledge, were repeated at Time 2, approximately 2 years later ($M$ time difference = 2.08 years, $SD$ = 0.12, range: 1.83–2.29). Note that participants completed other tasks in between these two testing points, which are reported elsewhere (Ricketts et al., 2007, 2008). All standardized measures were administered according to manual instructions. Data on educational attainment were collected at the end of primary school (Time 2) and approximately 5 years later at the end of compulsory education (Time 3).

### Nonverbal reasoning
Nonverbal reasoning was measured using the Matrix Reasoning subtest of the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999). This subtest assesses nonverbal reasoning using a pattern completion task in which participants are provided with a pattern that has a piece missing; their task is to select the missing piece from an array of five. WASI subtests yield a $t$-score ($M$ = 50, $SD$ = 10); for comparison with other measures, this was transformed into a standard score ($M$ = 100, $SD$ = 15). The WASI provides norms for individuals aged 6–89 years, and high internal consistency (split half reliability) is reported in the manual ($r$ = 0.86–0.96, depending on age group).

### Oral vocabulary
Oral vocabulary knowledge was measured using the Vocabulary subtest of the WASI (Wechsler, 1999). This is a measure of expressive vocabulary in which children are asked to verbally define words. Scores capture both depth and breadth of word knowledge, indexing the incremental nature of oral vocabulary knowledge. WASI subtests yield a $t$-score ($M$ = 50, $SD$ = 10); for comparison with other measures, this was transformed into a standard score ($M$ = 100, $SD$ = 15). The WASI provides norms for individuals aged 6–89 years, and high internal consistency (split half reliability) is reported in the manual ($r$ = 0.86–0.93, depending on age group).

### Decoding
Decoding (nonword reading) was assessed using the phonemic decoding efficiency (PDE) subtest of the test of word reading efficiency (TOWRE; Torgesen et al., 1999). In this test, children are

**Table 1 | Summary of performance on selection measures, follow up measures and oral vocabulary at Time 1 and Time 2.**

| Time point/measure | Poor comprehenders (n = 15) | | Controls (n = 15) | | Group comparisons | | |
|---|---|---|---|---|---|---|---|
| | M | SD | M | SD | F | p | Cohen's d |
| **Time 1 selection** | | | | | | | |
| Chronological age[1] | 9.21 | 0.30 | 9.26 | 0.28 | 0.22 | 0.64 | 0.17 |
| Nonverbal reasoning[2] | 103.50 | 7.54 | 103.50 | 6.43 | 0.00 | 1.00 | 0.00 |
| Decoding[2] | 107.67 | 13.11 | 108.27 | 9.68 | 0.02 | 0.88 | 0.05 |
| Reading comprehension[2] | 81.93 | 2.69 | 103.13 | 4.88 | 217.14 | <0.001 | 5.60 |
| **Time 2 follow up** | | | | | | | |
| Chronological age[1] | 11.30 | 0.31 | 11.34 | 0.32 | 0.14 | 0.71 | 0.12 |
| Nonverbal reasoning[2] | 101.90 | 6.98 | 103.50 | 8.63 | 0.31 | 0.58 | 0.20 |
| Decoding[2] | 106.40 | 14.94 | 107.27 | 10.66 | 0.03 | 0.86 | 0.07 |
| Reading comprehension[2] | 83.60 | 4.44 | 95.87 | 7.57 | 29.33 | <0.001 | 2.04 |
| **Oral vocabulary** | | | | | | | |
| Time 1[2] | 89.20 | 13.63 | 110.00 | 8.09 | 25.82 | <0.001 | 1.92 |
| Time 2[2] | 91.50 | 11.31 | 109.50 | 6.12 | 29.40 | <0.001 | 2.07 |

[1]Years; [2]Standard scores (M = 100, SD = 15).

asked to read a list of nonwords of increasing length and difficulty as quickly as they can. Efficiency is indexed by the number of nonwords decoded correctly in 45 s. The TOWRE produces standard scores (M = 100, SD = 15). The test provides norms for individuals aged 6–24 years, and its manual indicates a high level of test/re-test reliability (r = 0.89–0.91, depending on age group).

### Reading comprehension
Reading comprehension was assessed using the Neale Analysis of Reading Ability-II (NARA-II; Neale, 1997). In the NARA-II children read aloud passages of connected text and then answer comprehension questions relating to each passage. Some questions can be answered with reference to verbatim memory while others require inferences to be made (Bowyer-Crane and Snowling, 2005). The NARA-II comprises two parallel forms; children completed Form 1 at Time 1 and Form 2 at Time 2 to avoid practice effects. The NARA-II produces standard scores (M = 100, SD = 15) for reading comprehension. The test provides norms for children aged 6–12 years, and shows high internal consistency (Cronbach's α = 0.93–0.95, depending on age group). The manual reports high correlations between comprehension scores on the two parallel forms (r = 0.82).

### Educational attainment
In England, pupils sit national school assessments at the end of primary school at age 11 years (SAT-UK tests) and at the end of compulsory education at age 16 years (GCSEs or qualifications at an equivalent level). At Time 2, participants were in the final year of primary school and at the end of this year schools were contacted to obtain SAT-UK test results. Schools provided the level (from 2 to 5) at which all pupils (n = 30) were performing in English, maths, and science subjects (note that pupils no longer sit SAT-UK tests

for science). English results can be further decomposed into separate scores for reading and writing. Given the reading difficulties observed in the poor comprehenders, reading and writing scores were considered separately. Maths and science scores were considered to aid comparison with an earlier study (Cain and Oakhill, 2006). UK government targets stipulate that in order to be "secondary ready" (have the requisite knowledge and skill to manage the secondary curriculum) pupils should be operating at level 4 or above at the end of primary school. The UK government publishes data each year indicating how many children meet this target (UK Department for Education, 2012a). Not all pupils obtain a level 4 in each subject but the majority do; thus, a level 4 does not represent the average, instead, most children are expected to reach this level.

At Time 3, GCSE (or equivalent) results were obtained via the following process. Some primary schools provided information about secondary school destinations at Time 2. For the remaining participants, primary schools were contacted and asked to provide details of secondary school destinations. The secondary schools that consented to take part in the study distributed information sheets and consent forms to participants and, on the basis of informed consent, released GCSE results to the research team. This process yielded GCSE data for 20/30 participants. One secondary school and one participant did not consent to take part. For some of the remaining participants, home addresses had been provided by parents at Time 1 (but this was not compulsory for inclusion in the study). Where possible, participants for whom GCSE data had not been obtained from schools were contacted directly by post. This resulted in one participant sending information about GCSE results independently. Thus, GCSE results were available for 21/30 participants.

GCSE-level qualifications can be acquired for a wide range of curriculum subjects, including the SAT-UK subjects (English,

maths, science) as well as other subjects (e.g., foreign languages, history, geography, art). Pupils and schools work together to choose the number of qualifications a pupil undertakes and which subjects they study at this level. When GCSEs (or equivalents) are marked, grades are given (A*, A, B–G) that correspond to points (16–58, e.g., $A^* = 58$, $A = 52$, $B = 46$, $C = 40$). Grades fall into two levels, level 2 relates to grades A*–C, and level 1 to grades D–G. Grades and points determine, to some extent, post-16 destinations (further education, apprenticeships, employment opportunities, etc.). When the government reports on attainment for pupils in England at the end of compulsory education, two key variables of interest are whether children obtained five GCSEs (or equivalent) at level 2 (i.e., with grades between A* and C) and whether they have made "expected progress" since taking SAT-UK tests. Expected progress is only recorded for subjects taken at both SAT-UK level (now English, maths) and GCSE (English and maths are compulsory); within each subject this reflects a pupil obtaining a level 3, 4, or 5 at SAT-UK and then at least D, C, or B at GCSE, respectively. The UK government publishes data each year indicating how many children meet these targets (UK Department for Education, 2012b).

## RESULTS

### READING AND VOCABULARY AT TIME 1 AND TIME 2

**Table 1** summarizes age and performance (standard scores) on nonverbal reasoning, decoding, and reading comprehension measures at Time 1 (selection measures) and Time 2 as well as performance on an oral vocabulary measure at Time 1 and Time 2. **Table 1** also includes details of group comparisons (one-way ANOVA) for each variable. In line with selection and matching procedures, groups were closely matched for age, nonverbal reasoning and decoding at Time 1. This close correspondence between the two groups was maintained at Time 2. Groups differed on reading comprehension and oral vocabulary measures at Time 1 and Time 2, with large effect sizes observed (all Cohen's $d \geq 2$).

To investigate Matthew effects, data on reading comprehension and oral vocabulary were analyzed using a series of $2 \times 2$ ANOVAs; in each, group (poor comprehenders vs. controls) was included as an independent samples factor and time (Time 1 vs. Time 2) as a repeated samples factor. Both raw scores and standard scores for each variable (reading comprehension, vocabulary) were analyzed to probe changes in absolute score (number of comprehension questions correct, knowledge of vocabulary items) as well as norm-referenced scores (cf. Scarborough and Parker, 2003; Cain and Oakhill, 2011). Mean raw scores on reading comprehension and oral vocabulary tasks are depicted in **Figures 1A,C**, respectively; mean standard scores appear in **Table 1** but are replicated in **Figures 1B,D** for ease of comparison.

When reading comprehension raw score (max = 44) was the dependent variable (**Figure 1A**), the main effect of group was significant, $F(1,28) = 86.98$, $p < 0.001$, $\eta_p^2 = 0.76$, with controls outperforming poor comprehenders, as was the main effect of time, $F(1,28) = 71.88$, $p < 0.001$, $\eta_p^2 = 0.72$, with higher performance at Time 2. These main effects were qualified by a significant group x time interaction, $F(1,28) = 17.75$, $p < 0.001$, $\eta_p^2 = 0.39$. Tests of simple effects with Bonferroni correction revealed that
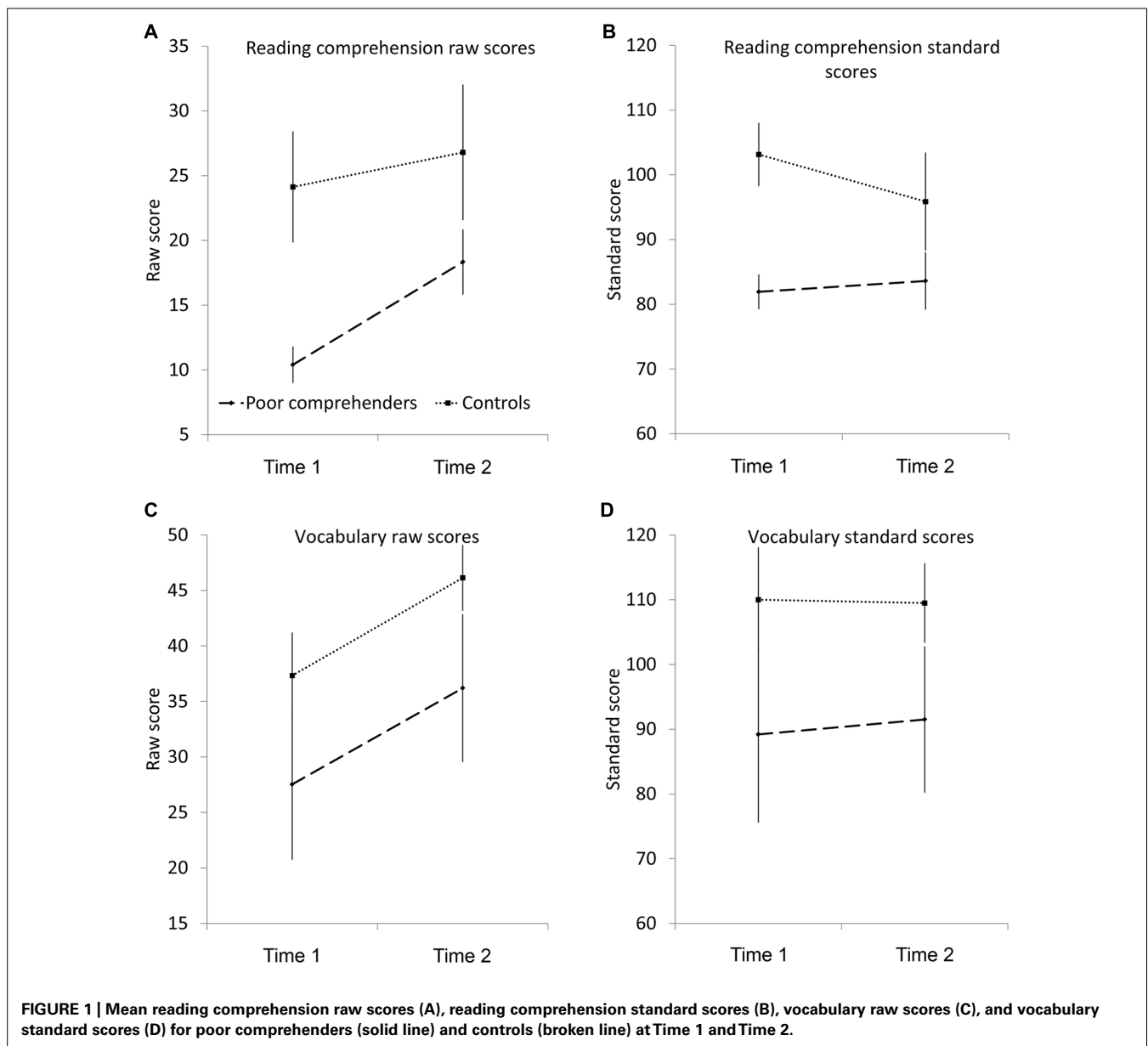
both groups showed a significant increase in raw score over time, but the poor comprehender group showed greater improvement. There were significant group differences in raw score at both time points but this was more marked at Time 1. When reading comprehension standard score was the dependent variable (**Figure 1B**) the main effects of group, $F(1,28) = 105.01$, $p < 0.001$, $\eta_p^2 = 0.79$, and time, $F(1,28) = 8.40$, $p < 0.01$, $\eta_p^2 = 0.23$, were also significant. Again, main effects were qualified by a significant group $\times$ time interaction, $F(1,28) = 21.37$, $p < 0.001$, $\eta_p^2 = 0.43$. Tests of simple effects with Bonferroni correction revealed that for the control group there was a significant decrease in the mean reading comprehension standard scores between Time 1 and Time 2, indicating that for this group reading comprehension performance was not developing in line with cross-sectional data from the test's normative sample. As would be expected from the test norms, means were stable across time (did not change significantly) for the poor comprehender group.

In line with our aim to consider development at the individual level, changes in individual reading comprehension scores are depicted in **Figure 2A** for reference. At Time 2, eight of the 15 poor comprehenders (53%) obtained reading comprehension standard scores that were at least one standard deviation below the test mean; all of these children still met the strict identification criteria adopted at Time 1 (see above). The remaining seven poor comprehenders obtained reading comprehension standard scores that were slightly greater than 85. At Time 2, most poor comprehenders still showed the large discrepancy between advanced decoding and lower reading comprehension that characterizes the poor comprehender profile ($M$ discrepancy = 22.80, $SD = 16.25$). One participant in the control group also met poor comprehender criteria at Time 2.

**Figures 1C,D** shows mean oral vocabulary raw scores (max = 80) and standard scores for poor comprehenders and controls at Time 1 and Time 2. The $2 \times 2$ ANOVA with oral vocabulary raw score as the dependent variable revealed significant main effects of group, $F(1,28) = 31.04$, $p < 0.001$, $\eta_p^2 = 0.53$, and time, $F(1,28) = 113.65$, $p < 0.001$, $\eta_p^2 = 0.80$, with controls outperforming poor comprehenders and higher performance at Time 2 than Time 1. The group $\times$ time interaction was not significant, $F(1,28) = 0.68$, $p = 0.42$, $\eta_p^2 = 0.02$, consistent with the parallel lines in **Figure 1**. With oral vocabulary standard score as the dependent variable, again there was a significant main effect of group, $F(1,28) = 34.14$, $p < 0.001$, $\eta_p^2 = 0.55$, but the main effect time, $F(1,28) = 0.28$, $p = 0.60$, $\eta_p^2 = 0.01$, and group $\times$ time interaction, $F(1,28) = 0.68$, $p = 0.42$, $\eta_p^2 = 0.02$, were not significant. Changes in individual vocabulary scores are depicted in **Figure 2B**.

### EDUCATIONAL ATTAINMENT AT 11 YEARS (TIME 2)

Government targets stipulate that children should be performing at or above level 4 in SAT-UK tests upon leaving primary education. In order to explore whether poor comprehenders show poor educational attainment at this point, the percentage of children in this group obtaining a level 4 across reading, writing, science and maths tests was compared to (1) the control group, and (2) national data. National data

**FIGURE 1 | Mean reading comprehension raw scores (A), reading comprehension standard scores (B), vocabulary raw scores (C), and vocabulary standard scores (D) for poor comprehenders (solid line) and controls (broken line) at Time 1 and Time 2.**

(UK Department for Education, 2012a) refer to children in England completing SAT-UK tests during the same year in which the present participants completed these tests (total $n \approx 584,500$). **Figure 3** illustrates the percentage of children in the poor comprehender group, control group and nationally who achieved a level 4 or above in reading (10 poor comprehenders: 67%, 15 controls: 100%, national data: 84%), writing (9 poor comprehenders: 60%, 15 controls: 100%, national data: 67%), science (12 poor comprehenders: 80%, 15 controls: 100%, national data: 88%), and maths (11 poor comprehenders: 73%, 12 controls: 80%, national data: 77%).

All participants in the control group performed at or above a level 4 in reading, writing, and science (but not maths). In each subject however, a lower number of poor comprehenders achieved a level 4 or above in comparison to controls, and this difference

was most marked for the reading and writing tests. Fisher's exact tests (all 2-tailed) showed that there was a significant association between comprehension group (poor comprehenders vs. controls) and attainment (below level 4 vs. level 4 or higher) for reading ($p = 0.04$) and writing ($p = 0.02$), but not for science ($p = 0.22$) or maths ($p = 1.00$). When comparing the poor comprehender group to national data, a lower percentage of poor comprehenders achieved a level 4 or above across all subjects; Fisher's exact tests revealed that the association between group (poor comprehender vs. national) and attainment (below level 4 vs. level 4 or higher) was significant for reading ($p = 0.01$) but not writing ($p = 0.39$), science ($p = 0.17$) or maths ($p = 0.50$). Finally, a higher percentage of the control group achieved a level 4 or above in comparison to the national data; Fisher's exact tests revealed that the association between group (control vs. national) and attainment (below level
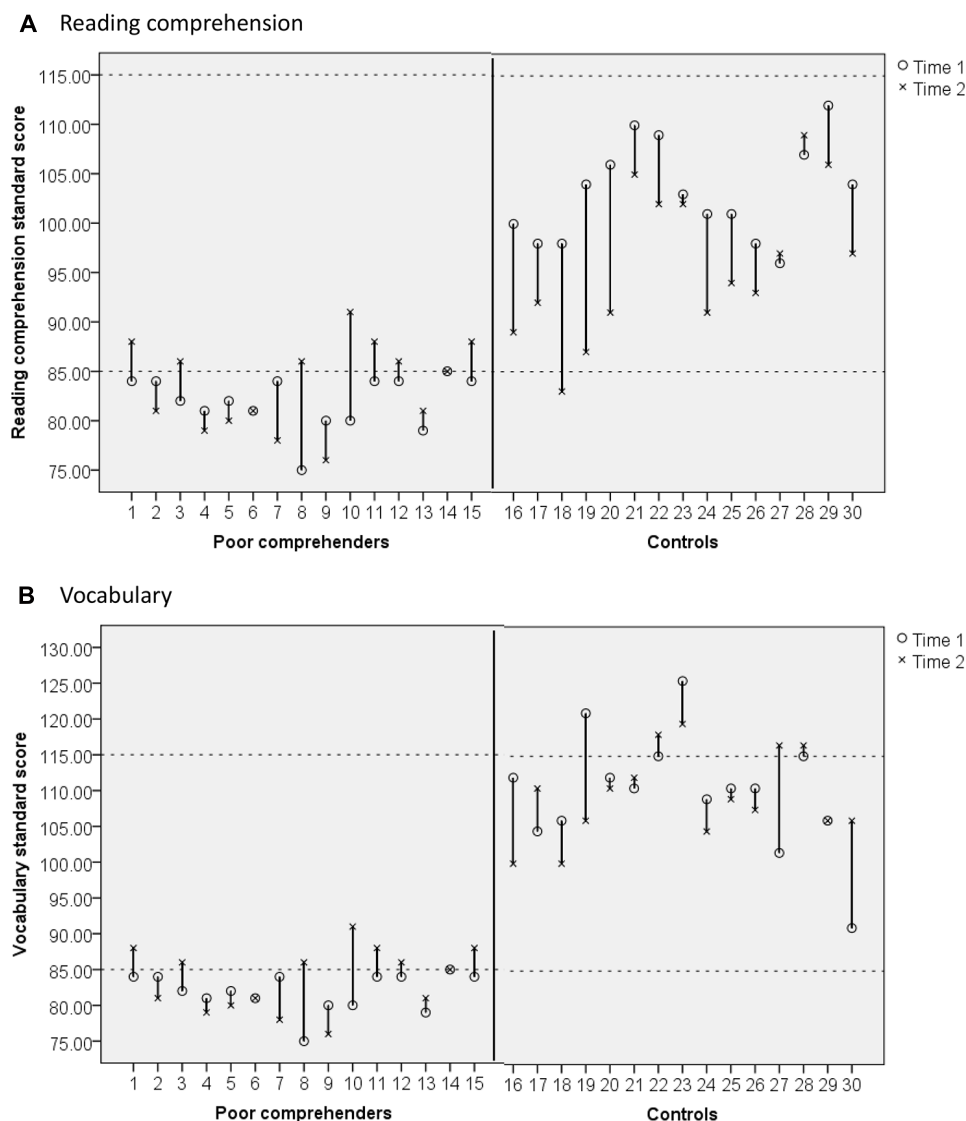
**FIGURE 2 | Change in (A) reading comprehension standard scores and (B) oral vocabulary standard scores between Time 1 (circles) and Time 2 (crosses) in each participating poor comprehender (participants 1–15) and control (participants 16–30).** Participant numbers are equivalent across **(A,B)** such that participant 1 in **(A)** is the sample child as participant 1 in **(B)**.
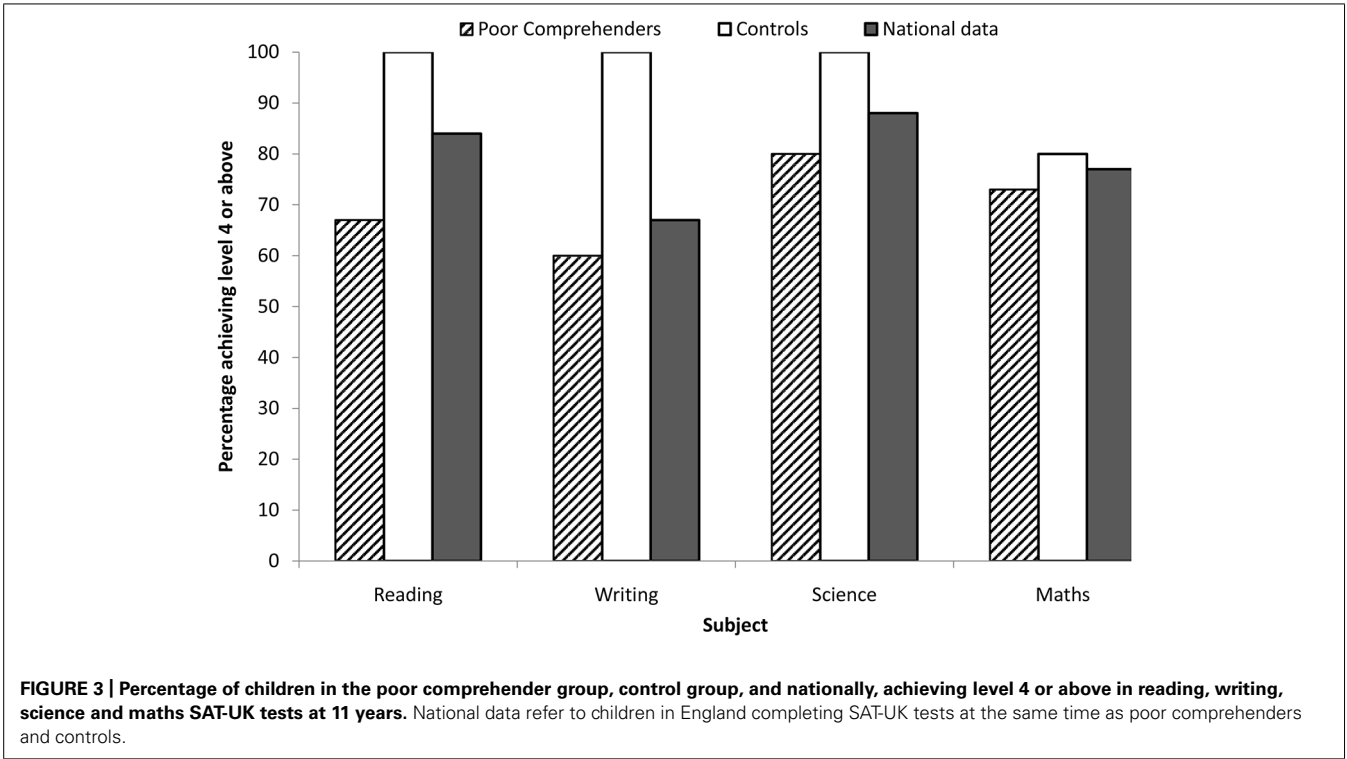
4 vs. level 4 or higher) was significant for writing ($p = 0.02$) but not reading ($p = 0.38$), science ($p = 0.38$) or maths ($p = 0.74$).

**EDUCATIONAL ATTAINMENT AT 16 YEARS (TIME 3)**
As mentioned above, data on educational attainment at 16 years were only available for 21 of the 30 participants. One-way ANOVA and Fisher's exact tests were conducted as appropriate, confirming that there were no systematic differences between those participants who were retained within the sample and those who were not on age, gender, nonverbal reasoning, reading, vocabulary and SAT-UK performance (all $p$s > 0.05). **Table 2** summarizes means and standard deviations for poor comprehender and control groups on GCSE exams (or equivalent), which occur at the end of compulsory schooling in the UK. **Table 2** indicates the total number

of qualifications taken, total points obtained and average points obtained. To mirror the SAT-UK test scores reported above, we also present average points obtained in English, maths, and science subjects (note that a maths points score was not available for one participant in the poor comprehender group). Compared to the controls, there were clear trends for the poor comprehenders to take fewer subjects at GCSE, obtain fewer points overall and perform less well on English. However, these group differences did not reach statistical significance (although effect sizes were small to moderate, see **Table 2**).

In a final set of analyses, we considered two key government targets (for details, see Materials and Methods section above). When the government report on attainment at the end of compulsory education, key indices are whether children obtain five or

**FIGURE 3 | Percentage of children in the poor comprehender group, control group, and nationally, achieving level 4 or above in reading, writing, science and maths SAT-UK tests at 11 years.** National data refer to children in England completing SAT-UK tests at the same time as poor comprehenders and controls.

**Table 2 | Summary of education outcomes at Time 3 (age 16 years).**

| Measure | Poor comprehenders ($n = 11$) | | Controls ($n = 10$) | | Group comparisons | | |
|---|---|---|---|---|---|---|---|
| | M | SD | M | SD | F | p | Cohen's d |
| Number of qualifications | 7.95 | 2.77 | 9.35 | 2.21 | 1.61 | 0.22 | 0.56 |
| Total points | 362.73 | 186.66 | 406.90 | 151.88 | 0.35 | 0.56 | 0.26 |
| Average points | 44.66 | 14.21 | 42.73 | 12.22 | 0.11 | 0.74 | 0.15 |
| Average English points | 37.82 | 8.81 | 41.40 | 16.84 | 0.38 | 0.54 | 0.28 |
| Average Science points | 41.11 | 12.11 | 42.70 | 9.79 | 0.11 | 0.75 | 0.15 |
| Average Maths points[1] | 40.87 | 13.64 | 41.27 | 14.78 | 0.00 | 0.95 | 0.03 |

[1] One participant in the poor comprehender group did not obtain a GCSE maths score therefore the mean for the poor comprehender group is based on 10 participants only.

more GCSEs (or equivalent) at level 2, and whether they make "expected progress" between SAT-UK and GCSE examinations in English and maths. In our sample, 6/11 children in the poor comprehender group (55%) achieved five or more level 2 grades (or equivalent), compared to 7/10 children in the control group (70%). A Fisher's exact test revealed that there was no significant association between comprehension group (poor comprehenders vs. controls) and whether or not participants achieved five or more level 2 grades ($p = 0.66$). We then compared the percentage of pupils in each comprehension group who obtained five or more level 2 GCSE grades (or equivalent) to the national percentage of pupils in England (83%; total $n \approx 561,300$) for the same calendar year (UK Department for Education, 2012b). For the comparison with the poor comprehender group, the Fisher's exact test indicated that there was a significant association between

group (poor comprehender vs. national) and attainment (five level 2 vs. not; $p = 0.03$). For the comparison between the controls and national data, this association was not significant ($p = 0.39$).

By ascertaining whether children made expected progress, it is possible to tap into the relationship between SAT-UK and GCSE performance. In English, 7/11 poor comprehenders (64%) and 7/10 controls (70%) made expected progress; in maths, 7/10 poor comprehenders (70%, one poor comprehender did not take maths GCSE) and 7/10 controls made expected progress. The same seven controls made expected progress across English and maths subjects. For poor comprehenders, there was almost complete overlap across subjects, with the exception of one poor comprehender making expected progress in English and not taking a maths GCSE (it is unclear why as English, maths and

science are compulsory), and another poor comprehender making expected progress in maths but not English. A Fisher's exact test revealed that there was no significant association between comprehension group (poor comprehenders vs. controls) and whether or not participants made expected progress for English ($p = 1.00$) or maths ($p = 1.00$). When these groups were compared to the number of pupils nationally who made expected progress in English (69%; total $n \approx 522{,}782$) and maths (70%; $n \approx 522{,}709$) over the same time frame, there were no significant associations for either poor comprehender (English: $p = 0.75$; maths: $p = 1.00$) or control (English: $p = 1.00$; maths: $p = 1.00$) groups.

## DISCUSSION

Despite a wealth of research investigating cognitive and linguistic skills in poor comprehenders in Italy, the UK and the US (e.g., Catts et al., 2006; Carretti et al., 2009; Nation et al., 2010), and the likely constraint that reading comprehension difficulties will place on educational progress, research on educational attainment was previously restricted to just one study, conducted in the UK with 11-year-old children (Cain and Oakhill, 2006). In the present study, data on national educational attainment tests in the UK were collected in order to explore whether poor comprehenders first recruited at age 9 years show poor educational outcomes at the end of primary school (age 11 years) and at the end of compulsory schooling (age 16 years). Data collected at ages 9 and 11 years also enabled investigation of reading and oral vocabulary development.

At 11 years, approximately a third of poor comprehenders failed to meet government targets on reading and writing tests and there was clear evidence for low achievement in reading compared to the national data set. Poor comprehenders showed lower scores on reading and writing tests compared to controls without a history of reading comprehension difficulties, despite groups being closely matched for age, general cognitive ability and decoding skill. Therefore, our findings point to a link between reading comprehension (and oral vocabulary) difficulties and poor educational attainment that cannot be explained by decoding or general cognitive ability. In the main, our study replicates Cain and Oakhill (2006), who showed differences in educational attainment on these tests between poor comprehenders and a similar control group. However, in contrast to Cain and Oakhill's (2011) study, differences between poor comprehenders and controls were restricted to English tests (i.e., group differences on reading and writing but not maths and science). Given marked heterogeneity in the profiles of children described as poor comprehenders (Nation et al., 2002; Cain and Oakhill, 2006; Floyd et al., 2006), differences between studies are perhaps to be expected.

At 16 years, evidence for low educational attainment in poor comprehenders was less clear. When poor comprehenders were compared to controls, there were no significant differences on any of the indices of achievement, although on almost all measures, poor comprehenders performed less well than controls. It is worth noting, however, that nearly one in two of our poor comprehenders failed to achieve five GCSEs at A* to C, compared to approximately one in six nationally. Taken together with findings from age 11 years, our study indicates that poor comprehenders

are at risk of educational failure at the end of primary school, and may also be at a disadvantage at the end of compulsory education.

Findings on attainment at 16 years should be treated with caution as data were only available for a subsample of poor comprehenders (11/15) and controls (10/15). Given the small sample size, and therefore limited power, it is perhaps not surprising that differences between poor comprehender and control groups were not statistically significant. In addition, we were not able to collect individual data on reading and other aspects of cognitive functioning at age 16 years, thus the reading (and oral vocabulary) status of participants at this point is unknown. Also unknown is whether any children had support during examinations (e.g., extra time, scribe). Nonetheless, to our knowledge, we provide the first study investigating educational attainment in poor comprehenders at the end of compulsory education. Further, our finding that children with a history of reading comprehension difficulties are less likely than pupils nationally to obtain five GCSEs at A* to C warrants further investigation: this is an index that is widely used by UK educational institutions and employers to make recruitment decisions and failing to obtain five GCSEs at A* to C is associated with greater risk of falling into the category of school leavers who are "Not in Employment, Education or Training" (NEET; UK Department for Education, 2010).

Alongside collecting data on educational attainment in poor comprehenders, we also tracked reading and vocabulary longitudinally. Reading and vocabulary measures were administered when poor comprehenders were identified at age 9 years and after a 2-year lag at age 11 years. Raw reading comprehension scores for poor comprehenders and controls increased significantly over time but this increase was more marked for the poor comprehenders (see **Figure 1**). For poor comprehenders, reading comprehension standard scores showed stability; with one or two exceptions, they showed little change over time. Controls' standard scores declined indicating that their improvements were not commensurate with the age-related differences reported for the test's normative sample. This is a surprising finding, and one that warrants further attention. Importantly though, the group difference in reading comprehension (raw and standard scores) maintained over time and the gap between low and high ability groups did not appear to widen (i.e., a Matthew effect), consistent with previous research (Scarborough and Parker, 2003; Cain and Oakhill, 2006, 2011; Elwér et al., 2013).

Mean oral vocabulary scores (raw, standard) for poor comprehenders were significantly lower than mean scores for controls at both Time 1 and Time 2. Over time, scores for the two groups showed parallel growth, with raw scores increasing and mean standard scores not changing significantly between ages 9 and 11 years (see **Figure 1**). Therefore, and in contrast to Cain and Oakhill (2011), we did not find evidence for Matthew effects in the oral vocabulary knowledge of poor comprehenders. Rather, they demonstrated poorer oral vocabulary knowledge than controls at Time 1, and this group difference was maintained (but did not increase) over time (cf. Scarborough and Parker, 2003). Given the discrepancy between our findings and those of Cain and Oakhill, it is worth noting that Cain and Oakhill (2011) identified their poor comprehenders using different criteria. In addition, markedly different measures of oral vocabulary were used across the studies.

Cain and Oakhill (2011) used a receptive measure, with scores determined by the breadth of oral vocabulary knowledge (i.e., how many words a child knows) whereas our expressive measure was more sensitive to the incremental nature of oral vocabulary, with scores capturing depth as well as breadth of knowledge. In order to investigate further whether poor comprehenders are at risk of Matthew effects for vocabulary, future research should aim to administer multiple measures of oral vocabulary, indexing vocabulary knowledge in relation to breadth, depth, and flexibility (e.g., understanding multiple meanings) and how this knowledge can be used.

In conclusion, we have replicated findings that poor comprehenders are at risk for poor educational attainment at the end of primary school (Cain and Oakhill, 2006). At this point, poor comprehenders were more likely to perform poorly, and fail to reach government targets, than controls and the national sample on literacy tests. We also extended this by providing preliminary evidence that some poor comprehenders show low educational outcomes at the end of compulsory education (16 years); compared to the national sample, poor comprehenders were less likely to obtain five or more A* to C GCSE grades (or equivalent). These findings indicate that more research on educational attainment in poor comprehenders is warranted. A key outstanding empirical question is *why* some poor comprehenders perform poorly in national school assessments. The complexity of these assessments means that there are a large number of factors that could constrain performance and given the heterogeneity of poor comprehenders, different factors could explain poor performance for different individuals. Further research is needed that tracks educational attainment in a more systematic and detailed way, and with large enough groups in order to investigate different trajectories. For instance, it would be of value to determine which factors (e.g., reading comprehension level, oral language abilities, ability to learn from reading, etc.) predict the likelihood that poor comprehenders will go on to perform poorly at school. A further complication for interpreting our findings is that SAT-UK and GCSE assessments are not directly comparable. For example, SAT-UK English tests measure reading ability directly whereas GCSE English assessments do not. Thus, there may be different reasons for poor performance at different educational stages. Future research that analyses the content of the tests taken could shed light on this issue, and probe the implications of this work for curriculum development and education in the UK. Finally, given that the extant literature comprises just two UK studies, future studies should aim to investigate links between poor reading comprehension and educational attainment in children outside of the UK. Difficulties with reading comprehension in childhood do not seem to guarantee poor educational outcomes and clearly there are a number of other variables that will influence national assessment scores. Taken together though, our findings do point to a link between reading comprehension difficulties in mid to late childhood and poor educational attainment further down the line.

## AUTHOR CONTRIBUTIONS

For the present study, Jessie Ricketts led in relation to study design, data collection, data analysis and the interpretation of the resulting dataset. Rachael Sperring contributed to data collection, data analysis and interpretation, and Kate Nation to study design, data analysis and interpretation. All authors contributed to manuscript preparation, with Jessie Ricketts taking the lead. All authors have approved the manuscript and take responsibility for all aspects of the work.

## ACKNOWLEDGMENTS

## REFERENCES

Bowyer-Crane, C. A., and Snowling, M. J. (2005). Assessing children's inference generation: what do tests of reading comprehension measure? *Br. J. Educ. Psychol.* 75, 189–201. doi: 10.1348/000709904X22674

Cain, K., and Oakhill, J. (2006). Profiles of children with specific reading comprehension difficulties. *Br. J. Educ. Psychol.* 76, 683–696. doi: 10.1348/0007099056X67610

Cain, K., and Oakhill, J. (2007). "Reading comprehension difficulties: Correlates, causes and consequences," in *Children's Comprehension Problems in Oral and Written Language: A Cognitive Perspective*, eds K. Cain and J. Oakhill (New York: Guildford Press), 41–75.

Cain, K., and Oakhill, J. (2011). Matthew effects in young readers: Reading comprehension and reading experience aid vocabulary development. *J. Learn. Disabil.* 44, 431–443. doi: 10.1177/0022219411410042

Cain, K., Oakhill, J., and Elbro, C. (2003). The ability to learn new word meanings from context by school-age children with and without language comprehension difficulties. *J. Child Lang.* 30, 681–694. doi: 10.1017/S0305000903005713

Cain, K., Oakhill, J., and Lemmon, K. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *J. Educ. Psychol.* 96, 671–681. doi: 10.1037/0022-0663.96.4.671

Carretti, B., Borella, E., Cornoldi, C., and De Beni, R. (2009). Role of working memory in explaining the performance of individuals with specific reading comprehension difficulties: A meta-analysis. *Learn. Individ. Differ.* 19, 246–251. doi: 10.1016/j.lindif.2008.10.002

Catts, H. W., Adlof, S. M., and Weismer, S. E. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *J. Speech Lang. Hear. Res.* 49, 278–293. doi: 10.1044/1092-4388(2006/023)

Clarke, P. J., Snowling, M. J., Truelove, E., and Hulme, C. (2010). Ameliorating children's reading comprehension difficulties: a randomised controlled trial. *Psychol. Sci.* 21, 1106–1116. doi: 10.1177/0956797610375449

Dockrell, J. E., Lindsay, G., and Palikara, O. (2011). Explaining the academic achievement at school leaving for pupils with a history of language impairment: previous academic achievement and literacy skills. *Child Lang. Teach. Therapy* 27, 223–237. doi: 10.1177/0265659011398671

Elwér, Å., Keenan, J. M., Olson, R. K., Byrne, B., and Samuelsson, S. (2013). Longitudinal stability and predictors of poor oral comprehenders and poor decoders. *J. Exp. Child Psychol.* 115, 497–516. doi: 10.1016/j.jecp.2012.12.001

Floyd, R., Bergeron, R., and Alfonso, V. (2006). Cattell–Horn–Carroll cognitive ability profiles of poor comprehenders. *Read. Writ.* 19, 427–456. doi: 10.1007/s11145-006-9002-5

Gough, P. B., and Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial Spec. Educ.* 7, 6–10. doi: 10.1177/074193258600700104

Harlaar, N., Cutting, L., Deater-Deckard, K., DeThorne, L. S., Justice, L. M., Schatschneider, C., et al. (2010). Predicting individual differences in reading comprehension: a twin study. *Ann. Dyslexia* 60, 265–288. doi: 10.1007/s11881-010-0044-7

Hulme, C., and Snowling, M. J. (2011). Children's reading comprehension difficulties: nature, causes, and treatments. *Curr. Dir. Psychol. Sci.* 20, 139–142. doi: 10.1177/0963721411408673

Kirby, J. R., and Savage, R. S. (2008). Can the simple view deal with the complexities of reading? *Literacy* 42, 75–82. doi: 10.1111/j.1741-4369.2008.00487.x

Muter, V., Hulme, C., Snowling, M. J., and Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: evidence from a longitudinal study. *Dev. Psychol.* 40, 665–681. doi: 10.1037/0012-1649.40.5.665

Nagy, W., Herman, P. A., and Anderson, R. C. (1985). Learning words from context. *Read. Res. Q.* 20, 233–253. doi: 10.2307/747758

Nation, K. (2005). "Children's reading comprehension difficulties," in *The Science of Reading: A Handbook*, eds M. J. Snowling, C. Hulme, and M. Seidenberg (Oxford: Blackwell), 248–265.

Nation, K., and Angell, P. (2006). Learning to read and learning to comprehend. *Lond. Rev. Educ.* 4, 77–88. doi: 10.1080/13603110600574538

Nation, K., Clarke, P., and Snowling, M. J. (2002). General cognitive ability in children with reading comprehension difficulties. *Br. J. Educ. Psychol.* 72, 549–560. doi: 10.1348/00070990260377604

Nation, K., Cocksey, J., Taylor, J. S., and Bishop, D. V. M. (2010). A longitudinal investigation of early reading and language skills in children with poor reading comprehension. *J. Child Psychol. Psychiatry* 51, 1031–1039. doi: 10.1111/j.1469-7610.2010.02254.x

Nation, K., Snowling, M. J., and Clarke, P. (2007). Dissecting the relationship between language skills and learning to read: semantic and phonological contributions to new vocabulary learning in children with poor reading comprehension. *Adv. Speech Lang. Pathol.* 9, 131–139. doi: 10.1080/14417040601145166

Neale, M. (1997). Neale analysis of reading ability – second revised british edition. Windsor: NFER.

Oakhill, J., and Cain, K. (2012). The precursors of reading ability in young readers: evidence from a four-year longitudinal study. *Sci. Stud. Read.* 16, 91–121. doi: 10.1080/10888438.2010.529219

Oakhill, J., Cain, K., and Bryant, P. (2003). The dissociation of word reading and text comprehension: evidence from component skills. *Lang. Cogn. Proc.* 18, 443–468. doi: 10.1080/01690960344000008

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

Ricketts, J., Bishop, D. V. M., and Nation, K. (2008). Investigating orthographic and semantic aspects of word learning in poor comprehenders. *J. Res. Read.* 31, 117–135. doi: 10.1111/j.1467-9817.2007.00365.x

Ricketts, J., Bishop, D. V. M., Pimperton, H., and Nation, K. (2011). The role of self-teaching in learning orthographic and semantic aspects of new words. *Sci. Stud. Read.* 15, 47–70. doi: 10.1080/10888438.2011.536129

Ricketts, J., Jones, C. R. G., Happé, F., and Charman, T. (2013). Reading comprehension in autism spectrum disorders: the role of oral language and social functioning. *J. Autism Dev. Disord.* 43, 807–816. doi: 10.1007/s10803-012-1619-4

Ricketts, J., Nation, K., and Bishop, D. V. M. (2007). Vocabulary is important for some, but not all reading skills. *Sci. Stud. Read.* 11, 235–257. doi: 10.1080/10888430701344306

Scarborough, H. S., and Parker, J. D. (2003). Matthew effects in children with learning disabilities: development of reading, IQ, and psychosocial problems from grade 2 to grade 8. *Ann. Dyslexia* 53, 47–71. doi: 10.1007/s11881-003-0004-6

Snowling, M. J., Adams, J. W., Bishop, D. V. M., and Stothard, S. (2001). Educational attainments of school leavers with a preschool history of speech-language impairments. *Int. J. Lang. Commun. Disord.* 36, 173–183. doi: 10.1080/13682820120976

Stanovich, K. E. (1986). Matthew effects in reading: some consequences of individual differences in the acquisition literacy. *Read. Res. Q.* 21, 360–406. doi: 10.1598/RRQ.21.4.1

Torgesen, J., Wagner, R., and Rashotte, C. (1999). *Test of Word Reading Efficiency*. Austin, TX: Pro-Ed.

Tunmer, W. E., and Chapman, J. W. (2012). The simple view of reading redux: vocabulary knowledge and the independent components hypothesis. *J. Learn. Disabil.* 45, 453–466. doi: 10.1177/0022219411432685

UK Department for Education. (2010). *Youth Cohort Study and Longitudinal Study of Young People in England*. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/218911/b01-2010v2.pdf [accessed September 17, 2013].

UK Department for Education. (2012a). *National Curriculum Assessments at Key Stage 2 in England: 2012*. Available at: https://www.gov.uk/government/publications/national-curriculum-assessments-at-key-stage-2-in-england-2012 [accessed September 17, 2013].

UK Department for Education. (2012b). *Revised GCSE and Equivalent Results in England: Academic Year 2011 to 2012*. Available at: https://www.gov.uk/government/publications/revised-gcse-and-equivalent-results-in-england-academic-year-2011-to-2012 [accessed September 17, 2013].

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. London: The Psychological Corporation.

# ADVANTAGES OF PUBLISHING IN FRONTIERS

**FAST PUBLICATION**

Average 90 days
from submission
to publication

**COLLABORATIVE
PEER-REVIEW**

Designed to be rigorous –
yet also collaborative, fair and
constructive

**RESEARCH NETWORK**

Our network
increases readership
for your article

**OPEN ACCESS**

Articles are free to read,
for greatest visibility

**TRANSPARENT**

Editors and reviewers
acknowledged by name
on published articles

**GLOBAL SPREAD**

Six million monthly
page views worldwide

**COPYRIGHT TO AUTHORS**

No limit to
article distribution
and re-use

**IMPACT METRICS**

Advanced metrics
track your
article's impact

**SUPPORT**

By our Swiss-based
editorial team