# Understanding the role of oral vocabulary in reading comprehension difficulties


**Danielle Colenbrander B. Lib Stud MSLP**

ARC Centre of Excellence in Cognition and its Disorders

Department of Cognitive Science

Macquarie University, Sydney, Australia


A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

December 2014

Table of Contents

## Summary

This thesis explores specific reading comprehension difficulties with a focus on the role of oral vocabulary. The first study illustrates the complexities associated with diagnosing reading comprehension difficulties by comparing two widely used assessments of reading comprehension. The second study explores the vocabulary, irregular word reading and oral language skills of a sample of children with poor reading comprehension. The third study presents the results of a randomized controlled trial looking at the effects of an oral vocabulary training program for poor comprehenders. The fourth study presents data from the same children as a repeated measures intervention study with multiple baselines, allowing an exploration of the results of the training program at an individual level.

This thesis extends the knowledge of the field by demonstrating that there is variability in the vocabulary skills of poor comprehenders at the individual level, and that poor comprehenders who do not have vocabulary difficulties nonetheless perform poorly on tasks of oral language above the word level. In addition, oral vocabulary training is demonstrated to be successful in improving reading comprehension skills for poor comprehenders with vocabulary deficits, implying a direct causal role for poor oral vocabulary skills in reading comprehension difficulties.

## Declaration

The work in this thesis is my own original work. It has not been submitted for a higher degree in any other university or institution. All of the work reported in this thesis was undertaken during the time I was enrolled as a PhD student at Macquarie University, under the supervision of Dr Saskia Kohnen, Professor Lyndsey Nickels and Dr Karen Smith-Lock. Ethics approval for the studies reported in this thesis was obtained from Macquarie University's Human Research Ethics Committee, Reference No. 201200038 (evidence of final approval is available in Appendix 7).

Signed:

Danielle Colenbrander

22nd December 2014

**Acknowledgements**

I would like to thank my wonderful supervisors, Dr. Saskia Kohnen, Professor Lyndsey Nickels and Dr. Karen Smith-Lock, for guiding me and supporting me through this process. In particular I would like to thank Saskia and Lyndsey, who supervised both my Masters project and my PhD. Their integrity, wisdom and passion for their work inspired me to do a PhD in the first place and I could not ask for better mentors.

Secondly, I would like to thank my family for their love and support, and for believing wholeheartedly that I could do this. I would especially like to thank my parents for showing by example what can be achieved with hard work and dedication. Because of the choices they made, I have had the opportunity to do what I loved, and I am so grateful.

Thirdly, I would like to thank my fellow PhD students and colleagues in the CCD for their friendship, help and advice over the years. When I had a question, there was always someone to answer it. When I had a problem, there was always someone who knew what to do. When I needed a good laugh, there was always someone to share it with. I am very lucky to have worked with such a lovely group of people, many of whom have become dear friends.

Last but certainly not least, I would like to thank Rob. This process brought us together and we have shared it from beginning to end. He has understood me, supported me, and challenged me in the best possible way. Without him the world would be a much less interesting place.

# Chapter One: Introduction

## Introduction

Learning to read is one of the most fundamental and important goals of education. The ability to read is crucial for academic success and everyday functioning – individuals with poor reading abilities are likely to fall behind their peers at school, which may limit their choices when it comes to future employment (Boetsch, Green, & Pennington, 1996; McLaughlin, Speirs, & Shenassa, 2012). Poor reading abilities may also lead to reduced self-esteem and perceptions of self-efficacy, which can in turn affect a child's motivation to learn and succeed (Boetsch, et al., 1996; Ingesson, 2007).

A significant minority of children have difficulties learning to read (that is, learning to sound out or recognize the written forms of words; Shaywitz, 1998). However, it is possible for a child to fail to comprehend written text even when his or her ability to read is age-appropriate: research has identified a group of children who can read aloud without difficulty, but who perform poorly on tasks of reading comprehension (e.g., Nation & Snowling, 1997; Stothard & Hulme, 1995; Yuill & Oakhill, 1991). These children are referred to as "less-skilled comprehenders" (e.g., Cain & Oakhill, 1999), "unexpected poor comprehenders" (e.g., Tong, Deacon, Kirby, Cain, & Parrila, 2011) or "poor comprehenders" (e.g., Nation & Snowling, 1997). The term "poor comprehenders" is used throughout this thesis for the purposes of brevity. Previous research has shown that such children make up 3-10% of the school-aged population, depending on the criteria used to define the sample (Hulme & Snowling, 2011; Nation & Snowling, 1997; Stothard & Hulme, 1992; Yuill & Oakhill, 1991).

While the existence of the poor comprehender profile is now widely acknowledged, there is little agreement as to how exactly it is defined (Clarke, Henderson, & Truelove, 2010; Elwer, Keenan, Olson, Byrne, & Samuelsson, 2013). Researchers agree that in order for a child to be designated a poor comprehender, the child must have age-appropriate word and/or text reading ability, and that there must be a discrepancy between the child's reading ability, and his or her ability to understand the text that he or she reads. However, the size of the discrepancy between reading ability and reading comprehension differs across studies. The authors of a recent paper (Clarke, Henderson, et al., 2010) determined that the average discrepancy between reading ability and reading comprehension in existing studies varied between 12 to 32 age-equivalent score months and 17 to 24 standard score points. In addition, some studies include age appropriate vocabulary skills (measured by tests of written and spoken word-picture matching) as part of their definition of the poor comprehender profile (e.g., Cain, 2006; Cain & Oakhill, 1999; Cain, Oakhill, & Lemmon, 2004), while in other studies, vocabulary skills are left to vary (e.g., Cain & Oakhill, 2006b; Hua & Keenan, 2014; Nation & Snowling, 1998, 1999; Ricketts, Bishop, & Nation, 2008; Ricketts, Sperring, & Nation, 2014).

A further source of variation across studies is the fact that studies differ in the way reading abilities are measured. In some cases, "reading ability" is defined as text reading ability (e.g., Cain & Oakhill, 1999; Stothard & Hulme, 1995), in others, it is defined as nonword reading accuracy or fluency (e.g., Clarke, Snowling, Truelove, & Hulme, 2010; Nation & Snowling, 2004; Nation, Snowling, & Clarke, 2007), real word reading accuracy or fluency (e.g., Hua & Keenan, 2014) or some combination of the above (e.g., Catts, Adlof, & Weismer, 2006; Tong, Deacon, & Cain, 2014; Tong, et al., 2011).

Relatedly, and crucially, reading comprehension itself is not always measured in the same way across studies. The majority of studies originating in the UK have used the Neale Analysis of Reading (NARA; Neale, 1999), a test of passage reading comprehension requiring verbal responses to open-ended questions (see for example Cain, 2006; Cain & Oakhill, 2006a; Cain & Oakhill, in press; Clarke, Snowling, et al., 2010; Nation, Cocksey, Taylor, & Bishop, 2010; Stothard & Hulme, 1995). Other studies have used a variety of different assessments such as the Qualitative Reading Inventory (Leslie & Caldwell, 2001), which also requires open-ended verbal responses; the Woodcock Reading Mastery Test Passage Comprehension subtest, which requires children to fill a blank with a suitable word

(Woodcock, 1987); or the Gray Oral Reading Test (Wiederholt & Bryant, 1992), which requires multiple choice responses. The choice of reading comprehension assessment is crucial because different tests will have different task demands and may tap different underlying skills (Cain & Oakhill, 2006a; Cutting & Scarborough, 2006; Keenan, Betjemann, & Olson, 2008; Keenan & Meenan, 2014; Kendeou, Papadopoulos, & Spanoudis, 2012; Nation & Snowling, 1997), therefore a child may be diagnosed with poor reading comprehension on one test, but not on another. For these reasons, it is not always straightforward to compare results across studies.

Nonetheless, there is evidence that poor reading comprehension skills can limit a child's academic performance (Cain & Oakhill, 2006b; Ricketts, et al., 2014). The difficulties experienced by poor comprehenders may be very subtle and may not be recognized in the classroom, particularly because poor comprehenders can read aloud without difficulty (Hulme & Snowling, 2011; Ricketts, et al., 2014). In addition, specific reading comprehension difficulties (that is, the reading comprehension difficulties demonstrated by poor comprehenders) may only become apparent when a child reaches the later primary school years, because it is at this point that the focus of the curriculum shifts from teaching children to read, to expecting that they will be able to learn from what they read (Elwer, et al., 2013; Hogan, Adlof, & Alonzo, 2014). For these reasons, it is important to develop an understanding of the nature of the difficulties experienced by children with specific reading comprehension disorders, so that we can determine the underlying causes and develop effective interventions for such children.

One influential theory which makes a claim about the causes and underlying difficulties associated with specific reading comprehension difficulties is the Simple View of Reading (Gough & Tunmer, 1986; Hoover & Gough, 1990). The Simple View states that successful reading comprehension is composed of two main areas of skill: decoding or reading abilities – originally defined as nonword reading abilities, although there is some debate as to the best way to assess reading skills within the Simple View (Garcia & Cain, 2013; Gough & Tunmer, 1986; Kirby & Savage, 2008); and oral language skills (sometimes referred to as "listening comprehension").

Note that for the purposes of this thesis, the term "reading ability" will be used to refer to both word and nonword reading abilities. The term "oral language skills" will be used to refer to the ability to understand and interpret spoken information, whether at the lexical, sentence or discourse level. It is worth noting that oral language is not a simple construct - it is comprised of multiple skills and processes, from knowledge of word meanings to sentence comprehension abilities and inferential abilities (Kirby & Savage, 2008).

It is also worth noting that there are differences between written language and oral (spoken) language. Compared to oral language, written language is often more "formal" and sophisticated. For example, written texts generally contain a greater number of infrequent words and complex sentences than conversational language (Oakhill & Cain, 2007, Cunningham, 2005). However, the development of written language abilities nonetheless depends on oral language skills (Oakhill & Cain, 2007). Consequently, in this thesis, the distinction between oral and written language does not refer to the difference between conversational and formal language registers. Rather, as in Hoover and Gough (1990), the distinction between oral language and written language refers simply to the method by which meaning is accessed – in other words, whether it is accessed via oral input, or via written input.

Given this, the Simple View predicts four types of readers – good readers, who have both good reading skills and good oral language skills; "garden variety" poor readers who have both poor reading and poor oral language skills; dyslexic readers, who have good oral language skills but poor reading abilities; and poor comprehenders, who have good reading skills but poor oral language skills (Gough & Tunmer, 1986; Hoover & Gough, 1990).

Therefore, the Simple View attributes poor comprehenders' reading comprehension difficulties to oral language difficulties.

Consistent with this prediction, poor comprehenders have been shown to have weaknesses in a wide variety of oral language skills, including oral vocabulary and semantics (Catts et al., 2006; Nation & Snowling, 1998, 1999), morphology (Tong et al., 2014; Tong et al., 2011) and syntax (Nation & Snowling, 2000; Tong et al., 2014). Any of these skills could potentially play a causal role in reading comprehension difficulties. Of these, vocabulary has received a large amount of attention in the reading comprehension literature (Cain & Oakhill, in press; Nation et al., 2007; Ouelette, 2006; Ouelette & Beers, 2010; Protopapas, Mouzaki, Sideridis, Kotsolakou, & Simos, 2013).

A theoretical perspective on the role of vocabulary skills within reading comprehension is provided by the Lexical Quality Hypothesis (Perfetti, 2007; Perfetti & Hart, 2001, 2002). According to the Lexical Quality Hypothesis, the ability to comprehend written text rests on the ability to read and understand individual words, which in turn liberates resources for higher-level processing efforts (Perfetti & Hart, 2001, 2002). Word knowledge is made up of three main constituents, phonological knowledge (the way a word sounds), orthographic knowledge (the written form of a word) and semantic knowledge (a word's meaning). A word has a high lexical quality when a person can efficiently access all three constituents in a coordinated manner (Perfetti, 2007; Perfetti & Hart, 2002). Lexical quality is a property of individual words, but some people may have fewer words of high lexical quality in their vocabulary than others (Perfetti & Hart, 2002). This may be the case for poor comprehenders.

The Simple View of Reading and the Lexical Quality Hypothesis are not mutually exclusive. If both the Simple View and the Lexical Quality Hypothesis are correct, we would expect poor comprehenders to have difficulties with the oral language aspects of vocabulary knowledge. In particular, they should have difficulties with semantics, but they should have no difficulties with the orthographic aspect of vocabulary knowledge. In fact, there is evidence that poor comprehenders have difficulty with semantics. At the group level, poor comprehenders have been shown to have weaknesses on tasks of semantics such as word-picture matching tasks, definition tasks, semantic priming and semantic fluency tasks (Catts et al., 2006; Nation, Clarke, Marshall, & Durand, 2004; Nation & Snowling, 1998, 1999), whilst having relative strengths in the ability to learn the orthographic forms of words (Ricketts et al., 2008). Poor comprehenders have also been shown to have strengths on tasks of phonological skills such as nonword reading, nonword repetition, phoneme deletion and creating spoonerisms (Catts et al., 2006; Nation et al., 2004; Stothard & Hulme, 1995).

Despite the fact that poor comprehenders (by definition) have intact nonword and even real word reading skills, some studies have shown that poor comprehenders as a group tend to have difficulties reading irregular words (words which do not follow the letter-sound rules of the English language and therefore have to be read by sight; Nation & Snowling, 1998; Ricketts et al., 2008; Ricketts, Nation, & Bishop, 2007). These poor irregular word reading skills are thought to be a consequence of poor comprehenders' weak semantic abilities (Nation & Snowling, 1998; Ricketts et al., 2007).

This hypothesis is consistent with a connectionist model of single word reading, the Triangle model (Plaut, McClelland, Seidenberg, & Patterson, 1996). Under the Triangle model, reading aloud is achieved by a distributed network of connections representing the three different constituents of word knowledge (semantics, orthography and phonology). When a person sees a word, this activates codes in a phonological pathway (made up of connections between orthography and phonology) and a semantic pathway (made up of connections between all three constituents of word knowledge). In this model, reading aloud is always accomplished via the activation of both routes, although the extent to which word reading relies on the phonological or semantic pathway varies. The semantic pathway plays a crucial role in reading irregular words because when mappings between phonology and orthography are inconsistent, the phonological pathway may not be able to converge on a correct solution. Consistent with this model, several studies exist which demonstrate a role for semantics in irregular word reading (Betjemann & Keenan, 2008; Keenan & Betjemann, 2007; Nation & Snowling, 1998; Ricketts et al., 2007).

However, other work has shown that semantic knowledge is not necessarily required for successful irregular word reading (Blazely, Coltheart, & Casey, 2005; Nation & Cocksey, 2009). An optional, rather than necessary, role for semantics in irregular word reading is specifically predicted by the Dual Route model of reading (Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001), which proposes that reading aloud is achieved by two separate pathways, a sub-lexical route which converts letters into sounds by means of grapheme-phoneme correspondence rules, and a lexical route which allows whole words to be recognized by sight. In this model, an irregular word cannot be read correctly via the sub-lexical route and must be read by stored forms of the sight words accessing the stored pronunciations of these words. This store, the phonological lexicon, can be accessed via semantics, but it can also be accessed directly from a word's visual form (its orthography). This model predicts that there could be poor comprehenders who have weak oral vocabulary skills but intact irregular word reading skills.

The role of semantics in irregular word reading has consequences for the Simple View of Reading. If semantics contributes not only to oral language aspects of the Simple View but also to irregular word (and indeed real word) reading, this challenges the Simple View's proposal that reading and oral language skills can be independent (Ouelette & Beers, 2010; Protopapas et al., 2013; Ricketts et al., 2007). However, the existence of children with weak oral vocabulary skills and intact irregular word reading would suggest that reading skills and oral language skills can be largely independent even when a child has semantic difficulties. This possibility has not yet been explored in the reading comprehension literature.

In addition to questions about the contribution of semantics to irregular word reading, many questions about the nature and role of oral vocabulary skills in reading comprehension difficulties exist. Firstly, some of the strongest evidence for the crucial role of vocabulary skills comes from training studies in which vocabulary skills were trained, resulting in improved reading comprehension (Beck, Perfetti, & McKeown, 1982; McKeown, Beck, Omanson, & Perfetti, 1983; McKeown, Beck, Omanson, & Pople, 1985). However, these studies involved a large amount of exposure to written discourse-level texts, and the effects of the training on semantic, phonological and orthographic aspects of word knowledge were not systematically assessed. Therefore, we do not know whether it was improvements in the semantic component of vocabulary knowledge which led to comprehension improvements, or whether improved phonological or orthographic knowledge played a role.

Finally, the vast majority of studies on comprehension difficulties have been carried out at the group level and have not explored individual variation. Considering that some poor comprehenders have been found to have intact vocabulary skills (Cain, 2006; Cain & Oakhill, 1999, 2006b; Cain et al., 2004), it seems crucial to consider the fact that vocabulary skills may be playing a causal role in vocabulary difficulties for some, but not all, poor comprehenders. While a small number of studies of individual difficulties in poor comprehenders do exist (Cain & Oakhill, 2006b; Cornoldi, De Beni, & Pazzaglia, 1996; Nation et al., 2004), none has specifically addressed the relative roles of all three components of vocabulary knowledge (semantics, orthography and phonology).

In summary, the purpose of this thesis is to tease apart the nature of the relationship between oral vocabulary and reading comprehension by examining the manner in which comprehension and vocabulary are assessed, by training oral vocabulary difficulties in a group of poor comprehenders, and by addressing the role of individual variation within the poor comprehender profile.

Chapter 2 addresses the difficulty of defining and measuring reading comprehension by looking at the similarities and differences between one of the most widely used assessments of reading comprehension, the NARA (Neale, 1999), and a recently developed assessment of reading comprehension, the York Assessment of Reading for Comprehension (YARC; Snowling et al., 2012).

In Chapter 3, the nature of oral vocabulary difficulties in reading comprehension is examined through the assessment of all three aspects of lexical knowledge (phonological, orthographic and semantic), allowing the exploration of individual variation in knowledge of these aspects. The prevalence of poor irregular word reading skills within a sample of poor comprehenders is also assessed and the findings are related to the role of semantics within irregular word reading and reading comprehension. The assessments that were designed specifically to test different aspects of oral language and reading skill for use in this chapter and Chapters 4 and 5, are described in detail in the Appendices to this thesis.

In Chapter 4, the causal role of oral vocabulary skills in reading comprehension is explored. A randomized controlled trial training vocabulary skills in poor comprehenders was conducted, using oral-language-based methods.

In Chapter 5, data from the training and wait-list control groups of the aforementioned Randomized Controlled Trial was combined. Data was analyzed at the individual level in order to explore the causal role of oral vocabulary difficulties and the efficacy of the training program at the individual level.

Finally, in Chapter 6, the findings of all the above studies are related to the predictions of the Simple View of Reading, the Lexical Quality Hypothesis, and the Triangle and Dual Route models of reading. Practical and theoretical implications of the findings are discussed and suggestions for future research are made.

## References


Beck, I., Perfetti, C., & McKeown, M. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology, 74*, 506-521. doi:10.1037/0022-0663.74.4.506

Betjemann, R. S., & Keenan, J. M. (2008). Phonological and semantic priming in children with reading disability. *Child Development, 79*, 1086-1102. doi:10.1111/j.1467-8624.2008.01177.x

Blazely, A. M., Coltheart, M., & Casey, B. J. (2005). Semantic impairment with and without surface dyslexia: Implications for models of reading. *Cognitive Neuropsychology, 22*, 695-717. doi:10.1080/02643290442000257

Boetsch, E. A., Green, P. A., & Pennington, B. F. (1996). Psychosocial correlates of dyslexia across the life span. *Development and Psychopathology, 8*, 539-562. doi:10.1017/S0954579400007264

Cain, K. (2006). Individual differences in children's memory and reading comprehension: an investigation of semantic and inhibitory deficits. *Memory*, *14*, 553-569. doi:10.1080/09658210600624481

Cain, K., & Oakhill, J. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing*, *11*, 489-503. doi:10.1023/A:1008084120205

Cain, K., & Oakhill, J. (2006a). Assessment matters: issues in the measurement of reading comprehension. *British Journal of Educational Psychology, 76*, 697-708. doi:10.1348/000709905X69807

Cain, K., & Oakhill, J. (2006b). Profiles of children with specific reading comprehension difficulties. *British Journal of Educational Psychology*, *76*, 683-696. doi:10.1348/000709905X67610

Cain, K., & Oakhill, J. (in press). Reading comprehension and vocabulary: Is vocabulary more important for some aspects of comprehension? *L'Annee Psychologique.*

Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology*, *96*, 671-681. doi:10.1037/0022-0663.96.4.671

Catts, H. W., Adlof, S. M., & Weismer, S. E. (2006). Language deficits in poor comprehenders: A case for the Simple View of Reading. *Journal of Language, Speech and Hearing Research, 49*, 278-293. doi:10.1044/1092-4388(2006/023)

Clarke, P. J., Henderson, L. M., & Truelove, E. (2010). The poor comprehender profile: Understanding and supporting individuals who have difficulties extracting meaning from text. In J. Holmes (Ed.), *Advances in Child Development and Behaviour* (Vol. 39, pp. 79-129). New York, Boston: Academic Press

Clarke, P. J., Snowling, M. J., Truelove, E., & Hulme, C. (2010). Ameliorating children's reading-comprehension difficulties: a randomized controlled trial. *Psychological Science, 21*, 1106-1116. doi:10.1177/0956797610375449

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed processing approaches. *Psychological Review*, *100*, 589-608. doi:10.1037/0033-295X.100.4.589

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

Cornoldi, C., De Beni, R., & Pazzaglia, F. (1996). Profiles of reading comprehension difficulties: An analysis of single cases. In C. Cornoldi & J. Oakhill (Eds.), *Reading Comprehension Difficulties: Processes and Intervention* (pp. 113-136). Mahwah, NJ: Lawrence Erlbaum Associates.

Cunningham, A. E. (2005). Vocabulary growth through independent reading and reading aloud to children. In E. H. Hiebert & M. L. Kamil (Eds.), *Teaching and learning vocbaulary: Bringing research into practice* (pp. 45-69). Mahwah, NJ: Lawrence Erlbaum Associates.

Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading, 10*, 277-299. doi: 10.1207/s1532799xssr1003_5

Elwer, S., Keenan, J. M., Olson, R. K., Byrne, B., & Samuelsson, S. (2013). Longitudinal stability and predictors of poor oral comprehenders and poor decoders. *Journal of Experimental Child Psychology*, *115*, 497-516. doi:10.1016/j.jecp.2012.12.001

Garcia, J. R., & Cain, K. (2013). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of Educational Research, 84*, 74-111. doi:10.3102/0034654313499616

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*, 6-10. doi:10.1177/074193258600700104

Hogan, T. P., Adlof, S. M., & Alonzo, C. N. (2014). On the importance of listening comprehension. *International Journal of Speech Language Pathology*, *16*, 199-207. doi:10.3109/17549507.2014.904441

Hoover, W. A., & Gough, P. B. (1990). The Simple View of Reading. *Reading and Writing, 2*, 127-160. doi:10.1007/BF00401799

Hua, A. N., & Keenan, J. M. (2014). The role of text memory in inferencing and in comprehension deficits. *Scientific Studies of Reading*, *18*, 1-17. doi:10.1080/10888438.2014.926906

Hulme, C., & Snowling, M. J. (2011). Children's Reading Comprehension Difficulties: Nature, Causes, and Treatments. *Current Directions in Psychological Science, 20,* 139-142. doi:10.1177/0963721411408673

Ingesson, S. G. (2007). Growing up with dyslexia: Interviews with teenagers and young adults. *School Psychology International, 28*(5), 574-591. doi:10.1177/0143034307085659

Keenan, J. M., & Betjemann, R. S. (2007). Comprehension of single words: The role of semantics in word identification and reading disability. In E. L. Grigorenko & A. J. Naples (Eds.), *Single-word reading: Behavioural and biological perspectives* (pp. 191-210). New York, NY: Taylor & Francis.

Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading, 12,* 281-300. doi:10.1080/10888430802132279

Keenan, J. M., & Meenan, C. E. (2014). Test differences in diagnosing reading comprehension deficits. *Journal of Learning Disabilities, 47*, 125-135. doi:10.1177/0022219412439326

Kendeou, P., Papadopoulos, T. C., & Spanoudis, G. (2012). Processing demands of reading comprehension tests in young readers. *Learning and Instruction, 22*, 354-367. doi:10.1016/j.learninstruc.2012.02.001

Kirby, J. R., & Savage, R. S. (2008). Can the Simple View of Reading deal with the complexities of reading? *Literacy, 42*, 75-82. doi:10.1111/j.1741-4369.2008.00487.x

Leslie, L., & Caldwell, J. (2001). *Qualitative Reading Inventory-3*. New York, NY: Addison Wesley Longman.

McKeown, M., Beck, I., Omanson, R., & Perfetti, C. (1983). The effects of long-term vocabulary instruction on reading comprehension: A replication. *Journal of Literacy Research, 15,* 3-18. doi:10.1080/10862968309547474

McKeown, M., Beck, I., Omanson, R., & Pople, M. (1985). Some effects of the nature and frequency of vocabulary instruction on the knowledge and use of words. *Reading Research Quarterly, 20*, 522-535. doi:10.2307/747940

McLaughlin, M. J., Speirs, K. E., & Shenassa, E. D. (2012). Reading disability and adult attained education and income: Evidence from a 30-year longitudinal study of a population-based sample. *Journal of Learning Disabilities, 47*, 374-386. doi:10.1177/0022219412458323

Nation, K., Clarke, P., Marshall, C. M., & Durand, M. (2004). Hidden language impairments in children: Parallels between poor reading comprehension and specific language impairment? *Journal of Speech, Language and Hearing Research*, *47*, 199-211. doi:10.1044/1092-4388(2004/017)

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

Nation, K., Cocksey, J., Taylor, J. S., & Bishop, D. V. (2010). A longitudinal investigation of early reading and language skills in children with poor reading comprehension. *Journal of Child Psychology and Psychiatry*, *51*, 1031-1039. doi:10.1111/j.1469-7610.2010.02254.x

Nation, K., & Snowling, M. J. (1997). Assessing reading difficulties: the validity and utility of current measures of reading skill. *British Journal of Educational Psychology, 67*, 359-370. doi:10.1111/j.2044-8279.1997.tb01250.x

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

Nation, K., & Snowling, M. J. (1999). Developmental differences in sensitivity to semantic relations among good and poor comprehenders: Evidence from semantic priming. *Cognition, 70*, B1-B13. doi:10.1016/S0010-0277(99)00004-9

Nation, K., & Snowling, M. J. (2000). Factors influencing syntactic awareness in normal readers and poor comprehenders. *Applied Psycholinguistics, 21*, 229-241. doi:10.1017/S0142716400002046

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

Nation, K., Snowling, M. J., & 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. *Advances in Speech-Language Pathology, 9*, 131-139. doi:10.1080/14417040601145166

Neale, M. D. (1999). *The Neale Analysis of Reading Ability Third Edition*. Melbourne, Australia: ACER.

Oakhill, J., & Cain, C. (2007). Introduction to comprehension impairment. In K. Cain & J. Oakhill (Eds.), *Children's comprehension problems in oral and written language: A cognitive perspective* (pp. 41-75). New York, NY: Guilford Press.

Ouelette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology, 98*, 554-566. doi:10.1037/0022-0663.98.3.554

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

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*, 357-383. doi:10.1080/10888430701530730

Perfetti, C., & Hart, L. (2001). The lexical basis of comprehension skill. In D. S. Gorfien (Ed.), *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity* (pp. 67-86). Washington, DC: American Psychological Association.

Perfetti, C., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro & P. Reitsma. (Eds.), *Precursors of funtional literacy* (pp. 189-213). Amsterdam/Philadelphia: John Benjamins.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*, 56-115. doi:10.1037/0033-295X.103.1.56

Protopapas, A., Mouzaki, A., Sideridis, G. D., Kotsolakou, A., & Simos, P. G. (2013). The role of vocabulary in the context of the Simple View of Reading. *Reading and Writing Quarterly, 29*, 168-202. doi:10.1080/10573569.2013.758569

Ricketts, J., Bishop, D. V., & Nation, K. (2008). Investigating orthographic and semantic aspects of word learning in poor comprehenders. *Journal of Research in Reading, 31*, 117-135. doi:10.1111/j.1467-9817.2007.00365.x

Ricketts, J., Nation, K., & Bishop, D. V. M. (2007). Vocabulary Is Important for Some, but Not All Reading Skills. *Scientific Studies of Reading, 11*, 235-257. doi:10.1080/10888430701344306

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

Shaywitz, S. E. (1998). Dyslexia. *The New England Journal of Medicine, 388*, 307-312. doi:10.1056/NEJM199801293380507

Snowling, M. J., Stothard, S. E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., . . . Hulme, C. (2012). *York Assessment of Reading for Comprehension: Passage Reading. Australian Edition.* London, UK: GL Assessment.

Stothard, S. E., & Hulme, C. (1992). Reading comprehension difficulties in children: The role of language comprehension and working memory skills. *Reading and Writing, 4,* 245-246. doi:10.1007/BF01027150

Stothard, S. E., & Hulme, C. (1995). A comparison of phonological skills in children with reading comprehension difficulties and children with decoding difficulties. *Journal of Child Psychology and Psychiatry, 36,* 399-408. doi:10.1111/j.1469-7610.1995.tb01298.x

Tong, X., Deacon, S. H., & Cain, K. (2014). Morphological and syntactic awareness in poor comprehenders: another piece of the puzzle. *Journal of Learning Disabilities, 47,* 22-33. doi:10.1177/0022219413509971

Tong, X., Deacon, S. H., Kirby, J. R., Cain, K., & Parrila, R. (2011). Morphological awareness: A key to understanding poor reading comprehension in English. *Journal of Educational Psychology*, *103*, 523-534. doi:10.1037/a0023495

Wiederholt, J., & Bryant, B. (1992). *Gray Oral Reading Test Third Edition*. Austin, TX: Pro-Ed.

Woodcock, R. (1987). *Woodcock Reading Mastery Tests - Revised*. Circle Pines, MN: AGS.

Yuill, N., & Oakhill, J. (1991). *Children's problems in text comprehension: An experimental investigation*. Cambridge, UK: Cambridge.

# Chapter Two:
# Similar but different - differences in comprehension diagnosis on the Neale Analysis of Reading Ability and the York Assessment of Reading for Comprehension

**Abstract**

Background: The York Assessment of Reading for Comprehension (YARC) is increasingly being used in preference to the Neale Analysis of Reading Ability (NARA) in educational and research settings. However, little is known about how the two assessments compare.

Aims: To determine the extent to which the NARA and YARC make the same reading comprehension diagnoses, and to explore the reasons for any differences.

Sample: Ninety-five Australian children aged 8 to 12.

Methods: Reading comprehension was assessed using the NARA and the YARC. Reading accuracy and fluency were also assessed. A Principal Components Analysis and a series of correlations determined the dependence of the NARA and YARC on reading accuracy and fluency abilities. An analysis of the linguistic characteristics of the NARA and YARC passages was also carried out.

Results: The NARA tended to diagnose a greater number of children as below average than the YARC. For poor readers, this was due to the NARA's greater dependence on reading accuracy. For good readers, this was due to the fact that children answered a greater number of comprehension questions on the NARA, and these questions were based on more challenging passages.

Conclusions: Although the NARA and YARC are similar in format, they do not always result in the same comprehension diagnosis. The YARC is less dependent on reading accuracy skills and may be more suitable for comprehension assessment of children with word reading difficulties. However, it may not detect subtle comprehension difficulties in children who have good reading accuracy.

## Introduction

It is widely acknowledged that tests of reading comprehension vary in the underlying skills that they assess (Cain & Oakhill, 2006; Cutting & Scarborough, 2006; Keenan, Betjemann, & Olson, 2008; Keenan & Meenan, 2014; Kendeou, Papadopoulos, & Spanoudis, 2012; Nation & Snowling, 1997). This has consequences for educational practice, because tests which tap different underlying skills may lead to different children being diagnosed with comprehension difficulties (Keenan & Meenan, 2014). It also has consequences for research – researchers may draw conflicting conclusions as a result of differences in the reading comprehension assessments they use (Cutting & Scarborough, 2006; Eason, Goldberg, Young, Geist, & Cutting, 2012; Keenan et al., 2008).

This paper discusses this issue in relation to two widely used assessments of reading comprehension, the Neale Analysis of Reading Ability Third Edition (NARA; Neale, 1999) and the York Assessment of Reading for Comprehension (YARC; Snowling et al., 2012). Until recently, the NARA was arguably the most widely used assessment of reading comprehension in Australia and the UK, in both educational and research contexts. However, since its release, the YARC is increasingly replacing the NARA in both these countries (GL Assessment, 2014; Howe, 2013; Ricketts, 2014).

Both the NARA and YARC measure passage comprehension and passage reading abilities and have a very similar format (Ricketts, 2014). However, it is nonetheless possible that they could differ in their rates of diagnosis of reading comprehension difficulties. This study aimed to determine whether or not this was the case for children in the upper primary years. We first demonstrate that the NARA and YARC differ in terms of reading comprehension diagnosis, and then explore the possible reasons for these differences. By doing so, we will enable educators, clinicians and researchers to make an informed decision if considering a transition from the NARA to the YARC.

# Method

## Participants

Ninety-five children between the ages of 8 and 12 participated in this study. Children were in Australian school Grades 3 to 6 (equivalent to four to seven years of schooling). Fifty-five of the children were female and 40 were male.

Sixty-five participants came from a Catholic primary school in a middle-class area of Sydney. These children were assessed during the screening phase of a training study which explored the effects of oral vocabulary intervention on reading comprehension (Colenbrander, Kohnen, Smith-Lock, & Nickels, 2015). Teachers of Grades 3, 4 and 5 classes were asked to nominate children who had poor reading comprehension skills relative to their reading abilities, and also to nominate children who had average reading accuracy and reading comprehension skills for their age. All participants had English as their primary language and had been attending school in Australia since Kindergarten. Permission slips were distributed to parents of the nominated children. All children who returned parental permission slips and gave verbal consent were assessed.

A further 30 participants were recruited via the Neuronauts Brain Science Club at Macquarie University, Sydney. Children and their parents who had voluntarily joined the club were sent a newsletter advertising various research participation options. Parents contacted the researcher directly if interested in participating. The 30 participants who joined the study via Neuronauts were initially recruited as controls for another study. These 30 children were in Grades 4, 5 and 6 and came from a variety of areas around Sydney, attending both government and private schools. They also had English as their primary language and had been attending school in Australia since Kindergarten. These children had no reported history of reading or language difficulties. All Neuronauts participants returned parental permission slips and gave verbal consent.

**Assessment**

Participants were assessed individually by the first author in a single session of approximately 45 minutes, administered either in a quiet room at the child's school or in a testing laboratory at Macquarie University. Participants were administered the NARA, followed by the Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1999), the Castles and Coltheart Reading Test 2 (CC2; Castles et al., 2009) and the YARC.

Form 1 of the NARA (Neale, 1999) was administered according to the standard instructions. Children read aloud a series of short passages and for each were asked several open-ended comprehension questions, to which they were required to provide a verbal response. Children's reading errors were corrected by the tester. If a child made a specified number of errors on a passage (16 or more errors for most passages), he or she was not asked any comprehension questions for that passage and the test was discontinued. If the child made less than the specified number of errors, comprehension questions were asked and the child progressed to the next passage. The time taken to read each passage was recorded. The test returned three scores, a comprehension score, a reading accuracy score and a reading rate score. Reliability and validity information is reported in the NARA manual (Neale, 1999).

There are three versions of the YARC, each catering to a different age level – YARC Early Years (Hulme et al., 2009), YARC Passage Reading (Snowling et al., 2012), and YARC Secondary (Snowling et al., 2010). This allows children to be tested on comparable assessments throughout their schooling. For the purpose of this research, we are focusing on YARC Passage Reading because like the NARA, this version is designed to be used with primary school children. We report data on the Australian standardization of the YARC. However, the differences between the Australian edition of the test and the UK edition are minor (three small wording changes were made; Snowling et al., 2012), and therefore our conclusions are likely to be broadly applicable to the UK edition of the test.

Form A of the YARC Passage Reading Australian Edition (Snowling et al., 2012) was administered according to standard instructions. On this test, children read passages aloud and answered open-ended comprehension questions, similar to the NARA. Accuracy errors were also corrected by the tester. However, the discontinuation rule was different. On the YARC, children read aloud an initial passage matched to their grade level. If a child made less than a specified number of accuracy errors (usually 20 errors) on that passage, s/he was asked a series of open-ended questions about the passage. If the child responded correctly to 5 or more of these questions, s/he was asked to read a passage one level higher than the starting passage, and asked comprehension questions about this passage. If a child did not meet these accuracy or comprehension cut-offs, s/he was administered a passage the level below the starting passage, until s/he had read 2 passages with at least 2 comprehension questions correct for at least one level. The test contained both fiction and non-fiction passages and children read at least one passage of each type. Like the NARA, the test returned a comprehension score, an accuracy score and a rate score. Reliability and validity information is available in the YARC manual (Snowling et al., 2012).

We also assessed children's word reading accuracy and fluency to provide a measure of their reading skills independently of the two reading comprehension assessments. The CC2 (Castles et al., 2009) was administered to test single word reading accuracy. The CC2 returns three scores, a regular word reading score, an irregular word reading score, and a nonword reading score. We report only the irregular word and nonword scores, which index lexical and sublexical reading skills respectively. Forty nonwords and 40 irregular words were presented for reading aloud interspersed with each other, and in order of increasing difficulty. A stopping rule of five consecutive errors applied to each item type.

Form B of the TOWRE (Torgesen et al., 1999) was administered to test word reading fluency. The TOWRE has two subtests. For the Sight Word Efficiency subtest, children were asked to read lists of words as fast as they could. For the Phonemic Decoding Efficiency subtest, children were instructed to read lists of nonsense words as fast as they could. The child's total score was the number of stimuli they read correctly within 45 seconds. Recently published Australian norms were used, as the original TOWRE norms have been shown to overestimate the performance of Australian children (Marinus, Kohnen & McArthur, 2013).

## Results

### Descriptive statistics

Means and standard deviations of standard scores for each measure are shown in Table 1. Standard scores (mean of 100 and standard deviation of 15) are used to allow us to compare the results of our sample to that of the published normative samples for each test.

**Table 1**

***Participant Standard Score Means and Classifications***

| | Standard score means | | | Classifications | | |
|---|---|---|---|---|---|---|
| Measure | Mean | SD | Range | Below average | Average | Above average |
| NARA Comprehension | 88.73 | 9.95 | 69 - 113 | 32 | 63 | 0 |
| YARC Comprehension | 96.42 | 10.86 | 75 - 125 | 17 | 73 | 5 |
| NARA Rate | 101.18 | 11.7 | 71 - 130 | 3 | 79 | 13 |
| YARC Rate | 97.69 | 12.39 | 70 - 126 | 13 | 74 | 8 |
| NARA Accuracy | 101.23 | 11.44 | 73 - 129 | 8 | 79 | 8 |
| YARC Accuracy | 94.46 | 10.8 | 70 - 117 | 19 | 75 | 1 |
| TOWRE Sight Word Efficiency | 100.01 | 12.84 | 69 - 127 | 14 | 69 | 12 |
| TOWRE Phonemic Decoding Efficiency | 98.09 | 14 | 65 - 139 | 18 | 66 | 11 |
| CC2 Irregular Words | 99.76 | 11.62 | 69 - 124 | 11 | 77 | 7 |
| CC2 Nonwords | 92.98 | 12.1 | 58 - 126 | 24 | 68 | 3 |

*Note.* Total number of participants was 95. The CC2 does not provide standard scores, therefore standard scores were converted from z-scores. Standard scores have a mean of 100 and a standard deviation of 15. Standard scores between 85 and 115 were considered to be within the average range. A standard score of 84 or below was considered to be below average. A score of 116 or above was considered above average.

**Reading accuracy and fluency.**

Teachers had been asked to nominate participants with no known history of reading difficulties, and with either average or below average reading comprehension. Consequently, as expected, the majority of the reading accuracy and fluency scores were close to the standardization sample mean (One-sample z-tests: NARA Accuracy, $z = 0.80$, $p = 0.42$, NARA Rate, $z = 0.77$, $p = 0.44$, YARC Rate, $z = -1.50$, $p = 0.13$, TOWRE Sight Word Efficiency, $z = 0.01$, $p = 0.42$, TOWRE Phonemic Decoding Efficiency, $z = -1.24$, $p = 0.21$; CC2 Irregular words, $z = -0.06$, $p = 0.88$). However, YARC Accuracy scores were significantly lower than the test's standardization sample, $z = -3.60$, $p < 0.001$, as were CC2 Nonword Reading scores, $z = -4.56$, $p < 0.001$. Indeed, a substantial proportion of individual children were below average on at least one reading accuracy or fluency score. For example, 24 participants (25% of the sample) were classified as below average on their CC2 nonword score (see Table 1). This suggests that classroom teachers may not have a good understanding of the difference between word reading difficulties and reading comprehension difficulties. It also raises the possibility that there are a number of children in mainstream classrooms who have undiagnosed word reading difficulties.

**Reading comprehension.**

Given that our sample was initially intended to identify children with specific reading comprehension difficulties (as well as age-matched controls), we expected that the mean standard scores for the reading comprehension assessments would be significantly lower than the standardization means (i.e. lower than 100). This was true for both the NARA, $z = -7.32$, $p < 0.001$, and the YARC, $z = -2.33$, $p = 0.02$. In addition, participants' standard scores on the NARA were significantly lower than their standard scores on the YARC, $t(94) = -7.22$, $p < 0.001$.

**To what extent do diagnoses of reading comprehension difficulties differ across tests?**

To answer this question, we calculated the number of children who could be classified as average, below average, or above average on their comprehension scores for both the NARA (Neale, 1999) and YARC (Snowling et al., 2012). As is standard in clinical practice and research (e.g., Catts, Hogan, & Fey, 2003; Elwer, Keenan, Olson, Byrne, & Samuelsson, 2013; Nation, Clarke, Marshall, & Durand, 2004; Nation, Snowling, & Clarke, 2007; Ricketts, Sperring, & Nation, 2014), children were considered to be within the average range if they obtained a standard score of between 85 and 115 (between the 16th and 85th percentile). Scores of 116 and above were considered above average, while scores of 84 and below (the bottom 15% of participants) were considered below average. The percentage agreement across tests was then calculated.

Seventy-one percent of children received the same comprehension diagnosis on both the NARA and the YARC. In other words, nearly a third of children received different diagnoses on the two tests. There was a significant difference in how the two tests classified participants, (Wilcoxon signed rank test $z = -3.78$, $p < 0.001$), with 19 children being diagnosed as average on the YARC but below average on the NARA, and five children being classified as above average on the YARC but average on the NARA. By contrast, only four children were diagnosed as below average on the NARA but average on the YARC, and no children were classified as above average on the NARA and average on the YARC (see Table 1).

We looked at each grade level separately in order to determine the distribution of comprehension classifications across grade levels (see Table 2). For this analysis, we combined Grade 5 and Grade 6 results because there were only seven Grade 6 participants (the pattern of results was the same as when Grade 5 and Grade 6 were considered separately).

**Table 2**

***Number of Participants in each Grade Receiving Average, Below Average or Above Average Diagnoses***

| | Grade 3 | | | Grade 4 | | | Grade 5/6 | | |
|---|---|---|---|---|---|---|---|---|---|
| Assessment | Below average | Average | Above average | Below average | Average | Above average | Below average | Average | Above average |
| NARA Comprehension | 18 | 8 | 0 | 9 | 25 | 0 | 5 | 30 | 0 |
| YARC Comprehension | 8 | 17 | 1 | 5 | 26 | 3 | 4 | 30 | 1 |
| NARA Accuracy | 5 | 21 | 0 | 1 | 29 | 4 | 2 | 30 | 3 |
| YARC Accuracy | 5 | 21 | 0 | 6 | 27 | 1 | 8 | 27 | 0 |
| NARA Fluency | 1 | 23 | 2 | 1 | 30 | 3 | 1 | 30 | 4 |
| YARC Fluency | 3 | 23 | 0 | 4 | 27 | 3 | 6 | 26 | 3 |

The NARA diagnosed significantly more children as below average than the YARC in Grades 3 and 4 (Wilcoxon signed rank test: Grade 3, z = -3.051, $p$ = 0.002; Grade 4, z = -2.646, $p$ = 0.008), but not in Grades 5 and 6 (z = -0.707, p = 0.480). However, a homogeneity test (Leach, 1979) showed that the pattern of diagnosis was not significantly different across grade levels, $X^2$ = 3.22, p = 0.20. Thus, although the trend towards a greater number of below average diagnoses on the NARA than on the YARC was less evident in Grades 5 and 6, there was no evidence to indicate that the trend in Grades 5 and 6 was significantly different from that in lower grades.

Together, these results show that the NARA diagnoses a greater number of children with reading comprehension difficulties (especially in the middle primary grades). The reasons for this will be explored below.

## What factors contribute to differences in reading comprehension diagnosis?

### Dependence on decoding abilities.

Previous research has shown that a child's comprehension score on the NARA is dependent to some extent on his or her word reading abilities (Spooner, Baddeley, & Gathercole, 2004). Children are discontinued on the NARA if they make a specified number of accuracy errors, therefore poor readers (children with poor reading accuracy) are likely to read fewer passages on the NARA than good readers (children with good reading accuracy) of the same age and answer fewer comprehension questions, which may limit their comprehension score (Spooner et al., 2004). However, on the YARC, children's comprehension scores will always be calculated based on the same number of questions, regardless of their level of reading accuracy. This could be one reason why the NARA diagnoses a greater number of children with reading comprehension difficulties.

In order to determine whether this was true, we calculated the average number of comprehension questions answered by good and poor readers in each grade and compared these results using Mann-Whitney U tests. Children were considered to be poor readers if their standard score on any of the reading accuracy or fluency tests was 85 or below (if they fell within the bottom 15% of participants). Children whose standard scores on all of the reading accuracy and fluency tests fell above 85 were considered to be good readers. As shown in Table 3, poor readers in all grades read significantly fewer passages and answered significantly fewer questions than good readers in those grades.

**Table 3**

***Average Number of Passages Read and Questions Attempted on the Neale Analysis of Reading (Neale, 1999)***

| | Good Readers | | | Poor Readers | | | |
|---|---|---|---|---|---|---|---|
| Grade | n | Mean | SD | n | Mean | SD | Mann-Whitney U |
| Passages Read | | | | | | | |
| 3 | 14 | 4 | 0.97 | 12 | 2.83 | 0.58 | <0.001** |
| 4 | 22 | 4.82 | 0.39 | 12 | 3.58 | 1.00 | <0.001** |
| 5/6 | 20 | 4.05 | 0.22 | 15 | 3.40 | 0.99 | 0.03* |
| Questions Attempted | | | | | | | |
| 3 | 14 | 33.71 | 7.80 | 12 | 20.33 | 5.52 | <0.001** |
| 4 | 22 | 38.55 | 3.16 | 12 | 28.67 | 7.97 | <0.001** |
| 5/6 | 20 | 32.40 | 1.79 | 15 | 26.93 | 8.07 | 0.03* |

*Note*. * $p < 0.05$ ** $p < 0.01$ (one tailed)

***Principal components analysis.***

Given that poor readers did in fact answer fewer comprehension questions on the NARA, we wished to further explore the possibility that the NARA comprehension score was more reliant on reading abilities than the YARC. We therefore used principal components analysis with oblique rotation to determine whether the NARA or YARC comprehension scores would load on a reading accuracy or fluency factor.

The analysis was conducted on the percentile ranks for the NARA and YARC comprehension, rate and accuracy scores, the TOWRE Sight Word Efficiency and Phonemic Decoding Efficiency scores, and the CC2 Irregular Word and Nonword reading scores. The Kaiser-Meyer-Olkin (KMO) measure verified that our sample size was adequate for an analysis of this type, KMO = 0.86 (Field, 2009). In addition, all KMO values for individual variables were >0.78, which is above the acceptable limit of 0.5 (Field, 2009). Bartlett's test of sphericity, $X^2(45) = 634.68$, $p < 0.001$, demonstrated that correlations between items were sufficiently large to justify the use of principal components analysis.

Three factors emerged with eigenvalues greater than 1. Together, these three factors accounted for 77.16% of the variance. Factor loadings are shown in Table 4.

According to Stephens (2002), given multiple comparisons and a sample size of approximately 100, a factor loading should be greater than 0.512 to be considered significant at $\alpha = 0.01$. Using this criterion, the CC2 Nonword and Irregular word scores, YARC and NARA accuracy scores, and TOWRE Phonemic Decoding Efficiency score formed a single factor, which we refer to as the accuracy factor. The NARA and YARC comprehension scores formed a comprehension factor, while the NARA rate score and TOWRE Sight Word Efficiency score formed a third factor, which we refer to as the fluency factor. The factor loadings for the YARC rate score did not reach significance on any of the three factors and therefore was not included in any single factor (loadings were 0.40 on the accuracy factor, 0.34 on the comprehension factor and 0.46 on the fluency factor).

**Table 4**

***Pattern Matrix Showing Factor Loadings of NARA, YARC, CC2 and TOWRE scores***

| Test score | Comprehension Factor | Accuracy Factor | Fluency Factor |
|---|---|---|---|
| NARA Comprehension | 0.72* | 0.18 | 0.05 |
| YARC Comprehension | 0.89* | -0.02 | -0.01 |
| NARA Accuracy | 0.11 | 0.84* | 0.05 |
| YARC Accuracy | 0.15 | 0.89* | -0.08 |
| CC2 Nonwords | -0.09 | 0.96* | -0.10 |
| CC2 Irregular Words | 0.10 | 0.66* | 0.07 |
| NARA Rate | 0.18 | -0.16 | 0.92* |
| YARC Rate | 0.34 | 0.40 | 0.46 |
| TOWRE Phonemic Decoding | -0.09 | 0.70* | 0.39 |
| TOWRE Sight Words | -0.20 | 0.26 | 0.78* |

* Significant at $\alpha = 0.01$

Thus, the principal components analysis showed that neither comprehension score loaded significantly on the accuracy or fluency factors. However, it is worth noting that the loading on the accuracy factor was greater for the NARA (0.182) than for the YARC (-0.024).

***Correlations.***

Given the results of the principal components analysis, a composite accuracy score was calculated by averaging CC2 Nonword, CC2 Irregular word, YARC accuracy, NARA accuracy, and TOWRE Phonemic Decoding Efficiency percentile ranks. Similarly, a composite fluency score was calculated by averaging NARA rate and TOWRE Sight Word

Efficiency percentile ranks. Pearson correlation coefficients for correlations between NARA and YARC comprehension scores and the accuracy and fluency composites were calculated.

**Table 5**

***Correlations between Neale Analysis of Reading (NARA) Comprehension Scores, York Assessment of Reading for Comprehension (YARC) Comprehension Scores, the Decoding Composite and the Fluency Composite***

| | NARA Comprehension | YARC Comprehension |
|---|---|---|
| NARA Comprehension | 1.00 | .48** |
| YARC Comprehension | .48** | 1.00 |
| Accuracy Composite | .47** | .31** |
| Fluency Composite | .35** | .26* |

*Note*. n = 95 *$p$ < 0.05 **$p$<0.01

The comprehension scores of both tests were moderately and significantly correlated with accuracy and fluency (see Table 5). The correlations between NARA comprehension and the accuracy and fluency composites were higher than the correlations with YARC comprehension. The differences between the two tests in the size of these correlations was marginally significant, z = 1.69, $p$ = 0.05 (Lee & Preacher, 2013). There was no difference between the size of the correlations between the two comprehension scores and fluency, z = 0.90, $p$ = 0.18 (Lee & Preacher, 2013).

In order to further explore the result for accuracy, we calculated the correlations between comprehension and accuracy scores separately for good and poor readers. For good readers, the correlation between NARA comprehension and the accuracy composite was small and non-significant. The correlation between YARC comprehension and accuracy was also small, but significant (see Table 6). For poor readers, both NARA and YARC comprehension scores correlated moderately and significantly with the accuracy composite.

**Table 6**

***Correlations between Comprehension Scores and the Accuracy Composite for Good and Poor Readers***

| | NARA Comprehension | YARC Comprehension |
|---|---|---|
| Good Readers | | |
| NARA Comprehension | 1.00 | .47** |
| YARC Comprehension | .47** | 1.00 |
| Accuracy Composite | .21 | .30* |
| Poor Readers | | |
| NARA Comprehension | 1.00 | 0.50** |
| YARC Comprehension | .50** | 1.00 |
| Accuracy Composite | .57** | .38* |

*Note*. Good readers: n = 56 Poor readers: n = 39 *$p < 0.05$ **$p<0.01$

The correlations between NARA comprehension and the accuracy composite were significantly different for good and poor readers, z = -2.01, $p = 0.02$, but not for the YARC, z = -0.419, $p = 0.34$ (Preacher, 2002). The differences between the two tests in the size of the correlations between comprehension and accuracy were not significant for good readers, z = -0.66, $p = 0.26$, but approached significance for poor readers, z = 1.36, $p = 0.09$ (Lee & Preacher, 2013).

In summary, there was some evidence that comprehension scores on the NARA were more reliant on reading accuracy than YARC comprehension scores. Breaking the results down by reading skill, there was evidence that the NARA comprehension scores of poor readers were more reliant on reading accuracy than those of good readers. However, for poor readers, the difference between NARA and YARC comprehension scores in terms of reliance on reading accuracy did not quite reach significance, and there was no difference for children who were good readers. In fact, nearly half of the children whose comprehension skills were classified as average on the YARC, but below average on the NARA, had average or above average reading accuracy and fluency scores (8 children out of 19). We therefore investigated whether the difference in diagnosis between the two tests may also be due to differences in the number of comprehension questions administered to children on the NARA and the YARC, regardless of the child's level of reading ability.

**Differences in number of comprehension questions answered on the NARA and the YARC by readers of all ability levels.**

As outlined above, the YARC comprehension score is always calculated on the basis of 16 questions asked about two passages, but the number of questions answered on the NARA is dependent on a child's text reading abilities. In the section above, we established that poor readers will answer fewer questions than good readers on the NARA. However, the format of the NARA means that all children, even poor readers, will generally attempt more comprehension questions than they will on the YARC. Table 3 (see above) shows that even the poorest readers in the study (poor readers in Grade 3) answered on average 20 comprehension questions on the NARA, and good readers answered on average over 30 questions each, as opposed to answering only 16 questions on the YARC. This may be part of the reason why the NARA diagnosed more children with comprehension difficulties. A larger number of questions may allow greater sensitivity in the detection of comprehension difficulties, particularly for good readers, who answer the highest number of questions.

The analysis reported above demonstrated that the tendency for the NARA to diagnose more children with poor comprehension was greater in the younger grades than in

**Table 7**

***Percentage of Good and Poor Readers Completing Each Passage Level***

| | Good readers | | | | | | Poor readers | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Grade 3 | | Grade 4 | | Grades 5 and 6 | | Grade 3 | | Grade 4 | | Grades 5 and 6 | |
| Passage level | NARA | YARC | NARA | YARC | NARA | YARC | NARA | YARC | NARA | YARC | NARA | YARC |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 58.33 | 8.33 | 0 | 0 | 6.67 | 0 |
| 2 | 100 | 42.86 | 86.36 | 0 | 5.00 | 0 | 100 | 50.00 | 91.67 | 0 | 13.33 | 0 |
| 3 | 100 | 100 | 100 | 31.82 | 100 | 0 | 83.33 | 91.67 | 100 | 41.67 | 93.33 | 6.67 |
| 4 | 92.86 | 57.14 | 100 | 100 | 100 | 40.00 | 33.33 | 50.00 | 91.67 | 100.00 | 86.67 | 60.00 |
| 5 | 78.57 | 0 | 100 | 68.18 | 100 | 100 | 8.33 | 0 | 41.67 | 58.33 | 73.33 | 93.33 |
| 6 | 50.00 | 0 | 95.45 | 0 | 100 | 60.00 | 0 | 0 | 31.33 | 0 | 66.67 | 40.00 |

*Note.* NARA = Neale Analysis of Reading (Neale, 1999). YARC = York Assessment of Reading for Comprehension (YARC). Participants aged 8 or 9 begin the NARA on passage level 2 and are awarded full credit for passage level 1 if they reach a basal level of accuracy on level 2 (no more than 2 errors). Participants aged 10 or over begin the NARA on passage level 3 and are awarded full credit for preceding passages if they reach the basal accuracy level. On the YARC, participants begin by reading the passage level appropriate to their age and then read the level above or below their starting level, depending on their accuracy and comprehension scores (see test manuals for further details).

Grades 5 and 6. It is therefore important to note that, in line with the NARA manual (Neale, 1999), children began on different levels of the NARA depending on their age. Children who were aged 8 and 9 began on Level 2 and children who were aged 10 and over began on Level 3. If children reached a basal level of accuracy on the starting passage, they continued to the next passage level and were credited full marks for any unadministered passage level (Level 1 and/or Level 2 depending on age). For this reason, good readers in Grades 3 and 4 generally answered slightly more questions than good readers in Grade 5 and 6, who were automatically awarded marks for Level 2 questions (see Table 3 above).

**Relative difficulty of passages.**

Another possible reason why the NARA appeared to diagnose more children with comprehension difficulties in Grades 3 and 4 was the fact that good readers in those grades who scored sufficiently well on earlier passages would be asked to read the most difficult passage levels of the NARA (passages 5 and 6; see Table 7). The content of these passages was likely to have been very challenging for children in Grades 3 and 4, but comparatively less challenging for older children.

We therefore examined the difficulty levels of NARA and YARC passages using CohMetrix version 3.0 (McNamara, Louwerse, Cai, & Graesser, 2011), a computational tool which analyzes texts on numerous different variables. Based on previous research, we selected a number of key variables to report in our analyses. We report measures of word length and frequency, because passages containing shorter and more frequent words are easier to decode and comprehend (Compton, Appleton, & Hosp, 2004; Freebody & Anderson, 1983; Graesser, McNamara, & Kulikowich, 2011; Menton & Hiebert, 1999; Miller et al., 2014; Ozuru, Rowe, O'Reilly, & McNamara, 2008). We also report measures beyond the word level which have been shown to play a role in passage difficulty, such as syntactic complexity (reported as sentence length in words), referential cohesion, and passage length

**Table 8**

***Passage Analyses***

| | Passages | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Neale Analysis of Reading | | | | | | York Assessment of Reading for Comprehension | | | | | |
| Measures | 1 | 2 | 3 | 4 | 5 | 6 | 1 | 2 | 3 | 4 | 5 | 6 |
| Number of sentences per passage | 4 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 19 | 9 | 13 | 11 |
| Number of words per passage | 26 | 52 | 73 | 97 | 117 | 141 | 66 | 98 | 155 | 182 | 192 | 222 |
| Mean number of words per sentence | 6.5 (1.0) | 6.5 (2.26) | 9.13 (2.98) | 12.13 (4.62) | 14.63 (7.87) | 17.63 (9.83) | 8.25 (3.04) | 12.25 (5.12) | 8.16 (3.53) | 20.22 (9.96) | 14.77 (7.77) | 20.18 (8.54) |
| Mean number of syllables per word | 1.15 (0.36) | 1.37 (0.65) | 1.41 (0.59) | 1.61 (0.77) | 1.73 (1.04) | 1.80 (1.05) | 1.27 (0.51) | 1.18 (0.39) | 1.33 (0.58) | 1.40 (0.69) | 1.37 (0.65) | 1.60 (0.92) |
| Mean number of letters per word | 3.39 (1.65) | 4.23 (2.08) | 4.52 (2.32) | 5.41 (2.52) | 5.12 (3.01) | 5.69 (2.82) | 3.96 (1.73) | 3.90 (1.56) | 4.29 (2.27) | 4.67 (2.21) | 4.55 (2.46) | 4.91 (2.54) |
| Mean log frequency for all words | 3.01 | 3.10 | 2.93 | 2.50 | 2.72 | 2.66 | 3.02 | 2.89 | 3.09 | 2.67 | 2.81 | 2.86 |
| Mean frequency for content words | 2.42 | 2.24 | 2.04 | 1.63 | 1.72 | 1.73 | 2.18 | 2.12 | 2.13 | 1.79 | 1.91 | 1.98 |
| Referential cohesion score (percentile) | 36.64 | 4.85 | 9.51 | 7.93 | 2.17 | 0.91 | 10.2 | 21.19 | 34.46 | 37.07 | 29.12 | 18.14 |

*Note.* Standard deviations are in parentheses. CohMetrix does not compute standard deviations for word frequency. A higher percentile of referential cohesion indicates a higher level of connection between words and ideas in the text (Graesser, McNamara and Kulikowich, 2011).

(Graesser et al., 2011; Keenan et al., 2008; McNamara, Graesser, & Louwerse, in press; Miller et al., 2014; Ozuru et al., 2008). Results are shown in Table 8.

Because different children completed different passages on the tests, it is difficult to generalize about minimum and maximum difficulty levels for each test. However, Table 8 shows that while YARC passages are generally longer in terms of number of words and sentences, levels 3 to 6 of the NARA contain on average longer and less frequent words, and the referential cohesion scores of these NARA passages are well below those of the YARC. Texts which are lower in referential cohesion may be more difficult to understand because readers are required to carry out inferences to connect relevant elements of the text (Britton & Gulgoz, 1991). This process is likely to be particularly challenging if readers are also faced with unfamiliar vocabulary items (Graesser et al., 2011; McNamara et al., in press). This suggests that the higher-level NARA passages may be more challenging than the higher-level YARC passages.

## Discussion and conclusions

This study assessed a sample of children in the upper primary years on the NARA (Neale, 1999) and the YARC (Snowling et al., 2012) to determine the extent to which the two tests differed in diagnosis of reading comprehension difficulties. Overall, the NARA tended to diagnose more children as having below average reading comprehension than the YARC, particularly in Grades 3 and 4 (4th and 5th year of schooling, ages 8 to10).

In line with previous research on differences between reading comprehension tests (Nation & Snowling, 1997; Spooner et al., 2004), we expected that a major reason for the differences in diagnosis across the two tests would be that the NARA relies to a greater extent on reading abilities. While our results showed a trend in this direction, differences between the NARA and the YARC in terms of reliance on reading ability did not quite reach significance.

Nevertheless, we did find that the NARA comprehension scores of poor readers were significantly more dependent on reading accuracy than the comprehension scores of good readers. This appears to be because poor readers answered significantly fewer comprehension questions on the NARA than good readers. However, greater dependence on word reading abilities cannot be the sole reason for differences in diagnosis between the tests: nearly half of the children who were diagnosed with below average comprehension on the NARA, but not on the YARC, had age appropriate reading accuracy and fluency skills.

In fact, our results indicate that the NARA may be a more sensitive test of reading comprehension than the YARC for children who are good readers. This is because, on the NARA, children will generally answer more questions from a larger number of passage levels, representing a wider range of difficulty. This is particularly apparent for good readers in Grades 3 and 4, who will answer 16 comprehension questions on two age-appropriate passages on the YARC, but will generally complete four or five passages on the NARA, including the highest, most challenging, passage.

In addition, the NARA passages tend to have lower levels of referential cohesion and contain longer and less frequent words than the YARC passages, and hence they are likely to be more challenging to understand. In other words, not only do children answer more questions on the NARA, but these questions relate to more difficult passages.

Of course, we cannot rule out the possibility that the NARA is in fact over-diagnosing reading comprehension difficulties. Some children who perform well on developmentally appropriate reading comprehension passages may perform poorly on the more challenging passages of the NARA, which may be taxing skills beyond those required for making suitable academic progress at the child’s grade level. Further research is required to determine how well children’s comprehension scores on the NARA and YARC reflect their academic performance in the classroom, and longitudinal research exploring the predictive abilities of the NARA and the YARC is also warranted.

Nonetheless, our results imply that when transitioning from the NARA to the YARC, clinicians and researchers should be alert to the possibility that the YARC may not detect milder or more subtle cases of comprehension difficulty. However, because the NARA comprehension scores of poor readers tend to depend on their reading abilities, the YARC is likely to be a better assessment option for clinicians or researchers working with populations of children with reading difficulties. The shorter administration time of the YARC is also an advantage in settings where assessment time is limited.

Given these results, it is important to note some limitations of this study. Firstly, this research was based on a sample of children collected for the purposes of another study and therefore our sample was not representative or random. Secondly, the study was carried out with children in Grades 3 to 6, while the NARA and the YARC are both designed for use with children in Grades 1 to 6, and results may differ in lower grades. Thirdly, the sample size was relatively small – in some cases, there were fewer than 35 participants per cell for the grade-by-grade analyses. Therefore, the results of these analyses should be treated with caution. Finally, our study was carried out on an Australian sample using the Australian edition of the YARC. However, given the minimal differences between the Australian and UK versions of the test (Snowling et al., 2012), it is likely that our findings are applicable beyond the Australian context. It is also worth noting that although the current sample was not strictly representative, it did contain children with a wide variety of different ability levels, enabling us to draw conclusions about both good and poor readers.

In summary, our data show that there are substantial differences in diagnosis between the NARA and the YARC. This means that the two tests are not interchangeable. The results of this study reinforce the recommendation that, wherever possible, a diagnosis of comprehension difficulty should not be made on the basis of a single comprehension test (Bowyer-Crane & Snowling, 2005; Cain & Oakhill, 2006; Keenan & Meenan, 2014). However, we do not suggest that a diagnosis of comprehension difficulty should only be made if a child scores poorly on more than one comprehension assessment. Rather, we suggest that multiple comprehension assessments can provide additional qualitative information about a child's comprehension skills. If a child receives different diagnoses across tests, the clinician or researcher should interpret this information based on their knowledge of the child, the child's reading accuracy or fluency skills, the testing circumstances, and the strengths or weaknesses of the assessments used, such as we have provided here. It is therefore hoped that the findings of this study will help researchers, schools and clinicians to make informed decisions about the use of the NARA and the YARC in particular, and reading comprehension assessment in general.

## References

GL Assessment. (2014). York Assessment of Reading for Comprehension: Improving literacy outcomes (City of Edinburgh Council). Retrieved 25 August 2014, from http://www.gl-assessment.co.uk/research-and-articles/york-assessment-reading-comprehension-improving-literacy-outcomes-city

Bowyer-Crane, C., & Snowling, M. J. (2005). Assessing children's inference generation: what do tests of reading comprehension measure? *British Journal of Educational Psychology, 75,* 189-201. doi:10.1348/000709904X22674

Britton, B. K., & Gulgoz, S. (1991). Using Kintsch's computational model to improve instructional text: Effects of repairing inference calls on recall and cognitive structures. *Journal of Educational Psychology, 83*, 329-345. doi:10.1037/0022-0663.83.3.329

Cain, K., & Oakhill, J. (2006). Assessment matters: issues in the measurement of reading comprehension. *British Journal of Educational Psychology, 76,* 697-708. doi:10.1348/000709905X69807

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

Catts, H. W., Hogan, T. P., & Fey, M. E. (2003). Subgrouping poor readers on the basis of individual differences in reading-related abilities. *Journal of Learning Disabilities, 36*, 151-164. doi:10.1177/002221940303600208

Colenbrander, D., Kohnen, S., Smith-Lock, K., & Nickels, L. (2015). Understanding the link between oral vocabulary and reading comprehension - a randomized controlled trial. Unpublished manuscript.

Compton, D. L., Appleton, A. C., & Hosp, M. K. (2004). Exploring the relationship between text-leveling systems and reading accuracy and fluency in second-grade students who are average and poor decoders. *Learning Disabilities Research and Practice, 19*, 176-184. doi:10.1111/j.1540-5826.2004.00102.x

Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading, 10*, 277-299. doi:10.1207/s1532799xssr1003_5

Eason, S. H., Goldberg, L. F., Young, K. M., Geist, M. C., & Cutting, L. E. (2012). Reader-text interactions: How differential text and question types influence cognitive skills needed for reading comprehension. *Journal of Educational Psychology, 104*, 515-528. doi:10.1037/a0027182

Elwer, S., Keenan, J. M., Olson, R. K., Byrne, B., & Samuelsson, S. (2013). Longitudinal stability and predictors of poor oral comprehenders and poor decoders. *Journal of Experimental Child Psychology, 115*, 497-516. doi:10.1016/j.jecp.2012.12.001

Field, A. (2009). *Discovering Statistics Using SPSS* (3rd ed.). London, UK: Sage.

Freebody, P., & Anderson, R. C. (1983). Effects of vocabulary difficulty, text cohesion and schema availability on reading comprehension. *Reading Research Quarterly, 18*, 277-294. doi:10.2307/747389

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher, 40*, 223-234. doi:10.3102/0013189x11413260

Howe, W. (2013). The use of assessments in the inclusive classroom. Paper presented at the Pedagogy in Practice Conference, Brisbane, Australia.

Hulme, C., Stothard, S. E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., & Snowling, M. J. (2009). *York Assessment of Reading for Comprehension (YARC): Early Reading*. London, UK: GL Assessment.

Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading, 12,* 281-300. doi:10.1080/10888430802132279

Keenan, J. M., & Meenan, C. E. (2014). Test differences in diagnosing reading comprehension deficits. *Journal of Learning Disabilities, 47*, 125-135. doi:10.1177/0022219412439326

Kendeou, P., Papadopoulos, T. C., & Spanoudis, G. (2012). Processing demands of reading comprehension tests in young readers. *Learning and Instruction, 22*, 354-367. doi:10.1016/j.learninstruc.2012.02.001

Leach, C. (1979). *Introduction to statistics: A nonparametric approach for the social sciences.* Chichester, UK: Wiley.

Lee, I. A., & Preacher, K. J. (2013). Calculation for the test of the difference between two dependent correlations with one variable in common [Computer software]. Retrieved from http://quantpsy.org

McNamara, D. S., Graesser, A. C., & Louwerse, M. M. (in press). Sources of text difficulty: Across the ages and genres. In J. P. Sabatini & E. Albo (Eds.), *Assessing reading in the 21st Century: Aligning and applying advances in the reading and measurement sciences.* Lanham, MD: R & L Education.

McNamara, D. S., Louwerse, M. M., Cai, Z., & Graesser, A. C. (2011). Coh-Metrix version 3.0 [Computer software]. Retrieved 20 December 2012, from http://cohmetrix.com

Menton, S., & Hiebert, E. H. (1999). *Literature anthologies: The task for first-grade readers.* Ann Arbor, MI: Centre for the Improvement of Early Reading Achievement (ERIC

Documentation Service No. ED436754). Retrieved from http://eric.ed.gov/?id=ED436754

Miller, A. C., Davis, N., Gilbert, J. K., Cho, S., Toste, J. R., Street, J., & Cutting, L. E. (2014). Novel approaches to examine passage, student and question effects on reading comprehension. *Learning Disabilities Research and Practice, 29*, 25-35. doi:10.1111/ldrp.12027

Nation, K., Clarke, P., Marshall, C. M., & Durand, M. (2004). Hidden language impairments in children: Parallels between poor reading comprehension and specific language impairment? *Journal of Speech, Language and Hearing Research, 47*, 199-211. doi:10.1044/1092-4388(2004/017)

Nation, K., & Snowling, M. J. (1997). Assessing reading difficulties: the validity and utility of current measures of reading skill. *British Journal of Educational Psychology, 67,* 359-370. doi:10.1111/j.2044-8279.1997.tb01250.x

Nation, K., Snowling, M. J., & 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. *Advances in Speech-Language Pathology, 9*, 131-139. doi:10.1080/14417040601145166

Neale, M. D. (1999). *The Neale Analysis of Reading Ability Third Edition*. Melbourne, Australia: ACER.

Ozuru, Y., Rowe, M., O'Reilly, T., & McNamara, D. S. (2008). Where's the difficulty in standardized reading tests: The passage or the question? *Behaviour Research Methods, 40*, 1001-1015. doi:10.3758/BRM.40.4.1001

Preacher, K. J. (2002). Calculation for the test of the difference between two independent correlation coefficients [Computer software]. Retrieved from http://quantpsy.org

Ricketts, J. (2014). *Reading: Research and assessment*. Retrieved 4 September 2014, from http://www.gl-assessment.co.uk/research-and-articles/reading-research-and-assessment-jessie-ricketts

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

Snowling, M. J., Stothard, S. E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., . . . Hulme, C. (2010). *York Assessment of Reading for Comprehension (YARC) Passage Reading Secondary Version*. London, UK: GL Assessment.

Snowling, M. J., Stothard, S. E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., . . . Hulme, C. (2012). *York Assessment of Reading for Comprehension: Passage Reading. Australian Edition*. London, UK: GL Assessment.

Spooner, A. L. R., Baddeley, A. D., & Gathercole, S. E. (2004). Can reading accuracy and comprehension be separated in the Neale Analysis of Reading Ability? *British Journal of Educational Psychology, 74*, 187-204. doi:10.1348/000709904773839833

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of Word Reading Efficiency*. Austin, TX: Pro-Ed.

# Chapter Three: Exploring the relationship between vocabulary skills and reading comprehension - a look at individual differences

**Abstract**

The relationship between reading comprehension and vocabulary skills is complex. In this study, the phonological, orthographic and semantic aspects of vocabulary knowledge were assessed in a sample of poor comprehenders and age-matched controls. In addition, assessments of syntax, listening comprehension and non-verbal reasoning were administered. Results were analyzed at both the group and individual level. It was found that the majority of poor comprehenders had vocabulary deficits, supporting the Lexical Quality Hypothesis. Vocabulary deficits predominantly took the form of impairments in semantic knowledge. However, a minority had good vocabulary skills. All participants scored below controls on at least one oral language task, including those who had good vocabulary skills, supporting the Simple View of Reading. Despite having poor semantic skills, the majority of participants had good irregular word reading skills, which challenges theories that propose that irregular word reading is reliant on semantic processing.

## Introduction

Approximately 10% of all primary school children have difficulties understanding written texts, despite being able to read aloud the words on the page (Nation & Snowling, 1997; Yuill & Oakhill, 1991). These children are known as poor comprehenders. There is a large body of research demonstrating that poor vocabulary skills are associated with reading comprehension difficulties (Catts, Adlof, & Weismer, 2006; Clarke, Snowling, Truelove, & Hulme, 2010; Landi & Perfetti, 2007; Nation, Clarke, Marshall, & Durand, 2004; Nation & Snowling, 1998, 1999), and some longitudinal and intervention research suggests that vocabulary difficulties may play a part in causing poor reading comprehension (e.g., Clarke, Henderson, & Truelove, 2010; Elwer, Keenan, Olson, Byrne, & Samuelsson, 2013). However, other studies have shown that there are poor comprehenders who do not have vocabulary difficulties (Cain & Oakhill, 1999; Tong, Deacon, & Cain, 2014). Thus, the role of vocabulary in reading comprehension is somewhat controversial.

One perspective on the role of vocabulary in reading comprehension is the Lexical Quality Hypothesis (Perfetti, 2007; Perfetti & Hart, 2002). According to this hypothesis, successful reading comprehension is reliant to a large extent on word knowledge – without the ability to read and understand the words in a text, one cannot comprehend it. In addition, good word-level knowledge allows processing resources to be allocated to higher-order comprehension processes (Perfetti & Hart, 2002).

The Lexical Quality Hypothesis states that word knowledge is comprised of three main constituents – phonological knowledge (the way a word sounds), orthographic knowledge (a word's written form), and semantic knowledge (what a word means). The quality of a lexical representation is determined by how well these different aspects of word knowledge are represented, how well they are integrated with each other and how efficiently they can be retrieved (Perfetti, 2007; Perfetti & Hart, 2002).

There is some evidence that poor comprehenders have relative strengths in phonological and orthographic abilities - they can perform as well as controls on phonological awareness and nonword repetition tasks and can learn new orthographic forms as well as controls (Catts et al., 2006; Nation et al., 2004; Ricketts, Bishop, & Nation, 2008). These findings are consistent with the fact that poor comprehenders, by definition, can read aloud without difficulty. However, they often have poor semantic skills, as measured using tasks such as synonym judgements, semantic fluency, semantic priming and verbal definition production (Nation & Snowling, 1998, 1999; Perfetti, 2007; Ricketts, Nation, & Bishop, 2007). They also tend to have difficulties learning new semantic representations (Nation, Snowling, & Clarke, 2007; Ricketts et al., 2008). This implies that poor vocabulary in poor comprehenders is a result of semantic difficulties, and suggests that semantic difficulties may therefore be a cause of poor reading comprehension (Nation et al., 2007; Ricketts et al., 2008).

A related finding is the fact that poor comprehenders have generally been reported to have poor irregular word reading abilities (Nation & Snowling, 1998; Ricketts et al., 2007). This is consistent with the Triangle model of reading (Plaut, McClelland, Seidenberg, & Patterson, 1996), which posits that reading aloud is achieved by a network of distributed phonological, orthographic and semantic codes. Under this model, all three aspects of word knowledge are required for successful reading aloud, but contributions from semantics are particularly important for irregular word reading, because irregular words have inconsistent mappings from orthography to phonology.

However, some studies have shown that it is possible for irregular words to be read without access to semantics (Blazely, Coltheart, & Casey, 2005; Nation & Cocksey, 2009). While inconsistent with the Triangle model, this finding is easily accommodated with the Dual Route model of reading (Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). In this model, nonwords and regular words are read via a sub-lexical route which converts letters into sounds using grapheme-phoneme correspondence rules. Irregular words are read by accessing a store of phonological representations, which can be accessed either via the word's meaning (semantics) or directly from the word's orthographic form. This model predicts the existence of poor comprehenders who have semantic difficulties, but intact irregular word reading abilities. This possibility has not yet been investigated in the literature.

In addition, the role of semantic difficulties in reading comprehension is challenged by the existence of studies in which poor comprehenders were selected to have adequate semantic skills for their age as measured on tasks such as written and spoken word-picture matching (e.g., Cain & Oakhill, 1999; Tong et al., 2014). These contrasting findings are not surprising if we consider that vocabulary can be measured in multiple different ways, some of which may tap semantic skills to a greater extent than others. Tasks used to assess vocabulary skills in children with reading comprehension difficulties have included word-picture matching (e.g., Catts et al., 2006; Silva & Cain, 2014), definition production (e.g., Clarke, Snowling, et al., 2010; Nation, Cocksey, Taylor, & Bishop, 2010), synonym production, picture naming, verbal analogies, and figurative word usage (e.g., Cain & Oakhill, in press; Nation & Snowling, 1998; Ouelette, 2006), as well as combinations of the above. These tasks differ in terms of the depth of semantic knowledge required (Ouelette, 2006) and they also rely to different degrees on expressive language and reasoning abilities.

Research has suggested that tasks assessing semantic depth (how much a child knows about a word) are more closely associated with reading comprehension skills than tasks assessing the breadth of semantic knowledge (the number of words a child knows; Ouelette, 2006; Tannenbaum, Torgesen, & Wagner, 2006). However, depth and breadth are not completely separable, and most semantic tasks will tap both to some degree (Cain & Oakhill, in press; Nation & Cocksey, 2009; Tannenbaum et al., 2006).

The use of different assessments is unlikely to be the only reason for differences between studies. Most research examining the links between reading comprehension and vocabulary has analyzed patterns at the group level. However, studies of individual differences in poor comprehenders (Cornoldi, de Beni & Passaglia, 1996; Nation et al., 2004; Cain & Oakhill, 2006) show that they are a heterogeneous population with different profiles of weakness. Notably, in the Nation et al. (2004) and Cain and Oakhill (2006) studies, there was evidence that some poor comprehenders had difficulties with vocabulary skills, while others did not. There was also evidence that some poor comprehenders had oral language deficits beyond the word level (for example, on tasks of syntax, listening comprehension and figurative language; Cain & Oakhill, 2006; Nation et al., 2004).

However, no study has yet teased apart the contribution of different aspects of vocabulary skill by examining individual poor comprehenders using a range of tasks designed to assess semantics and other aspects of word knowledge in multiple different ways. In addition, no study has looked at the irregular word reading skills of poor comprehenders at the individual level. These were the aims of the study described here: we wished to examine the vocabulary skills of individual poor comprehenders in detail, using a variety of tasks assessing semantic, orthographic and phonological aspects of word knowledge.

Using these measures, we were interested in how many of our poor comprehenders would display vocabulary deficits, and whether any other language deficits were evident in these children above the word level. We also wished to determine whether the vocabulary deficits of poor comprehenders were evident primarily at the semantic level, which would support previous claims that the vocabulary difficulties shown by poor comprehenders are primarily due to semantic weaknesses (Nation et al., 2007; Ricketts et al., 2008). Further, we were interested in how many of our participants would display irregular word reading deficits, and whether these would always be associated with semantic deficits. Finally, we wished to know whether tasks assessing semantic depth would be more sensitive to semantic difficulties in poor comprehenders than tasks assessing primarily semantic breadth, given demonstrated links between semantic depth and reading comprehension (Ouelette, 2006; Tannenbaum et al., 2006).

## Method

### Recruitment and screening

An initial sample of participants was recruited from a Catholic primary school in a middle-class area of Sydney. Teachers of classes in Grades 3, 4 and 5 (4th to 6th year of schooling) were asked to nominate children with average word reading abilities for their age and average or below average reading comprehension skills. Consent forms were distributed to the parents of the nominated children. Sixty-five children who returned consent forms and gave verbal consent participated in screening assessment.

Participants were screened for reading comprehension using Form 1 of the Neale Analysis of Reading Ability (NARA; Neale, 1999) and Form A of the York Assessment of Reading for Comprehension Passage Reading Australian Edition (YARC; Snowling et al., 2012). They were screened for reading accuracy using the Castles and Coltheart Reading Test 2 (CC2; Castles et al., 2009).

The NARA (Neale, 1999) is a test of reading comprehension during which participants read a series of passages aloud and are asked a series of open-ended comprehension questions about the passages. The number of passages read is determined by a child’s age and passage reading accuracy.

The YARC (Snowling et al., 2012) also requires children to read passages aloud and answer open-ended questions. However, on the YARC children generally read aloud and answer questions on two passages. Passage levels are determined by the child’s age, reading ability and comprehension ability.

The CC2 (Castles et al., 2009) is a test of single word reading accuracy. Forty nonwords and 40 irregular words are presented to children interspersed with each other, and in order of difficulty . Children read the words or nonwords aloud from cards. A stopping rule of five consecutive errors applied to each item type.

At screening, children were also assessed on the Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1999), a test of word reading fluency. This was not used as a diagnostic measure, but was used to explore whether there were any differences in fluency skills between the two groups. The TOWRE contains two subtests, a Sight Word Efficiency subtest, on which children read lists of words as fast as they can, and a Phonemic Decoding Efficiency subtest, on which children read lists of nonsense words as fast as they can. The child’s total score is equal to the number of words they read correctly within 45 seconds. Because TOWRE norms were collected in the US and have been shown to overestimate the performance of Australian children, recently published Australian norms (Marinus, Kohnen & McArthur, 2013) were used.

In order to be considered a poor comprehender, a child's reading accuracy scores on both CC2 subtests had to be within the average range (standard scores between 85 and 115, z-scores between 1 and -1). In addition, they had to have a reading comprehension standard score of less than 85 on either the NARA, the YARC, or both. This score also had to be at least one standard deviation below their lowest accuracy score.

Controls were children for whom all reading accuracy and reading comprehension scores fell within the average range.

Screening revealed 13 participants who fitted our criteria for specific reading comprehension difficulties. Nine participants from the Catholic school sample met our criteria for controls. As the number of participants was insufficient to carry out some statistical analyses, we recruited further participants through the Neuronauts Brain Science Club of the ARC Centre for Excellence in Cognition and its Disorders (CCD) at Macquarie University, Sydney. This is a club for children and parents interested in participating in cognitive research. Children and their parents subscribe to receive a newsletter advertising various research participation options. Parents contacted the first author directly if interested in participating in the study.

Of the 30 Neuronauts screened, 11 met control criteria and could be matched to poor comprehender participants in terms of age and grade level at the time of language and cognitive assessment (see below).

Thus, our final sample consisted of 13 (11 female) poor comprehenders and 20 (9 female) reading-accuracy matched controls. Children were in Australian school grades 4 to 6 at the time of experimental testing and were aged between 9 and 11. All participants spoke English as their primary language and had been attending school in Australia since Kindergarten. They attended a variety of government, Catholic and independent schools. None had any history of reading or oral language difficulties.

Means standard scores for poor comprehenders and controls on each screening measure are shown in Table 1 along with the results of Mann-Whitney U tests comparing the groups. As expected, poor comprehenders had significantly lower standard scores than controls on the reading comprehension measures, but not on the reading measures.

**Table 1**

***Mean Standard Scores of Poor Comprehenders and Controls at Screening***

| Test | Poor Comprehenders | Controls | Mann-Whitney U |
|---|---|---|---|
| NARA Text Reading Comprehension | 81.00 (6.10) | 97.36 (7.90) | <0.001** |
| YARC Text Reading Comprehension | 89.92 (9.05) | 100.70 (5.44) | 0.001* |
| NARA Text Reading Accuracy | 104.31 (8.62) | 104.95 (5.59) | 0.33 |
| YARC Text Reading Accuracy | 100.15 (6.90) | 99.00 (7.16) | 0.76 |
| CC2 Nonword Reading Accuracy | 100.66 (5.63) | 96.36 (7.90) | 0.13 |
| CC2 Irregular Word Reading Accuracy | 106.60 (8.25) | 103.41 (5.64) | 0.21 |

*Note.* Standard deviations are in parentheses. *$p < 0.05$ **$p<0.01$

**Language and cognitive assessment**

Participants were assessed on a battery of assessments testing a range of aspects of vocabulary skill (phonological, orthographic and semantic). Participants were assessed on both standardized tests and experimenter-designed tasks assessing children's knowledge of orthography, phonology and semantics of the same words, so that we could determine the relative strength of different aspects of vocabulary (Perfetti & Hart, 2002). Semantic skills were assessed using a variety of different tasks which varied in terms of depth of semantic processing and expressive language demands.

Participants were also assessed on two tasks assessing oral language beyond the word level, so that it was possible to determine whether their deficits extended to broader oral language skills. In addition, participants were assessed on non-verbal working memory and reasoning tasks to ensure that their reading comprehension difficulties were not a consequence of more general intellectual difficulties.

Participants were assessed individually by the first author and were tested either in a quiet room at their school or in a testing laboratory at Macquarie University. Assessment took approximately two and a half hours per child. Parents gave written consent for their children to participate and children gave verbal consent. Children were given rest breaks throughout assessment.

Tests with auditory responses were audio recorded and scored from these recordings. All tests were scored by four trained research assistants who were blind to group membership. One primary research assistant scored approximately 60% of all the assessments, while the three others scored the remaining 40%. For the definition production and listening comprehension assessments (see below), the primary research assistant double-scored one randomly selected test from each other rater. This amounted to 9% of the total data from each assessment. Percentage agreement between the primary research assistant and each other research assistant was then calculated, and these figures were averaged to constitute the percentage of inter-rater agreement for these tests.

A description of all assessments is provided below. Further details of the development of all experimental tests, including information on test-retest reliability and internal consistency, are reported in Appendices 1 to 6 in Colenbrander (2015).

**Lexical semantics.**

We measured children's knowledge of lexical semantics (word meanings) in three different ways, requiring different levels of semantic processing depth (Ouelette, 2006) and different levels of expressive language skill.

***Definition Production (Colenbrander, 2015).***

This task assessed children's ability to give verbal definitions for 61 words (see Appendix A): a word was said aloud and the child was asked to tell the tester anything they knew about the word's meaning, for example "Prevent. Do you know what 'prevent' means?". A maximum of three points was awarded for each definition, with full marks being awarded for a well expressed, semantically correct definition. Further details of the scoring criteria are outlined in Colenbrander, Kohnen, Smith-Lock & Nickels (2015). This task tapped deep knowledge of word meanings and the ability to express this knowledge orally. Inter-rater agreement was 89.62%.

***Definition Recognition (Colenbrander, 2015).***

After being asked to provide a verbal definition for each target word as required by the previous task, the child was asked to say which of three orally presented definitions best matched the target word. One distractor was closely semantically related to the target definition, while the other was semantically unrelated. Both distractors referred to the same part of speech as the target definition. For example, for the target word "enthusiasm", the multiple choice options were the feeling of being excited about something, the feeling of being very happy and the feeling of being alone. The multiple choice question was administered whether or not the child had been able to produce a correct definition. The task assessed semantic knowledge of the same words as the definition task, but without the expressive language demands, and less depth of semantic knowledge was required for a correct response.

***Word-Picture matching: Peabody Picture Vocabulary Test Fourth Edition (PPVT-IV; Dunn & Dunn, 2007).***

On this test, children heard a word and saw an array of four pictures. They were required to point to the picture that represented the word. This standardized task is widely used in the reading comprehension literature, and therefore allows comparison with previous research. Like the word-definition matching task, this task required relatively shallow semantic processing and did not depend on expressive language skill. However, it also required less receptive language skill given that the choice was between pictures rather than verbal definitions. This task provided a measure of the breadth of a child's word knowledge (in other words, the size of their receptive vocabulary; Ouelette, 2006) as it tested a range of words of increasing difficulty.

**Conceptual Semantics.**

***Picture-picture association (Colenbrander, 2015).***

To determine whether or not children's difficulties with lexical semantics extended to non-linguistic concept knowledge (conceptual semantics), we used a picture-picture association task. Three pictures were displayed on a computer screen, one target picture at the top of the screen (such as a raincloud) and two stimulus pictures on the bottom of the screen (such as an umbrella and a broom). The child's task was to select which of the two stimulus pictures 'went with' (were associated with) the target picture by pressing an appropriate key. Pictures were presented on a laptop computer using E-Prime 2.0 software (Psychology Software Tools, 2012). Both accuracy and reaction time were recorded. Some of the items were taken from the Squirrel Nut Test (Pitchford, Funnell, Ellis, Green, & Chapman, 1997) while others were taken from a semantic association test developed by Biran & Friedmann, (2007). Some items were associated due to function or collocation (cow - milk, hair - comb), while others were categorically related (rollerskate - boot). There were 38 items.

**Word retrieval.**

***Picture naming: Assessment of Comprehension and Expression 6-11 Naming test (Adams, Cooke, Crutchley, Hesketh, & Reeves, 2001).***

Children were shown a picture and were asked to name the item they saw in the picture. There were 25 items. This task assessed a child's ability to recognize a picture, activate the appropriate semantics, retrieve the phonological form of the item's name and articulate it.

**Phonology.**

***Lexical Phonology: Auditory lexical decision (Colenbrander, 2014).***

A series of words and nonwords (the same as those used in the verbal definition production and word-definition matching tasks discussed above) were read aloud to each child. Nonwords were formed by changing one phoneme of these words (for example, the nonword counterpart for "impair" was "impore"). The child's task was to state whether the stimulus was a word or not by saying "word" or "not a word". The auditory lexical decision task was always administered before the definition and multiple choice tasks. This task was designed to assess children's knowledge of the phonological form of the experimental words, though some children may also have accessed semantics in order to complete the task (see Betjemann & Keenan, 2008).

***Phonological short-term memory: Automated Working Memory Assessment (AWMA) Nonword Recall (Alloway, 2007).***

On this task, children heard a sequence of nonwords and had to recall the nonwords in the correct order. Normally, this task is presented by computer. However, given that the Nonword Recall items were presented by a speaker with a UK accent, we felt that some of the items might be confusing for Australian children. Therefore, the Nonword Recall items were presented by the tester. Apart from this, assessment followed the same procedures as the computerized version of the test. This task was intended to assess children's broader phonological skills independent of the influence of semantics.

**Orthography.**

***Irregular word reading: CC2 Irregular Words (Castles et al., 2009).***

The CC2 Irregular Word reading subtest administered during screening also functioned as an assessment of children's orthographic skills, because irregular words cannot be accurately sounded out and can only be read correctly by accessing the word's orthographic form.

***Ability to read experimental words: Vocabulary reading task (Colenbrander, 2015).***

Of course, knowledge of a word's phonological form is required for reading aloud, and children may also access semantic knowledge while reading regular and irregular words (Keenan & Betjemann, 2007; Nation & Snowling, 1998; Ricketts et al., 2007). Therefore, we assessed children's ability to read the experimental words used in the definition, multiple choice and auditory lexical decision tasks. This allowed us to contrast orthographic, phonological and semantic knowledge for the same words.

**Syntax.**

***Sentence-Picture matching (Colenbrander, 2015).***

The syntax task was adapted from a similar task designed for use with Hebrew-speaking children (Friedmann & Novogrodsky, 2002). Pictures depicting three characters (for example, two women and a girl) were shown on a computer screen using Microsoft Powerpoint. Children then heard a sentence relating to the picture. The child's task was to point to the correct referent for each sentence. Ten of the sentences were subject wh-questions ("Which lady is pinching the girl?"), ten were object wh- questions ("Which lady is the girl pinching?"), ten were subject relatives ("Point to the lady that is pinching the girl") and ten were object relatives ("Point to the lady that the girl is pinching"). Sentences types were presented intermixed. The object sentences (both the questions and relative clauses) were of particular interest as they have been shown to be particularly difficult for children with semantic impairments (Friedmann & Levy, 2009; Friedmann & Novogrodsky, 2004, 2011; van der Lely, Jones, & Marshall, 2011). However, subject sentences of similar length and semantic content were included so that low scores on the object questions could be attributed to syntactic difficulties rather than processing factors unrelated to syntax.

**Broader Oral Language Skills.**

***Listening comprehension (Colenbrander, 2015).***

Listening comprehension was assessed as a broad measure of general oral language comprehension skills including syntax, semantics, inferential abilities etc. Passages 4, 5 and 6 of Form 2 of the NARA (Neale, 1999) were read aloud to participants. After hearing the passages, children were asked 8 open-ended questions about each passage (24 questions in total). Inter-rater agreement was 98.61%.

**Memory and Reasoning.**

***Spatial working memory:  AWMA Spatial Recall Subtest (Alloway, 2007).***

Evidence has shown that verbal working memory skills are linked to reading comprehension (e.g., Cain, 2006; Nation, Adams, Bowyer-Crane, & Snowling, 1999). However, there is debate as to whether poor comprehenders perform poorly on verbal working memory tasks due to the language demands of the task, or due to underlying working memory difficulties (Nation et al., 1999; Pimperton & Nation, 2010; Van Dyke, Johns, & Kukona, 2014). We therefore assessed non-verbal working memory in order to explore the prevalence of such difficulties within our sample.

The child saw a series of two shapes on the screen of a laptop computer. In each pair of shapes, the shape on the right had a red dot above it. The child was required to say whether or not the shape on the right was the same as the shape on the left, and then recall the spatial location of the red dot for each pair, in the order that they appeared. The number of shape pairs increased with each trial. This task returned a Processing score (number of correct similarity judgements) and a Recall score (whether or not the child was able to remember the location of the red dots in order).

***Nonverbal IQ: Matrix Reasoning subtest of Wechsler Abbreviated Scale of Intelligence Second Edition (WASI-II; Wechsler, 2011).***

We wished to rule out general reasoning or intellectual deficits as a possible reason for reading comprehension difficulties. Therefore, we assessed nonverbal reasoning abilities. On this test, children were asked to identify which of 5 pictures represented the next step in a visual matrix.

## Analyses

We analyzed results at both the group and individual levels in order to determine what trends were evident at the group level, and whether these held for individual participants.

### Group-level comparisons.

Our sample contained participants who were different ages and in different school grades. Therefore, in order to combine the data, we regressed each child's raw score for each measure on their age and age squared. The resulting standardized residuals were a measure of each child's performance on a particular task relative to other children, once the influence of age had been removed (Hua & Keenan, 2014). These standardized residuals were transformed into standard scores with a mean of 100 and a standard deviation of 15. These standard scores were used in all subsequent group-level analyses and will henceforth be referred to as "sample standard scores" to distinguish them from standard scores obtained from standardized tests.

The conceptual semantics test returned both accuracy and reaction time data. We analyzed these separately. Analyses of reaction time data were carried out using each participant's mean reaction time (RT). Only RTs from correct trials were included. RTs more than three standard deviations from each participant's mean were excluded, resulting in a loss of 2.6% of the data.

For many of the measures, data did not meet assumptions of normality or equality of variance. Therefore, Mann-Whitney U tests were carried out to compare the scores of poor comprehenders to those of controls for each measure.

**Individual-level comparisons.**

The oral language difficulties of poor comprehenders may be subtle and difficult to detect (Catts, 2009; Nation et al., 2004). Poor comprehenders may score in the low end of the average range on a standardized task and it may therefore appear that they do not have a deficit on that task – but in fact, the poor comprehender's score may be significantly lower than the standard scores obtained on the task by children who have average reading comprehension skills. For this reason, we compared the scores of individual poor comprehenders to the scores of average comprehenders carefully matched for reading accuracy, using statistical methods specifically developed for the purpose. Poor comprehenders' scores were compared to those of a group of matched controls on each task using a modified t-test procedure implemented in a computer programme, SinglimsES.exe (Crawford, Garthwaite, & Porter, 2010; Crawford & Howell, 1998), which is freely available online (http://homepages.abdn.ac.uk/j.crawford/pages/dept/SingleCaseMethodsComputerPrograms.HTM).

The SinglimsES test is designed to be used with control samples of less than 50 participants and is accurate for samples as small as 5 control participants (Crawford & Howell, 1998). This test calculates how unusual a particular case's score is likely to be within a relevant control population, extrapolated from the test scores of the control sample. This is expressed as the percentage of the estimated control population whose scores would be expected to fall below the given case's score.

When 10% of the control population's scores were estimated to fall below a poor comprehender's score, we considered the child to be impaired on that particular skill. This is equivalent to approximately 1.3 standard deviations below the mean. In other words, we considered a child to have a deficit on a particular skill when 90% of the control population would be expected to obtain a score higher than that of the poor comprehender.

As well as reporting the percentage of control participants falling below a case's score, SinglimsES reports $p$ values for the difference between the case of interest and the control sample. In our sample, when a poor comprehender's score falls below that of 5% of the control population, this is equivalent to $p < 0.05$, and when their score falls below that of 10% of the population, $p$ is between 0.05 and 0.10. This means that our choice of the 10% cut-off entails acceptance of an alpha level of 0.10. However, given the small size of our control sample (6 to 7 participants per grade), the level of power in our study is low (Crawford & Howell, 1998). Therefore, the higher risk of Type I errors associated with the adoption of a higher alpha level is mitigated by the study's low power. Conversely, one of the overarching aims of this study was to be able to identify subtle, difficult-to-detect oral language difficulties, and adoption of an alpha level of 0.05 in conjunction with our small sample size is likely to lead to an unacceptably high risk of Type II errors. For these reasons, we believe an alpha level of 0.10 is warranted.

Because the scores of poor comprehenders were compared to those of grade- and age-matched controls, all individual-level comparisons were calculated using raw scores (or mean RT in the case of conceptual semantics reaction time data). The only exceptions to this were reading comprehension, word reading accuracy and fluency – these tasks were assessed at screening, which took place 6 months before the other assessments for children from the Catholic school sample, but only one to two weeks before for children from the Neuronauts sample. These comparisons were therefore calculated using standard scores.

# Results

## Group Level Results

Results for poor comprehenders and controls at screening are displayed in Table 1 above. Mean sample standard scores (created from raw scores regressed on age and age squared) and standard deviations for both groups on all measures are shown in Table 2.

**Table 2**

***Mean Sample Standard Scores of Poor Comprehenders and Controls on Oral Language and Cognitive Tasks***

| Measure | Poor Comprehenders | Controls | Mann-Whitney U |
|---|---|---|---|
| Screening measures | | | |
| NARA Comprehension[a] | 89.00 (9.66) | 107.10 (12.74) | <0.001** |
| YARC Comprehension[a] | 89.54 (12.51) | 106.80 (11.53) | <0.001** |
| CC2 Nonword Reading[a] | 103.69 (12.61) | 97.6 (15.39) | 0.22 |
| CC2 Irregular Word Reading[a] | 102.23 (16.70) | 98.6 (13.06) | 0.50 |
| Language, cognitve and fluency measures | | | |
| TOWRE Sightword Efficiency[a] | 100.08 (15.79) | 100.00 (14.23) | 0.81 |
| TOWRE Pseudoword Decoding[a] | 101.08 (17.54) | 99.35 (12.55) | 0.81 |
| Vocabulary reading | 96.08 (3.95) | 102.55 (14.55) | 0.25 |
| PPVT-IV | 92.38 (12.45) | 105.05 (13.70) | 0.02* |
| Definition production task | 90.46 (7.40) | 106.05 (14.72) | 0.002** |
| Definition recognition task | 91.62 (11.24) | 105.40 (14.02) | 0.01* |
| Conceptual semantics task accuracy | 93.46 (14.22) | 104.25 (13.48) | 0.03* |
| Conceptual semantics task reaction time | 104.54 (9.60) | 97.10 (16.66) | 0.12 |
| ACE 6-11 Naming | 98.23 (16.39) | 101.37 (13.33) | 0.82 |
| Auditory lexical decision task | 93.00 (13.13) | 104.40 (13.93) | 0.02* |
| AWMA Nonword Recall | 98.92 (15.20) | 95.58 (27.05) | 0.91 |
| Syntax task - object sentences[b] | 92.38 (17.84) | 104.95 (9.67) | 0.02* |
| Listening comprehension task | 90.46 (12.17) | 106.35 (12.63) | <0.001** |
| AWMA Spatial Recall Processing | 99.15 (14.14) | 100.56 (15.16) | 0.92 |
| AWMA Spatial Recall | 95.69 (7.89) | 103.22 (17.39) | 0.28 |
| WASI-II Matrix Reasoning | 94.69 (10.11) | 103.5 (16.02) | 0.09 |

*Note*. Standard scores shown here were derived from the experimental sample and not from standardized tests. Age differences were controlled for by regressing raw scores on age and age squared. Standard deviations are in parentheses. [a] Assessed at screening. [b] All participants were at ceiling on subject questions, therefore these were not analyzed. *$p < 0.05$ **$p<0.01$

**Semantics and word retrieval.**

There was no difference between the groups on the reaction time measure of the conceptual semantics task, nor on the naming task. However, there were significant differences on the PPVT-IV, definition production and definition recognition tasks, and on the accuracy measure of the conceptual semantics task. Thus, as a group, the semantic knowledge of the poor comprehenders was poorer than that of the controls.

**Phonology.**

There was no significant difference between the groups on the nonword recall task, indicating that the poor comprehenders' general phonological skills were not different from those of controls. However, there was a significant difference between the groups on the auditory lexical decision task, indicating that as a group, the poor comprehenders' knowledge of lexical phonology was weaker than that of the controls.

**Orthography.**

There was no difference between the groups on CC2 Irregular Word Reading or on the vocabulary reading task, indicating that the orthographic skills of the poor comprehenders were as good as those of the controls despite their poor reading comprehension.

**Broader Oral Language.**

There were significant differences between the groups on the tasks of syntax and listening comprehension. Thus, the poor comprehenders seemed to have difficulties with oral language skills beyond the word level.

**Memory and Reasoning Abilities.**

Mean spatial recall scores were lower for poor comprehenders than for controls, but the difference was not significant. The difference between the groups on the Matrix Reasoning task approached, but did not reach, significance. Thus, although the mean scores for poor comprehenders were lower than that of controls, there was no clear evidence that the memory and reasoning abilities of poor comprehenders were worse than those of controls.

### Summary.

At the group level, the poor comprehenders appeared to have difficulties with the semantic and phonological aspects of word knowledge, as well as difficulties with conceptual semantic knowledge, syntax and broader oral language skills. However, the poor comprehenders had relative strengths in orthographic skills and non-verbal cognitive abilities.

## Individual-level results

The number of control participants at each grade level is displayed in Table 3, along with the mean age of controls. Table 4 presents the estimated percentage of the population of children from which the control children are drawn who would score worse than a poor comprehender for each measure (as calculated using the Singlims statistics). A child whose score falls below the bottom 10% of the control population is considered to have difficulties

**Table 3**

***Mean Age of Control Participants***

| | Grade 4 | Grade 5 | Grade 6 |
|---|---|---|---|
| Number of participants | 7 | 6 | 6 |
| Mean age (years:months) | 9:7 | 10:5 | 11:4 |
| Standard deviation (months) | 2.37 | 2.00 | 3.00 |

Note. One control participant in Grade 5 and one in Grade 6 were not tested on the AWMA due to equipment failure. In addition, one control child in Grade 5 was not tested on the ACE 6-11 due to testing interruptions. Thus, there is one less control participant for each of these comparisons. However, there were never less than 5 controls on any one measure (5 is the minimum number of controls required for reliability of the statistical analysis; Crawford & Howell, 1998).

# Table 4

## *Singlims Results*

| Measure | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 | PC7 | PC8 | PC9 | PC10 | PC11 | PC12 | PC13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Grade | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 6 | 6 |
| Age (years:months) | 9:1 | 9:6 | 9:0 | 8:11 | 9:5 | 9:9 | 9:8 | 10:0 | 9:10 | 10:2 | 10:4 | 11:5 | 11:3 |
| Screening measures | | | | | | | | | | | | | |
| NARA Comprehension[a] | 1.36* | 1.12* | 0.92* | 2.52* | 0.26* | 4.76* | 3.11* | 2.07* | 3.54* | 0.54* | 3.54* | 2.62* | 34.39 |
| YARC Comprehension[a] | 0.81* | 2.62* | 18.36 | 9.05^ | 0.35* | 0.16* | 0.52* | 91.16 | 28.15 | 28.15 | 0.71* | 35.55 | 0.71* |
| CC2 Nonword Reading[a] | 60.76 | 77.28 | 74.07 | 81.55 | 56.52 | 23.76 | 46.36 | 91.00 | 77.06 | 79.42 | 23.54 | 88.41 | 63.71 |
| CC2 Irregular Word Reading[a] | 2.95* | 22.94 | 49.68 | 39.07 | 90.32 | 80.62 | 41.38 | 99.36 | 99.87 | 97.86 | 93.91 | 50.00 | 9.64^ |
| Language, cognitive and fluency measures | | | | | | | | | | | | | |
| TOWRE Sightword Efficiency[a] | 60.76 | 37.59 | 81.53 | 77.97 | 6.30* | 81.53 | 96.12 | 94.56 | 54.79 | 75.65 | 80.25 | 39.99 | 30.02 |
| TOWRE Pseudoword Decoding[a] | 64.56 | 44.97 | 13.67 | 62.30 | 11.16 | 13.67 | 44.97 | 99.95 | 84.13 | 84.13 | 39.16 | 70.23 | 86.63 |
| Vocabulary reading task | 13.50 | 19.06 | 55.15 | 22.45 | 16.08 | 11.29 | 60.19 | 57.31 | 5.82^ | 22.16 | 0.20* | 75.73 | 58.30 |
| PPVT-IV | 3.92* | 7.77^ | 12.15 | 0.94* | 6.82^ | 50.91 | 60.23 | 53.98 | 2.33* | 4.12* | 1.53* | 10.54 | 45.82 |
| Definition production task | 8.04^ | 7.62^ | 9.93^ | 6.85^ | 7.22^ | 12.23 | 25.25 | 26.27 | 14.40 | 12.66 | 9.72^ | 7.28^ | 12.37 |
| Definition recognition task | 7.49^ | 6.31^ | 6.38* | 1.62* | 1.91* | 20.00 | 39.28 | 23.83 | 14.33 | 23.83 | 18.61 | 0.81* | 21.47 |
| Conceptual semantics task accuracy | 2.45* | 0.51* | 50.00 | 50.00 | 13.22 | 86.78 | 13.22 | 98.87 | 63.94 | 98.87 | 3.06* | 8.86^ | 8.86^ |
| Conceptual semantics task reaction times | 68.60 | 6.18^ | 9.07^ | 0.88* | 28.24 | 82.50 | 19.74 | 56.07 | 20.16 | 14.85 | 40.40 | 39.19 | 11.90 |
| ACE 6-11 Naming | 14.56 | 14.56 | 1.60* | 0.80* | 0.13* | 47.11 | 91.35 | 77.47 | 43.74 | 14.63 | 15.27 | 31.63 | 82.15 |
| Auditory lexical decision task | 17.24 | 7.02^ | 9.55^ | 9.55^ | 5.15^ | 32.94 | 49.40 | 33.81 | 14.48 | 14.48 | 14.48 | 12.31 | 0.01* |
| AWMA Nonword Recall | 81.43 | 72.91 | 18.57 | 18.57 | 81.43 | 72.91 | 72.91 | 15.37 | 50.00 | 50.00 | 0.31* | 20.46 | 63.75 |
| Syntax task - object sentences[b] | 18.89 | 0.26* | 45.54 | 45.54 | 45.54 | 18.89 | 6.21^ | 38.42 | 0.19* | 38.42 | 10.97 | 0.46* | 77.71 |
| Listening comprehension task | 9.69^ | 3.85* | 7.14^ | 9.69^ | 5.24^ | 9.69^ | 76.31 | 7.71^ | 15.44 | 1.32* | 1.86* | 10.85 | 31.27 |
| AWMA Spatial Recall Processing | 27.81 | 41.17 | 35.13 | 43.25 | 3.64* | 13.32 | 4.19* | 44.39 | 17.89 | 39.25 | 48.57 | 22.79 | 93.01 |
| AWMA Spatial Recall | 26.83 | 29.65 | 31.13 | 29.65 | 17.44 | 25.49 | 19.50 | 45.66 | 9.90^ | 33.34 | 39.33 | 25.70 | 86.70 |
| WASI-II Matrix Reasoning | 7.85^ | 18.07 | 29.52 | 5.88^ | 23.30 | 23.30 | 13.81 | 61.11 | 30.55 | 74.81 | 45.50 | 14.61 | 22.39 |

*Note.* Shaded areas represent scores which fall below less than 10% of the estimated control population's scores. [a]Assessed at screening [b]Subject sentences not reported as participants were at ceiling ^$p < 0.10$ *$p < 0.05$

with that skill (see Method section). Mean raw scores of controls and raw scores of individual poor comprehenders are shown in Appendices B and C.

**Semantics and Word Retrieval.**

Ten of the 13 poor comprehenders (77%) scored below the control population on at least one lexical or conceptual semantics task, and of these 10 children, 7 scored below the cut-off on multiple semantic tasks. Thus, the majority of children in our sample appeared to have some level of semantic difficulty. For approximately half of the poor comprehenders, semantic difficulties extended to inaccurate or slow performance on the task of conceptual semantics. Three of the children (23%) who had difficulties with semantics also scored below controls on the word retrieval task (ACE II Naming). However, three children (23%) scored as well as controls on all the tasks tapping semantics, suggesting that there are poor comprehenders who do not have semantic difficulties.

**Phonology.**

Only one participant (PC11) scored below controls on the nonword recall task, indicating that for the majority of poor comprehenders, broader phonological skills are a relative strength. However, 5 poor comprehenders (38%) scored below the cut-off on the auditory lexical decision task. This suggests that these children have difficulties with learning, storing or retrieving lexical representations which extend beyond semantics.

**Orthography and Reading Skills.**

Only two participants scored below controls on the vocabulary reading task. Orthographic knowledge therefore appears to be a relative strength for the majority of poor comprehenders. This is especially true considering that 6 participants who scored below controls on the definition production and recognition tasks, and 5 who scored below controls on the auditory lexical decision task, scored above the cut-off when asked to read the same words aloud. Note that the experimental words were a mixture of regularly and irregularly spelled words, so participants could sound out the regular words even if they had not encountered them before.

Turning to the CC2 Irregular Word Reading results, only two children scored below controls (both of these children were within the normal range using the test's norms). These children also scored below cut-off on the conceptual semantics assessment, so it is plausible that their irregular word reading weaknesses might be caused by semantic weaknesses. Three participants had intact semantic skills and intact irregular word reading skills. However, interestingly, there were 8 participants who scored poorly on one or more semantics assessments, but whose irregular word reading skills were as good as controls. This implies that poor semantics does not necessarily lead to poor irregular word reading (see Blazely et al., 2005; Nation & Cocksey, 2009).

In contrast to irregular word reading, not one participant scored below controls on CC2 Nonword Reading. Thus, decoding accuracy is a clear strength for poor comprehenders. However, one participant scored below these age matched controls the TOWRE Sight Word Efficiency subtest, despite the fact that her score fell within the normal range using the test's norms.

These results demonstrate that for a minority of poor comprehenders, subtle deficits in irregular word reading and reading fluency may be present, and these may play a role in the child's reading comprehension difficulties. However for the majority, poor word reading is not a contributing factor.

**Broader Oral Language.**

Nine poor comprehenders (69%) scored below controls on the listening comprehension task and four scored below controls on the syntax task. Taken together, all but one of the poor comprehenders (92%) demonstrated an oral language difficulty above the word level.

For the majority of the children, impaired semantic and/or lexical processing could be the source of their problem on these tasks. However, there were three participants who did not display any semantic or lexical difficulties. Consequently, we were interested in whether or not they would demonstrate oral language difficulties beyond the word level. In fact, participants PC6 and PC8 scored below controls on the listening comprehension task, while PC7 scored below controls on the syntax task. Thus, each of these participants demonstrated a subtle deficit on a sentence- or discourse-level oral language task in the context of intact lexical-level abilities.

#### Memory and Reasoning Abilities.

Two participants scored below controls on the AWMA Spatial Recall Processing measure, one participant scored below controls on the Spatial Recall measure, and two further participants scored below controls on the Matrix Reasoning measure. Therefore, a minority of poor comprehenders do appear to have subtle deficits in broader cognitive abilities.

#### Summary.

Compared to controls who had average reading comprehension, the majority of poor comprehenders displayed deficits on tasks of lexical or conceptual semantics, while a small proportion also demonstrated deficits on tasks tapping orthography and lexical phonology. Thus, the majority of poor comprehenders appeared to have difficulties at the lexical level, primarily in the semantic domain. By contrast, they had relative strengths in word reading, broader phonological skills, spatial working memory and non-verbal reasoning skills.

Three participants displayed lexical-level skills on par with controls. However, all three of these participants had oral language deficits at the sentence or discourse level. In total, 12 of the 13 poor comprehenders scored below controls on either the syntax task, the listening comprehension task or both. In fact, all 13 poor comprehenders scored below controls on at least one task of oral language, whether at the lexical, sentence, or discourse level.

Overall, the results of the individual comparisons show that while the pattern of impairments across poor comprehenders is heterogeneous, consistent with previous studies (Cain & Oakhill, 2006; Cornoldi, De Beni, & Pazzaglia, 1996; Nation et al., 2004), all poor comprehenders had some degree of oral language deficit.

## Discussion

Despite being the subject of much research, the role of vocabulary within reading comprehension is controversial. While the results of some studies have pointed to a central role for vocabulary and semantic deficits in all children with specific reading comprehension difficulties (e.g. Nation & Snowling 1998), other findings support the view that oral vocabulary difficulties play a role in the reading comprehension deficits of some, but not all, poor comprehenders (Cain & Oakhill, 2006; Clarke, Snowling et al., 2010). In the current study, we wished to build on this work by exploring the vocabulary skills of poor comprehenders in detail. We assessed semantic, orthographic and phonological aspects of word knowledge in order to tease apart the contributions of the different aspects of word knowledge. Furthermore, we assessed semantic skills on a variety of different tasks in order to explore whether certain tasks would be more sensitive to the vocabulary deficits of poor comprehenders than others, and assessed irregular word reading skills so that we could explore associations between irregular word reading and semantics. Finally, we analyzed data at both the group and individual level, so that we could explore individual variation within the sample of poor comprehenders.

At the group level, poor comprehenders scored more poorly than controls on all semantic tasks – these included tasks of word-picture matching, definition recognition and production (lexical semantics), and picture-picture matching (conceptual semantics). However, as a group, poor comprehenders' difficulties with vocabulary were not restricted to semantics – they also scored more poorly than controls on a task tapping knowledge of phonological word form. Although we cannot rule out the fact that this task may reflect semantic skills to some degree (e.g., see Betjemann & Keenan, 2008), this finding raises the possibility that poor comprehenders may have weaker lexical phonology. Thus, at the group level, poor comprehenders did appear to have vocabulary deficits, apparent in both syntactic and phonological aspects of word knowledge. These results are consistent with the Lexical Quality Hypothesis (Perfetti, 2007; Perfetti & Hart, 2002).

Although poor comprehenders, as a group, had weaknesses in the phonological aspect of word knowledge, their general phonological abilities (as measured by nonword recall) were as good as those of controls. They also demonstrated relative strengths in orthographic knowledge, word reading abilities, spatial working memory and non-verbal reasoning. However, they scored more poorly than controls on tasks of syntax and listening comprehension. Together, these results demonstrate that poor comprehenders seem to have oral language difficulties at the lexical, sentence and discourse level, in the context of broadly age-appropriate word reading and cognitive abilities.

These results are consistent with the Simple View of Reading (Gough & Tunmer, 1986; Hoover & Gough, 1990), a theoretical view which posits that reading comprehension is the product of word reading (usually referred to as "decoding") and oral language skills. Because the word reading skills of poor comprehenders are by definition unimpaired, this model predicts that specific reading comprehension difficulties are due to poor oral language skills. Consequently, our group-level results are compatible with both the Simple View and The Lexical Quality Hypothesis, because poor comprehenders' lexical-level deficits were restricted to the oral language domain (that is, their lexical deficits were semantic and phonological, rather than orthographic).

Deficits in the quality of the semantics of lexical representations appear to be more pervasive than in the phonological aspect of lexical representations. At the individual level, the majority of poor comprehenders (77%) scored below controls on at least one of the semantics tasks, while a third of this subset of poor comprehenders scored below controls on the auditory lexical decision task, and only two scored below controls on the vocabulary reading task.

Although participants who scored poorly on one test of semantics also tended to score poorly on other tests of semantics, this was not always the case. No single assessment of semantics was clearly more sensitive than the others. For example, 7 poor comprehenders scored below controls on the PPVT-IV word-picture matching test, which is often considered a measure of semantic breadth (Tannenbaum et al., 2006), and 7 poor comprehenders (not all the same children) scored below controls on the definition task, a measure of semantic depth with much greater expressive language demands. This was unexpected, as measures of semantic depth have been found to be more closely related to reading comprehension (Ouelette, 2006) and we also expected that the definition task would be more difficult for poor comprehenders given its reliance on expressive language. Given these results, it may be that the source of vocabulary difficulties is not the same for all poor comprehenders. However, there are a variety of reasons why a child could succeed or fail on a task of vocabulary. Research with larger samples should explore the interactions between vocabulary depth, task demands and factors such as child motivation and behavior during testing.

Nonetheless, it is worth noting that of the 7 children who scored below controls on the PPVT-IV, not one scored below the average range using the test's norms. This supports previous research suggesting that poor comprehenders' oral language difficulties are subtle and may not always be evident on standardized tests (Catts et al., 2006; Nation et al., 2004). Our findings also concur with those of a recent study by Spaulding, Hosmer, & Schechtman, (2013), in which it was found that the PPVT-IV was not sensitive to vocabulary deficits in a group of children with specific language impairment. Spaulding et al. (2013) stated that this may be because the PPVT-IV norming sample contains a relatively high proportion of children with developmental disorders, with the result that the vocabulary difficulties of children with language impairment appear less extreme than they might if children were compared to a sample containing only typically developing children (Spaulding et al., 2013).

Despite our thorough assessment of vocabulary knowledge, there were three poor comprehenders who scored as well as controls on all the tasks tapping lexical skills. These three children, as well as the vast majority of poor comprehenders in our study (12 out of 13), scored below controls on either the syntax task, the listening comprehension task, or both. This implies that these children have oral language deficits beyond the lexical level. Although our data does not allow us to explore causal hypotheses, it is plausible that these lexical and oral language deficits played a role in causing the reading comprehension deficits of these children. The prevalence of oral language deficits in our sample of poor comprehenders is broadly consistent with the Simple View of Reading.

It is interesting to note that there were four poor comprehenders who did not score below the cut-off on the listening comprehension task. This is surprising given the fact that the Simple View specifically predicts listening comprehension difficulties (Gough & Tunmer, 1986; Hoover & Gough, 1990), and three out of the four participants (Participants 7, 9 and 12) scored in the below average range for their age on the written version of the same task[1]. A pattern of poor reading comprehension coupled with relatively good listening comprehension is usually observed when a child has word reading difficulties in the context of intact oral language skills. This is not the case for the three participants in question, because they scored as well as, or better than, the majority of control participants on all measures of reading accuracy and fluency, but below controls on at least one oral language measure.

Rather, it appears that the listening comprehension task was very challenging for both poor comprehenders and controls. Children heard the three most difficult passages of the NARA Form 2, and accuracy scores were very low – the maximum raw score achieved by the oldest (Year 6) control children was 15, and the mean was 10 out of a possible 24 marks

[1] The fourth participant scored as well as controls on both the written and oral versions of the NARA. She was included in the study because of her poor YARC comprehension score.

(42%). Therefore, it is likely that floor effects were operating, masking the extent of the difference between the poor comprehenders and controls. Future research should ensure that listening comprehension tasks are extensively piloted to ensure that they are an appropriate level of difficulty.

Our findings with regard to poor comprehenders' irregular word reading skills were also somewhat unexpected. Studies of reading comprehension seldom measure irregular word reading abilities (Ouelette & Beers, 2010). However, the ability to read irregular words is just as crucial for reading comprehension as the ability to sound out regular words (Ouelette & Beers, 2010; Ricketts et al., 2007). In fact, it is arguable whether the term "poor comprehenders" should be applied to children who have irregular word reading difficulties, because they do not in fact have age appropriate reading accuracy.

Weak irregular word reading skills in poor comprehenders have been attributed to semantic difficulties (Nation & Snowling, 1998; Ricketts et al., 2008; Ricketts et al., 2007), because at the group level, poor comprehenders' phonological and orthographic skills are generally found to be intact (Catts et al., 2006; Nation et al., 2004; Ricketts et al., 2008; Stothard & Hulme, 1995). Indeed, two of our poor comprehenders did score below controls on irregular word reading (despite being within normal limits on the standardized test), and both of these children also had semantic deficits. However, crucially, 8 of our participants had intact irregular word reading skills but poor semantic skills. Our study is the first to demonstrate this pattern in poor comprehenders.

This finding implies that semantic deficits do not necessarily cause irregular word reading deficits (Blazely et al., 2005; Nation & Cocksey, 2009). Such a finding is consistent with the Dual Route model, which posits that a word's phonology can be accessed directly from its stored orthographic form, and thus, successful irregular word reading can be achieved independently of semantics (Coltheart et al., 1993; Coltheart et al., 2001). However, our findings can also be explained by the Triangle model (Plaut et al., 1996). Under this model, word reading is achieved via the activation of semantic, orthographic and phonological nodes. Words can be read aloud via different pathways of activation (orthography-phonology-semantics, orthography-phonology and orthography-semantics). The division of labour between these pathways varies depending on a variety of factors including the type of word being read (Plaut et al., 1996; Harm & Seidenberg, 2004). Under this view, children could compensate for semantic weaknesses by relying to a greater degree on the orthography-phonology pathway when reading aloud.

Despite group-level findings that orthography tends to be a relative strength for poor comprehenders (Ricketts et al., 2008), it is possible (under the Dual Route view) that some poor comprehenders could have impairments to the orthographic lexicon (with or without concurrent semantic difficulties). In other words, they might have difficulty accessing representations in the orthographic lexicon, or have poor-quality orthographic representations. Deficits in the orthographic lexicon have been identified in Hebrew speakers using tasks of orthographic lexical decision (Friedmann & Lukov, 2008). A similar task has been administered to English-speaking poor comprehenders (Ricketts et al., 2007), but data was not analyzed at the individual level. Future research with larger samples should examine the prevalence of both semantic and orthographic weaknesses in poor comprehenders at the individual level, using sensitive measures of both skills, and relate these findings to the prevalence of irregular word reading difficulties. It may be that the presence or absence of irregular word difficulties in poor comprehenders is related to the severity of semantic deficits, and/or the presence or absence of co-occuring orthographic deficits.

At the group level, the choice of selection criteria in a study will affect whether or not group differences in irregular word reading abilities are found. If poor comprehenders are matched to controls on a measure of nonword reading alone, it is possible that the poor comprehender sample will contain a number of children who have poor irregular word reading skills. This in turn is likely to result in group differences in irregular word reading between poor comprehenders and controls, and such a selection procedure may also result in a higher number of poor comprehenders with semantic deficits. On the other hand, matching poor comprehenders to controls on both nonword and irregular word reading is likely to reduce the overall number of participants with semantic deficits – but our results have shown that this selection procedure certainly does not eliminate such children altogether.

In addition to irregular word and nonword reading accuracy, we assessed reading fluency. Only one poor comprehender scored below controls on the sight word fluency task. The role of fluency within the Simple View of reading is somewhat controversial, with some studies finding that fluency is important for reading comprehension over and above decoding abilities (Silverman, Speece, Harring, & Ritchey, 2013) and others studies finding that it is not (Adlof, Catts, & Little, 2006). Nonetheless, both fluency and irregular word reading difficulties were rare within our sample, supporting group-level findings that in general, poor comprehenders have intact reading aloud abilities.

It is important to note that, like the oral language weaknesses found in our sample, the fluency and irregular word reading weaknesses of poor comprehenders were very subtle – all poor comprehenders scored within the average range on the CC2 and TOWRE subtests when test norms were used. It was only when their reading and fluency scores were compared to those of average comprehender controls that weaknesses became visible in some of our participants.

In summary, the results for the majority of poor comprehenders were consistent with the predictions of the Lexical Quality Hypothesis. However, three participants scored as well as controls on all measures of lexical skill, whilst scoring below controls on at least one measure of oral language beyond the lexical level. The Lexical Quality Hypothesis states that lexical skills are necessary for successful reading comprehension, but does not claim that lexical skills alone are sufficient for good reading comprehension (Perfetti & Hart, 2002). Therefore, the existence of children with oral language deficits above the word level does not argue against the Lexical Quality Hypothesis as such. All of our participants demonstrated poorer performance than controls on at least one oral language task, whether at the lexical, sentence or discourse level, and the majority had no deficits on nonword or irregular word reading skills. Thus, our findings support the independence of reading abilities and oral language as predicted by the Simple View of Reading, and are also compatible with the Lexical Quality Hypothesis.

Of course, it is crucial to note that our study is correlational. We measured poor comprehenders' skills at a single point in time. Some of the deficits we identified may therefore be incidental correlates of comprehension rather than causal factors. In addition, our results represent profiles of poor comprehenders in the upper primary grades. The picture may well be different at different stages of development. For example, some deficits may play a role in the development of reading comprehension but may no longer be playing a key role by the time the child reaches the upper primary grades (Castles, Kohnen, Nickels, & Brock, 2014; Oakhill & Cain, 2012; Ouelette & Beers, 2010). Training and longitudinal studies are required to explore the causal roles of oral language skills in reading comprehension.

Nonetheless, our study makes several important contributions to the literature. We have replicated previous findings demonstrating that poor comprehenders are a heterogeneous population (Cain & Oakhill, 2006; Cornoldi et al., 1996; Nation et al., 2004), and extended these findings by showing that although the majority of poor comprehenders do have vocabulary deficits, some poor comprehenders do not, even when vocabulary skills are thoroughly assessed using multiple different assessments. We have shown that the majority of poor comprehenders have deficits with the semantic aspects of word knowledge, and that some also have difficulties recognizing the phonological forms of words (though they do not have difficulties with phonology in general). We have also shown that all poor comprehenders in our sample had some level of oral language deficit, though they did not all have the same type of oral language deficit. This emphasizes the importance of considering multiple causal factors for reading comprehension difficulties (Cain & Oakhill, 2006; Pimperton & Nation, 2014).

The heterogeneity of the poor comprehender population emphasizes the importance of a careful examination of which remediation programs work best for individual children (Cain & Oakhill, 2006). At the group level, our results suggested that semantic skills were an area of particular weakness for poor comprehenders, implying that semantic skills should be targeted in programs of reading comprehension remediation. However, when we looked at the individual level, it became clear that three of our participants did not have semantic difficulties, and therefore, a training program targeting semantic skills may not be helpful for these children. Thus, analysis of the predictors of the effectiveness of training programs should be carried out at the individual level as well as at the group level.

Furthermore, our study has shown that language deficits in poor comprehenders are subtle and in many cases will not be detected using standard scores from standardized tests (Catts et al., 2006; Nation et al., 2004). This implies that studies of individual differences in poor comprehenders should compare poor comprehenders to samples of matched controls with average reading comprehension, rather than to standardized test norms.

Finally, our study demonstrates that the use of different vocabulary assessments will result in different outcomes. We used multiple measures of semantic skill which varied in the level of semantic depth required and in their expressive language demands. We found that no one single test diagnosed semantic difficulties across all poor comprehenders, and no test was clearly more sensitive than another. This highlights the importance of using multiple measures of semantic skills (or indeed, any oral language skill) wherever possible when attempting to rule out whether or not children have difficulties on a particular skill.

In conclusion, this study has highlighted some crucial theoretical and methodological points in relation to individual differences in oral language skill across poor comprehenders.

Future studies should explore the causal roles of oral language difficulties through longitudinal and training studies, analyzing results at both the group and individual levels.

**Acknowledgements:**

We wish to thank the staff, students and parents at the participating school and Marielle Smith, Cornelia van Werkum, Kate Ross and Elissa Greco for their assistance with test scoring.

## References


Adams, C., Cooke, R., Crutchley, A., Hesketh, A., & Reeves, D. (2001). *Assessment of Comprehension and Expression 6-11*. Windsor, UK: nferNelson.

Adlof, S. M., Catts, H. W., & Little, T. D. (2006). Should the Simple View of Reading include a fluency component? *Reading and Writing, 19*, 933-958. doi:10.1007/s11145-006-9024-z

Alloway, T. P. (2007). *Automated Working Memory Assessment*. London, UK: Pearson.

Betjemann, R. S., & Keenan, J. M. (2008). Phonological and semantic priming in children with reading disability. *Child Development, 79*, 1086-1102. doi:10.1111/j.1467-8624.2008.01177.x

Biran, M., & Friedmann, N. (2007). *'Shmita vehar Sinai': picture and word association test.* Tel Aviv, Israel: Tel Aviv University.

Blazely, A. M., Coltheart, M., & Casey, B. J. (2005). Semantic impairment with and without surface dyslexia: Implications for models of reading. *Cognitive Neuropsychology, 22*, 695-717. doi:10.1080/02643290442000257

Cain, K. (2006). Individual differences in children's memory and reading comprehension: an investigation of semantic and inhibitory deficits. *Memory, 14*, 553-569. doi:10.1080/09658210600624481

Cain, K., & Oakhill, J. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing, 11*, 489-503. doi:10.1023/A:1008084120205

Cain, K., & Oakhill, J. (2006). Profiles of children with specific reading comprehension difficulties. *British Journal of Educational Psychology, 76*, 683-696. doi:10.1348/000709905X67610

Cain, K., & Oakhill, J. (in press). Reading comprehension and vocabulary: Is vocabulary more important for some aspects of comprehension? *L'Annee Psychologique*.

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

Castles, A., Kohnen, S., Nickels, L., & Brock, J. (2014). Developmental disorders: what can be learned from cognitive neuropsychology? *Philosophical Transactions of the Royal Society B, 369*, 20130407. doi:10.1098/rstb.2013.0407

Catts, H. W. (2009). The narrow view of reading promotes a broad view of reading comprehension. *Language, Speech and Hearing Services in Schools, 40*, 178-183. doi:10.1044/0161-1461(2008/08-0035)

Catts, H. W., Adlof, S. M., & Weismer, S. E. (2006). Language deficits in poor comprehenders: A case for the Simple View of Reading. *Journal of Language, Speech and Hearing Research, 49*, 278-293. doi:10.1044/1092-4388(2006/023)

Clarke, P. J., Henderson, L. M., & Truelove, E. (2010). The poor comprehender profile: Understanding and supporting individuals who have difficulties extracting meaning from text. In J. Holmes (Ed.), *Advances in Child Development and Behaviour* (Vol. 39, pp. 79-129). New York, Boston: Academic Press.

Clarke, P. J., Snowling, M. J., Truelove, E., & Hulme, C. (2010). Ameliorating children's reading-comprehension difficulties: a randomized controlled trial. *Psychological Science, 21*, 1106-1116. doi: 10.1177/0956797610375449

Colenbrander, D. (2015). *Understanding the role of oral vocabulary in reading comprehension difficulties.* (Doctoral dissertation, Macquarie University, Sydney, Australia).

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed processing approaches. *Psychological Review*, *100*, 589-608.  doi:10.1037/0033-295X.100.4.589

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

Cornoldi, C., De Beni, R., & Pazzaglia, F. (1996). Profiles of reading comprehension difficulties: An analysis of single cases. In C. Cornoldi & J. Oakhill (Eds.), *Reading Comprehension Difficulties: Processes and Intervention* (pp. 113-136). Mahwah, NJ: Lawrence Erlbaum Associates.

Crawford, J. R., Garthwaite, P. H., & Porter, S. (2010). Point and interval estimates of effect sizes for the case-controls design in neuropsychology: Rationale, methods, implementations, and proposed reporting standards. *Cognitive Neuropsychology, 27*, 245-260. doi:10.1080/02643294.2010.513967

Crawford, J. R., & Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *The Clinical Neuropsychologist (Neuropsychology, Development and Cognition: Section D), 12*, 482-486. doi:10.1076/clin.12.4.482.7241

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test Fourth Edition*. Minneapolis, MN: Pearson.

Elwer, S., Keenan, J. M., Olson, R. K., Byrne, B., & Samuelsson, S. (2013). Longitudinal stability and predictors of poor oral comprehenders and poor decoders. *Journal of Experimental Child Psychology, 115*, 497-516. doi:10.1016/j.jecp.2012.12.001

Friedmann, N., & Levy, H. (2009). Treatment of syntactic movement in syntactic SLI: A case study. *First Language*, *29*, 15-50. doi:10.1177/0142723708097815

Friedmann, N., & Lukov, L. (2008). Developmental surface dyslexias. *Cortex, 44*, 1146-1160. doi:10.1016/j.cortex.2007.09.005

Friedmann, N., & Novogrodsky, R. (2002). *BAMBI: Battery for assessment of syntactic abilities in children*. Tel Aviv, Israel: Tel Aviv University.

Friedmann, N., & Novogrodsky, R. (2004). The acquisition of relative clause comprehension in Hebrew: A study of SLI and normal development. *Journal of Child Language, 31*, 661-681. doi:10.1017/S0305000904006269

Friedmann, N., & Novogrodsky, R. (2011). Which questions are most difficult to understand?: The comprehension of Wh questions in three subtypes of SLI. *Lingua, 121*, 367-382. doi:10.1016/j.lingua.2010.10.004

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*, 6-10. doi:10.1177/074193258600700104

Harm, M., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. P*sychological Review, 111,* 662–720. doi: 10.1037/0033-295X.111.3.662

Hoover, W. A., & Gough, P. B. (1990). The Simple View of Reading. *Reading and Writing, 2*, 127-160. doi:10.1007/BF00401799

Howard, D., & Nickels, L. (2005). Separating input and output phonology: semantic, phonological, and orthographic effects in short-term memory impairment. *Cognitive Neuropsychology, 22*, 42-77. doi:10.1080/02643290342000582

Hua, A. N., & Keenan, J. M. (2014). The role of text memory in inferencing and in comprehension deficits. *Scientific Studies of Reading, 18,* 1-17. doi:10.1080/10888438.2014.926906

Keenan, J. M., & Betjemann, R. S. (2007). Comprehension of single words: The role of semantics in word identification and reading disability. In E. L. Grigorenko & A. J. Naples (Eds.), *Single-word reading: Behavioural and biological perspectives* (pp. 191-210). New York, NY: Taylor & Francis.

Landi, N., & Perfetti, C. A. (2007). An electrophysiological investigation of semantic and phonological processing in skilled and less-skilled comprehenders. *Brain and Language, 102*, 30-45. doi:10.1016/j.bandl.2006.11.001

Nation, K., Adams, J. W., Bowyer-Crane, C., & Snowling, M. J. (1999). Working memory deficits in poor comprehenders reflect underlying language impairments. *Journal of Experimental Child Psychology, 73*, 139-158. doi:10.1006/jecp.1999.2498

Nation, K., Clarke, P., Marshall, C. M., & Durand, M. (2004). Hidden language impairments in children: Parallels between poor reading comprehension and specific language impairment? *Journal of Speech, Language and Hearing Research, 47*, 199-211. doi:10.1044/1092-4388(2004/017)

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

Nation, K., Cocksey, J., Taylor, J. S., & Bishop, D. V. (2010). A longitudinal investigation of early reading and language skills in children with poor reading comprehension. *Journal of Child Psychology and Psychiatry, 51*, 1031-1039. doi:10.1111/j.1469-7610.2010.02254.x

Nation, K., & Snowling, M. J. (1997). Assessing reading difficulties: the validity and utility of current measures of reading skill. *British Journal of Educational Psychology, 67*, 359-370. doi:10.1111/j.2044-8279.1997.tb01250.x

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

Nation, K., & Snowling, M. J. (1999). Developmental differences in sensitivity to semantic relations among good and poor comprehenders: Evidence from semantic priming. *Cognition, 70*, B1-B13. doi:10.1016/S0010-0277(99)00004-9

Nation, K., Snowling, M. J., & 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. *Advances in Speech-Language Pathology, 9,* 131-139. doi:10.1080/14417040601145166

Neale, M. D. (1999). *The Neale Analysis of Reading Ability Third Edition*. Melbourne, Australia: ACER.

Oakhill, J., & Cain, K. (2012). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading, 19,* 91-121. doi:10.1080/10888438.2010.529219

Ouelette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology, 98,* 554-566. doi:10.1037/0022-0663.98.3.554

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

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11,* 357-383. doi:10.1080/10888430701530730

Perfetti, C., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of funtional literacy* (pp. 189-213). Amsterdam/Philadelphia: John Benjamins.

Pimperton, H., & Nation, K. (2010). Suppressing irrelevant information from working memory: Evidence for domain-specific deficits in poor comprehenders. *Journal of Memory and Language, 62,* 380-391. doi:10.1016/j.jml.2010.02.005

Pimperton, H., & Nation, K. (2014). Poor comprehenders in the classroom: teacher ratings of behavior in children with poor reading comprehension and its relationship with individual differences in working memory. *Journal of Learning Disabilities, 47,* 199-207. doi:10.1177/0022219412454172

Pitchford, N. J., Funnell, E., Ellis, A. W., Green, S. H., & Chapman, S. (1997). Recovery of spoken language processing in a 6-year-old child following a left hemisphere stroke: A longitudinal study. *Aphasiology, 11*, 83-102. doi:10.1080/02687039708248457

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*, 56-115. doi:10.1037/0033-295X.103.1.56

Ricketts, J., Bishop, D. V., & Nation, K. (2008). Investigating orthographic and semantic aspects of word learning in poor comprehenders. *Journal of Research in Reading, 31*, 117-135. doi:10.1111/j.1467-9817.2007.00365.x

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

Silva, M., & Cain, K. (2014). The relations between lower and higher level comprehension skills and their role in prediction of early reading comprehension. Advance online publication. *Journal of Educational Psychology*. doi:10.1037/a0037769

Silverman, R. D., Speece, D. L., Harring, J. R., & Ritchey, K. D. (2013). Fluency has a role in the Simple View of Reading. *Scientific Studies of Reading, 17*, 108-133. doi: 10.1080/10888438.2011.618153

Snowling, M. J., Stothard, S. E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., . . . Hulme, C. (2012). *York Assessment of Reading for Comprehension: Passage Reading. Australian Edition.* London, UK: GL Assessment.

Spaulding, T. J., Hosmer, S., & Schechtman, C. (2013). Investigating the interchangeability and diagnostic utility of the PPVT-III and PPVT-IV for children with and without SLI. *International Journal of Speech Language Pathology, 15*, 453-462. doi:10.3109/17549507.2012.762042

Stothard, S. E., & Hulme, C. (1995). A comparison of phonological skills in children with reading comprehension difficulties and children with decoding difficulties. *Journal of Child Psychology and Psychiatry, 36*, 399-408. doi:10.1111/j.1469-7610.1995.tb01298.x

Tannenbaum, K. R., Torgesen, J. K., & Wagner, R. K. (2006). Relationships between word knowledge and reading comprehension in third-grade children. *Scientific Studies of Reading, 10*, 381-398. doi:10.1207/s1532799xssr1004_3

Tong, X., Deacon, S. H., & Cain, K. (2014). Morphological and syntactic awareness in poor comprehenders: another piece of the puzzle. *Journal of Learning Disabilities, 47*, 22-33. doi:10.1177/0022219413509971

Psychology Software Tools. (2012). E-Prime 2.0. Pittsburgh, PA: Psychology Software Tools.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of Word Reading Efficiency*. Austin, TX: Pro-Ed.

van der Lely, H. K., Jones, M., & Marshall, C. R. (2011). Who did Buzz see someone? Grammaticality judgement of wh-questions in typically developing children and children with Grammatical-SLI. *Lingua, 121*, 408-422. doi:10.1016/j.lingua.2010.10.007

Van Dyke, J. A., Johns, C. L., & Kukona, A. (2014). Low working memory capacity is only spuriously related to poor reading comprehension. *Cognition, 131*, 373-403. doi:10.1016/j.cognition.2014.01.007

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence Second Edition*. San Antonio, TX: NCS Pearson.

Yuill, N., & Oakhill, J. (1991). *Children's problems in text comprehension: An experimental investigation*. Cambridge, UK: Cambridge.

## Appendix A

Words used in Definition, Multiple Choice, Auditory Lexical Decision

and Vocabulary Reading Tasks

abrupt
abundance
ally
assist
amateur
benefactor
brute
coerce
conceal
confide
dubious
dispute
enthusiasm
envy
foe
frantic
gradual
idle
lunge
mock
naïve
nimble
offend
petty
ploy
precise
prevent
request
scarce
scorn
seize
serene
shrewd
shudder
swindle
temptation
tense
triumph
tyrant
urge
vague

anxious
conform
conquer
expedition
fatigue
fury
harsh
immense
impair
invader
mingle
modest
observe
pursue
rebel
regret
unique
vacant
vandal
vicious

# Appendix B

Mean Raw Scores on All Measures – Controls

| Measure | Grade 4 Controls | Grade 5 Controls | Grade 6 Controls |
|---|---|---|---|
| Age (years:months) | 9:7 (2.37) | 10:5 (2.00) | 11:4 (3.00) |
| Screening measures | | | |
| NARA Comprehension[a] | 98.86 (6.08) | 98.71 (7.13) | 98.83 (6.15) |
| YARC Comprehension[a] | 103.43 (5.21) | 99.71 (5.68) | 98.67 (3.94) |
| CC2 Nonword Reading[a] | 96.70 (7.37) | 98.41 (7.73) | 96.7 (7.37) |
| CC2 Irregular Word Reading[a] | 105.44 (4.48) | 101.86 (3.87) | 102.85 (7.57) |
| Language, cognitive and fluency measures | | | |
| TOWRE Sightword Efficiency[a] | 103.29 (6.47) | 105.57 (10.65) | 106.67 (12.71) |
| TOWRE Pseudoword Decoding[a] | 101.86 (6.10) | 106.29 (7.91) | 104.67 (5.44) |
| Vocabulary reading task | 46 (6.93) | 53.43 (2.77) | 51.17 (3.48) |
| PPVT-IV | 170.71 (11.35) | 182.00 (8.98) | 175.00 (8.39) |
| Definition production task | 53.29 (24.14) | 65.71 (19.03) | 66.33 (17.91) |
| Definition recognition task | 40.29 (7.53) | 45.70 (4.56) | 45.50 (3.77) |
| Conceptual semantics task accuracy | 37 (0.76) | 36.86 (0.35) | 37.17 (0.69) |
| Conceptual semantics task reaction times | 2273.60 (301.14) | 2228.90 (386.72) | 2008.01 (397.89) |
| ACE 6-11 Naming | 18.14 (1.73) | 20.33 (1.89) | 19.67 (3.04) |
| Auditory lexical decision task | 95.14 (8.34) | 103.43 (5.18) | 109 (1.41) |
| AWMA Nonword Recall | 14 (2.83) | 16.00 (1.63) | 15.83 (2.91) |
| Syntax task - object sentences[b] | 19.14 (1.12) | 19.29 (0.88) | 19.33 (0.75) |
| Listening comprehension task | 10.57 (4.20) | 12.29 (3.61) | 10.33 (4.15) |
| AWMA Spatial Recall Processing | 54.17 (16.41) | 85.33 (33.23) | 83.33 (33.61) |
| AWMA Spatial Recall | 29.33 (20.01) | 25.67 (5.41) | 24.67 (6.16) |
| WASI-II Matrix Reasoning | 18.71 (4.46) | 19.29 (2.25) | 18.33 (2.62) |

*Note*. Standard deviations are in parentheses. [a] Assessed at screening. [b] All participants were at ceiling on subject questions, therefore these were not analyzed.

## Appendix C

Raw Scores on All Measures – Poor Comprehenders

| Measure | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 | PC7 | PC8 | PC9 | PC10 | PC11 | PC12 | PC13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Grade | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 6 | 6 |
| Age (years:months) | 9:1 | 9:6 | 9:0 | 8:11 | 9:5 | 9:9 | 9:8 | 10:0 | 9:10 | 10:2 | 10:4 | 11:5 | 11:3 |
| Screening measures | | | | | | | | | | | | | |
| NARA Comprehension[a] | 80 | 79 | 78 | 83 | 71 | 86 | 84 | 79 | 82 | 71 | 82 | 82 | 96 |
| YARC Comprehension[a] | 85 | 90 | 98 | 95 | 81 | 77 | 83 | 109 | 96 | 96 | 79 | 97 | 83 |
| CC2 Nonword Reading[a] | 99 | 103 | 102 | 104 | 98 | 91 | 96 | 111 | 105 | 106 | 92 | 105 | 97 |
| CC2 Irregular Word Reading[a] | 94 | 102 | 105 | 104 | 112 | 110 | 104 | 116 | 122 | 112 | 109 | 103 | 91 |
| Language, cognitive and fluency measures | | | | | | | | | | | | | |
| TOWRE Sightword Efficiency[a] | 106 | 101 | 110 | 109 | 91 | 110 | 118 | 127 | 107 | 114 | 116 | 103 | 99 |
| TOWRE Pseudoword Decoding[a] | 112 | 101 | 94 | 104 | 93 | 94 | 107 | 139 | 111 | 111 | 103 | 108 | 112 |
| Vocabulary reading task /61 | 37 | 39 | 47 | 40 | 38 | 36 | 48 | 54 | 48 | 51 | 40 | 54 | 52 |
| PPVT-IV | 145 | 151 | 155 | 132 | 149 | 171 | 174 | 183 | 158 | 162 | 155 | 162 | 174 |
| Definition production task /183 | 12 | 11 | 16 | 9 | 10 | 20 | 35 | 52 | 42 | 40 | 36 | 33 | 41 |
| Definition recognition task /61 | 27 | 26 | 24 | 18 | 19 | 33 | 38 | 42 | 40 | 42 | 41 | 31 | 42 |
| Conceptual semantics task accuracy /38 | 35 | 34 | 37 | 37 | 36 | 38 | 36 | 38 | 37 | 38 | 36 | 36 | 36 |
| Conceptual semantics task reaction times | 2109 | 2850 | 2760 | 3319 | 2470 | 1947 | 2569 | 2162 | 2610 | 2715 | 2336 | 2132 | 2584 |
| ACE 6-11 Naming /25 | 16 | 16 | 13 | 12 | 9 | 18 | 21 | 22 | 20 | 18 | 18 | 18 | 23 |
| Auditory lexical decision task /122 | 86 | 80 | 82 | 82 | 78 | 91 | 95 | 101 | 97 | 97 | 97 | 107 | 93 |
| AWMA Nonword Recall | 17 | 16 | 11 | 11 | 17 | 16 | 16 | 14 | 16 | 16 | 8 | 13 | 17 |
| Syntax task - object sentences[b] | 18 | 14 | 19 | 19 | 19 | 18 | 17 | 19 | 15 | 19 | 18 | 16 | 20 |
| Listening comprehension task /24 | 4 | 1 | 3 | 4 | 2 | 4 | 14 | 6 | 8 | 1 | 2 | 4 | 8 |
| AWMA Spatial Recall Processing | 43 | 60 | 47 | 51 | 14 | 32 | 16 | 80 | 49 | 75 | 84 | 54 | 147 |
| AWMA Spatial Recall | 15 | 17 | 18 | 17 | 7 | 14 | 9 | 25 | 17 | 23 | 24 | 20 | 33 |
| WASI-II Matrix Reasoning /30 | 11 | 14 | 16 | 10 | 15 | 15 | 13 | 20 | 18 | 21 | 19 | 15 | 16 |

*Note*. Standard deviations are in parentheses. [a] Assessed at screening. [b] All participants were at ceiling on subject questions, therefore these were not analyzed.

# Chapter Four: Understanding the link between oral vocabulary and reading comprehension – a randomized controlled trial

## Abstract

A number of studies have trained vocabulary skills in children and achieved improvements in reading comprehension. These results have been taken to imply that there may be a causal link between vocabulary skills and reading comprehension. However, in these studies, vocabulary training was either delivered in the context of discourse-level texts, or in conjunction with training of other oral language skills. Therefore, the specific role played by oral vocabulary skills in reading comprehension has not yet been explored. For this reason, we carried out a randomized controlled trial of oral vocabulary intervention with eleven children with poor reading comprehension. Participants were randomized to a training (n=5) or wait-list control group (n=6). Children received 30 minutes of oral vocabulary training at school, three times a week for 8 weeks. The trained group made significant gains on oral vocabulary measures and on an experimenter-designed reading comprehension task containing some of the trained words. However, improvements did not generalize to untrained words or to a standardized measure of reading comprehension, nor was any generalization observed on broader measures of oral language. The results support a direct causal link between vocabulary skills and reading comprehension.

## Introduction

Research has shown that there is a subgroup of children who, despite being able to read aloud at an age-appropriate level, have difficulties understanding what they have read (Catts, Adlof & Weismer, 2006; Nation, 2005; Nation & Snowling, 1998; Yuill & Oakhill, 1991). These children are often referred to as poor comprehenders. It is estimated that approximately 10% of school-age children fit this profile (Nation & Snowling, 1997). Because of their ability to read aloud accurately and fluently, poor comprehenders may be difficult to identify in school and clinical contexts (Clarke, Snowling, Truelove & Hulme, 2010; Nation, Clarke, Marshall & Durand, 2004; Nation, Cocksey, Taylor & Bishop, 2010). Difficulties with reading comprehension can have serious consequences for academic achievement (Cain & Oakhill, 2006b; Ricketts, Sperring & Nation, 2014). Therefore, it is important to understand the causes of these reading comprehension deficits, as well as what methods are best able to remediate them.

The Simple View of Reading is a widely cited theory which makes a claim about the possible cause of specific reading comprehension difficulties (Hoover & Gough, 1990; Gough & Tunmer, 1986). According to the Simple View, reading comprehension is the product of decoding skills and listening comprehension (oral language) skills. There is little consensus as to the definition of “decoding” within the Simple View, but for the purposes of this study, we adopt a relatively strict definition and consider a child’s decoding skills to be intact if they have age-appropriate nonword and irregular word reading skills as well as age-appropriate nonword and sight word reading fluency (see “Screening measures” below). Thus, if a child is able to decode (read) successfully and has good oral language skills, he or she will have good reading comprehension, but if either decoding or oral language skills are below average, the child will have difficulties comprehending printed text. In other words, in this view, poor reading comprehension in the presence of intact decoding skills is caused by deficits in oral language skills.

However, the claim that oral language difficulties may cause specific reading comprehension deficits is not as simple as it may at first appear. Oral language is comprised of multiple skills and processes, from knowledge of word meanings and the comprehension of word and sentence structure, to inferential abilities (Kirby & Savage, 2008). Thus, difficulty in different components of oral language could be responsible for the reading comprehension problems of poor comprehenders, and there is unlikely to be one type of oral language deficit which is responsible for the reading comprehension difficulties of every child. Indeed, poor comprehenders tend to have very heterogeneous profiles of language skill (Cain & Oakhill, 2006b; Cornoldi, de Beni & Passaglia, 1996; Nation et al., 2004).

One component of language which has been widely studied in relation to reading comprehension is vocabulary. Evidence from a recent large, randomized controlled trial by Clarke, Snowling, Truelove and Hulme (2010) indicated a possible causal role for oral vocabulary difficulties in reading comprehension. In this trial, 84 poor comprehenders from 23 different schools were randomly assigned to one of four conditions: oral language training, text comprehension training, a combined program containing elements of both oral language and text comprehension training, and a wait-list control. Oral vocabulary training was administered as part of both the oral language and combined programs.

Children in all three training programs made gains relative to the control group immediately after intervention. However, 11 months after the intervention was completed, the oral language group continued to make gains in reading comprehension and these gains were greater than either of the other two intervention groups, providing support for the hypothesis that oral language skills play a causal role in reading comprehension. In addition, a mediation analysis revealed that increases in vocabulary scores partially accounted for the effects of the oral language program, and completely accounted for the effects of the combined program. Thus, the results of the study suggest that oral vocabulary deficits may play a key role in specific reading comprehension difficulties.

Nonetheless, it is important to note that in the Clarke et al. (2010) study, oral vocabulary training formed only a part of the oral language and combined training programs, and was intermixed with other types of oral language or text comprehension training. Therefore, conclusions could not be drawn regarding the unique role of oral vocabulary.

Other studies have empirically tested the effects of vocabulary training on reading comprehension. Although some of these studies failed to find improvements in reading comprehension as a result of the training (e.g., Baumann, Carr Edwards, Font, Tereshinski, Kame'enui, & Olejnik, 2002), many did lead to improved reading comprehension scores (e.g., Beck, Perfetti, & McKeown, 1982; McKeown, Beck, Omanson & Perfetti, 1983; McKeown, Beck, Omanson & Pople, 1985; Nash & Snowling, 2006).

One of the successful vocabulary training methods used in these studies was the Rich Instruction approach (sometimes also referred to as “multiple-context learning”; Beck, McKeown & Kucan, 2002; Beck et al., 1982; McKeown et al., 1983, 1985).

Under Rich Instruction, children not only learn word meanings but are also exposed to words in a variety of meaningful contexts and engage with the meanings of words through discussion. The aim of this approach is to help children develop richly detailed, flexible and generalizable knowledge of word meaning and word usage (Beck et al., 2002). This is the method used for the vocabulary training elements of the Clarke et al. (2010) study.

Although Rich Instruction seems promising as a tool to improve reading comprehension, it is not clear that it was oral vocabulary gains that drove the reported comprehension improvements. Rich Instruction has been empirically tested in several studies (e.g., Beck et al., 1982; McKeown et al., 1983; McKeown et al., 1985) in which words were taught in the context of written texts, using both written and oral language activities. Because children were exposed to the words in written texts, improvement in comprehension might have been partially due to increased experience with discourse-level (narrative or expository) texts. Similarly, training involved exposure to the written vocabulary items. As a result, it is possible that Rich Instruction training could have led to gains in decoding of the target word, not just word comprehension. However, the ability to read the trained words was not specifically assessed in these studies, so it is not possible to determine if this was the case. Finally, of course, it is possible that the oral language aspects of the training program (hearing and talking about the meanings of the words) were crucial to comprehension improvement.

A further question about Rich Instruction relates to the fact that the training led not only to improved knowledge of trained words but also to generalization to untrained words on a multiple choice vocabulary definition task (Beck et al., 1982). The authors stated that these generalization effects may have operated through increased “word awareness” or metalinguistic skills (Beck et al., 1982). If this were the case, then the hypothesized links between vocabulary skills and reading comprehension might not be operating through vocabulary per se, but through improvements in metalinguistic awareness or broader oral language skills.

In addition to these theoretical questions, there are some methodological issues that should be noted. Firstly, the comprehension measures in these studies of Rich Instruction were only administered at post-test, and trained children’s results were compared to those of control children (Beck et al., 1982; McKeown et al., 1983; McKeown et al., 1985). Therefore, it is not possible to determine whether groups showed equal comprehension prior to training, making it difficult to interpret the results. Secondly, the studies were carried out on typically developing children in mainstream classrooms, so we do not know whether the Rich Instruction method is also effective for children with specific comprehension difficulties.

Thus, if we are to dissect the causal relationship between oral vocabulary and reading comprehension, it is necessary to determine not only that vocabulary training improves reading comprehension, but also how vocabulary training programs do so. While there is preliminary evidence that vocabulary training improves reading comprehension, we do not know whether this is because children made improvements in oral vocabulary knowledge, because they made improvements in reading ability, or because vocabulary training improves metalinguistic or broader oral language skills.

With this in mind, the current study aimed to explore the effects of oral vocabulary training on reading comprehension in a population of children with specific reading comprehension difficulties. We aimed to answer the following questions:

1) Will oral vocabulary training lead to improvements in oral vocabulary knowledge for children with specific comprehension difficulties?
2) Will oral vocabulary training lead to improvements in reading comprehension?
3) If oral vocabulary training is successful, are these improvements a result of improved knowledge of word meanings, improved word reading abilities, or improved metalinguistic/broader oral language skills?

We expected that oral vocabulary training would be successful in improving oral vocabulary knowledge. In addition, in line with the results of Clarke et al. (2010), Beck et al. (1982) and McKeown et al. (1983, 1985), we expected that training oral vocabulary skills would lead to improvements in reading comprehension. We hypothesized that such improvements in reading comprehension would be primarily driven by increased knowledge of the meanings of trained words.

## Method

Methods are described according to CONSORT 2010 and CONSORT Extension for Nonpharmacologic Treatment Interventions guidelines (Shulz, Altman & Moher, 2010; Boutron, Moher, Altman, Schulz & Ravaud, 2008a; Boutron, Moher, Altman, Schulz & Ravaud, 2008b). The study was approved by the Macquarie University Human Research Ethics Committee. Children and their parents gave informed verbal and written consent.

### Participants and sample size

All participants came from a single Catholic primary school in a middle-class area of Sydney, Australia. In July 2012, classroom teachers of Grades 3, 4 and 5 were asked to nominate children to participate in screening assessment. Because the population of interest was children who have reading comprehension difficulties but intact word reading skills, teachers were asked to nominate children they thought had poor comprehension but average word reading skills for their age. Teachers were also asked to nominate children who had average reading comprehension skills for their age to act as controls for a separate study. Consent forms for participation in screening assessment were sent out to the parents of these children and 70 were returned. These 70 participants were then screened between August and November 2013 in order to determine their eligibility for the training study.

**Screening measures**

Children's reading accuracy was assessed using the nonword and irregular word subtests of the Castles and Coltheart Reading Test 2 (CC2; Castles et al., 2009). Forty nonwords and 40 irregular words were presented to children interspersed with each other, in order of difficulty (i.e. words decreased in frequency as the test went on and nonwords increased in length as the test went on). Children were asked to read the stimuli aloud. A stopping rule of five consecutive errors applied to each item type. Scores were reported as z scores with a mean of 0 and a standard deviation of 1. Scores between -1 and 1 are considered to be within the average range.

Reading fluency was assessed on the two subtests of the Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner & Rashotte, 1999). For the Sight Word Efficiency subtest, children were instructed to read a list of words as fast as they could. For the Phonemic Decoding Efficiency subtest, children were asked to read a list of nonwords as fast as they could. The child's total score consisted of the number of words they read correctly within 45 seconds. Recently published Australian norms were used, as the original TOWRE norms were collected in the US and tend to overestimate the performance of Australian children (Marinus, Kohnen & McArthur, 2013). Standard scores between 85 and 115 were considered to be within the average range.

Reading comprehension was screened using the Neale Analysis of Reading Ability Third Edition Form 1 (NARA; Neale, 1999) and the York Assessment of Reading Comprehension Passage Reading – Australian Edition Form A (YARC; Snowling et al., 2012). The NARA has been extensively used in the reading comprehension literature (Bowyer-Crane & Snowling, 2005; Cain & Oakhill, 2006a; Nation & Snowling, 1997), thus use of the NARA allows comparability to previous studies. However, the NARA has somewhat dated norms and NARA comprehension scores have been shown to be dependent on a child's word reading abilities (Spooner, Baddeley & Gathercole, 2004). Therefore, we also used the YARC, as it is a recently developed test which has very recent Australian norms (Snowling et al., 2012).

On both tests, children read short passages aloud and answered questions about the passages. Both tests returned a text reading accuracy score, a reading comprehension score (total number of questions answered correctly) and a fluency score. For the purposes of this study we utilized only the accuracy and reading comprehension scores. This was because accuracy errors on both tests were prompted by the examiner, therefore a child's fluency score might not reflect the fluency level they achieved while reading texts independently. For both tests, standard scores between 85 and 115 were considered to be within the average range[2].

Children were eligible to participate in the training study if:

- CC2, TOWRE and accuracy scores on NARA and YARC were within the average range
- Either one or both of the NARA or the YARC reading comprehension scores were below the average range[3]
- At least one of their NARA or YARC reading comprehension score(s) were at least one standard deviation below CC2, TOWRE and NARA and YARC accuracy scores.

---

[2] NARA norms do not provide standard scores, so percentile ranks on this test were converted to standard scores.

[3] Note that 7 children scored below average on the NARA but not on the YARC, and only one scored below average on the YARC but not on the NARA. This finding is not unexpected given the fact that different reading comprehension assessments may result in difference comprehension diagnoses (Keenan, Betjemann & Olson, 2008; Keenan & Meenan, 2014). For further discussion of this topic, see Chapter Two of this thesis.

Using these criteria, 14 participants were eligible for the training study. Consent forms for training study participation were sent out to these children. One child left the school before the training program commenced and a further two declined to participate.

Therefore, the final sample size for the training study was 11 participants, one of whom was male. Screening scores for the 11 children who ultimately participated in the training study are shown in Table 1. All children were born in Australia and were fluent speakers of English. Two of these children (one in the training group and one in the control group) spoke a language other than English at home with their parents. All other participants spoke English at home.

A developmental questionnaire was administered by telephone to parents of these 11 children. According to parental report, all children had normal or corrected-to-normal vision. One child in the control group had a slight hearing loss in one ear as a consequence of an operation for cholesteatoma removal in the previous year and prior to this had a history of glue ear. Her mother reported that this did not seem to affect her day-to-day functioning. None of the children had ever been to see a speech therapist and none had any history of reading difficulties, although six of the parents (three in the control group and three in the training group) did mention some concern with their child’s academic progress. Concerns reported included a general feeling that the child was “falling behind”, concerns with the child’s abilities to organize themselves and sustain attention, and concerns with reasoning and deduction in longer mathematical problems.

**Table 1**

***Participant Screening Results***

| Measure | Training group | | | | | Control group | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Participant | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| Age at screening | 9;7 | 9;10 | 8;11 | 8;6 | 8;11 | 10;9 | 10;11 | 9;9 | 9;4 | 8;5 | 8;10 |
| Grade | 4 | 4 | 3 | 3 | 3 | 5 | 5 | 4 | 4 | 3 | 3 |
| Sex | F | F | F | F | F | F | F | F | F | F | M |
| NARA accuracy standard score | 129 | 96 | 91 | 102 | 104 | 109 | 111 | 103 | 103 | 101 | 102 |
| NARA comprehension standard score | 79 | 83 | 71 | 83 | 79 | 96 | 82 | 71 | 82 | 78 | 80 |
| YARC accuracy standard score | 117 | 99 | 94 | 103 | 104 | 95 | 100 | 101 | 101 | 109 | 96 |
| YARC comprehension standard score | 109 | 79 | 81 | 95 | 90 | 83 | 97 | 96 | 96 | 98 | 85 |
| CC2 nonword z-score | 0.73 | -0.53 | 0.83 | 0.29 | 0.2 | -0.22 | 0.34 | 0.38 | 0.33 | 0.14 | -0.07 |
| CC2 irregular word z-score | 1.09 | 0.62 | -0.13 | 0.27 | 0.11 | -0.63 | 0.19 | 0.83 | 1.48 | 0.36 | -0.38 |
| TOWRE nonword fluency standard score | 139 | 103 | 93 | 104 | 101 | 112 | 108 | 111 | 111 | 94 | 112 |
| TOWRE sightword fluency standard score | 127 | 116 | 91 | 109 | 101 | 99 | 103 | 114 | 107 | 110 | 106 |

*Note:* NARA = Neale Analysis of Reading Third Edition. YARC = York Assessment of Reading for Comprehension Passage Reading – Australian Edition. CC2 = Castles and Coltheart Reading Test 2. TOWRE = Test of Word Reading Efficiency

## Trial Design

This was a parallel study in which eligible children were randomly assigned to either a training group or a wait-list control group (see Table 2). Children were assessed on outcome measures at four time points. Assessment at times 1 and 2 (Baseline 1 and Baseline 2) occurred before and after a baseline period of no training. This baseline period provided a measure of how much children's scores on the outcome measures were affected by factors other than training, such as maturation, test-retest effects and/or regression to the mean (McArthur, Castles, Kohnen, Larsen, Jones, Anandakumar & Banales, 2013). Oral vocabulary training occurred between times 2 and 3 for the training group, with a follow-up period occurring between times 3 and 4 (see Table 2). The wait-list control group received training between times 3 and 4. Results of training for the wait-list control group are reported elsewhere (Colenbrander, Nickels, Smith-Lock, & Kohnen, 2015).

**Table 2**

***Study Design***

| Time | Training group | Control group |
|---|---|---|
| **Test 1** | **Baseline 1** | |
| 10 weeks | No intervention | No intervention |
| **Test 2** | **Baseline 2** | |
| 10 weeks | Oral vocabulary training | No intervention |
| **Test 3** | **Post-Test** | |
| 10 weeks | No intervention | Oral vocabulary training |
| **Test 4** | **Follow-Up** | |

## Intervention

In Term 2, 2013, children in the training group undertook 30 minutes of oral vocabulary training, three times a week for seven weeks. This was followed by a two-week school holiday break during which the children received no training, and one week of revision at the start of Term 3. Thus, children participated in eight weeks of training over a 10-week period. Children in the control group attended class as usual and received no extra intervention over this period. Note that training began approximately six months after the completion of screening assessment, and therefore participants were in Grades 4, 5 and 6 when they undertook training.

Training took place at school during school hours and was delivered by the first author, a trained speech pathologist. Children were taken out of class to participate in training in a single group of five children. Training sessions took place in a room inside the school's library. Roughly half of the training sessions were held in the morning and half in the afternoon – the timing of sessions was dependent on the school timetable.

Prior to the study, one child in the training group and two children in the control group had been receiving homework help from tutors outside of school hours. All parents agreed to cease any reading or language tutoring for the duration of the study.

### Lesson structure and content.

Lessons were taught from pre-prepared scripts. Example scripts are available from the first author. In weeks one to seven, children were taught five to six words per week over three lessons following the same structure (see Appendix A). Week eight was a revision week and no new words were learnt. Forty-one words were taught in total (see "Word selection" below for details).

Words were taught according to the Rich Instruction approach as outlined by Beck, McKeown and Kucan (2002, 2008). In this method, words are usually taught in the context of written texts. However, in the current study, words were taught using spoken language: meaningful context was provided by spoken sentences rather than by written narratives and expository text. Children did see the written forms of words, but exposure to written forms was minimized and incidental in nature. Children saw written words during the initial word learning phase – words were written on the board along with their meanings. Children also copied the words and definitions into their notebooks (which they had access to only during the first few minutes of the lesson), and printed flashcards of the words were used during one game per lesson. Apart from this, the training program was delivered using entirely oral language methods.

**Word selection.**

The Rich Instruction method advocates the teaching of "Tier 2" words (Beck et al., 2002). These are defined as relatively sophisticated words that have a wide applicability and are frequently used by mature language users, such as "persist" or "dubious". They are neither so basic that all speakers of a language know them (such as Tier 1 words like "bed" or "table"), nor so domain-specific and infrequent that most people will never need to use them (like "tachycardia" or "meninges"; Beck et al. 2002). We aimed to select a group of Tier 2 words that would be useful and appropriate for our participants, who were between the ages of 8 and 12.

Words were selected in the following manner: From the 30,000 words in an Age of Acquisition database (Kuperman, Stadthagen-Gonzalez & Brysbaert, 2012), we eliminated any words with an age of acquisition below 8, given that participating children should have been exposed to the majority of these words, because the youngest child in the study was 8 years and 11 months old at pre-testing. We then eliminated any words for which no frequency value was available (using N-Watch; Davis, 2005) and finally selected the words that best fitted the definition of Tier 2 words (n=151).

The 151 potential target words were grouped into broad semantic categories, because words in the aforementioned studies of the Rich Instruction method were taught in semantic categories (Beck et al., 1982; McKeown et al., 1983, 1985). Any words which did not fit into a semantic category with at least three other words from the sample were rejected. We then excluded any words for which it was not possible to devise simple, child-friendly definitions of the words using the Collins Cobuild Advanced Learner's English Dictionary (Sinclair, 2006).

The final word list comprised 82 items. These words, grouped into their semantic categories, were allocated into two sets – one set to be trained and the other to be assessed, but not trained, as a measure of generalization (to assess whether the effects of training were specific to improved knowledge of word directly taught, or whether the training was teaching more generic word-learning skills which would generalize to new words). These two groups of words were matched on measures of age of acquisition (obtained from Kuperman et al., 2012), as well as measures of imageability, written and spoken frequency, subjective familiarity, number of phonemes, number of syllables and number of letters obtained from N-Watch (Davis, 2005). We also ensured that both the trained and untrained word sets contained the same proportion of verbs, nouns and adjectives.

One set of 41 words was randomly allocated to be trained. The other set of 41 words was reduced to 20 words in order to shorten overall assessment time. These 20 words were re-matched to the trained words on all the above measures. Thus, the final list consisted of 41 trained and 20 untrained words, 61 in total (see Appendix B).

**Outcome measures**

We were primarily interested in whether the oral vocabulary training program would lead to improvements in oral vocabulary knowledge – specifically knowledge of the meaning of trained words – and whether it would lead to improvements in reading comprehension. Therefore, our primary outcome measures consisted of measures of word meaning knowledge and two assessments of reading comprehension skill. We were also interested in the reasons why oral vocabulary training led to improvements in comprehension skill. We tested the children's ability to read the trained words, and we also tested their knowledge of the phonological form of the words. In order to explore effects on metalinguistic and broader oral language skills, we assessed morphological awareness and broader oral language skills.

**Primary outcome measures.**

***Semantic knowledge.***

The trained and untrained words were assessed using a task which aimed to examine the children's knowledge of the meaning of the words. The task had two components: definition production and definition recognition (Colenbrander, 2015).

For the definition production component of the task, children were asked the meaning of a given word, for example "Assist. What does assist mean?" Children were given general prompts, such as "Just have a guess," but no specific prompts were given.

Definitions were given a score between 0 and 3. A score of 0 indicated a completely incorrect definition or a "don't know" response. A score of 1 was given if a child was able to use the word correctly in a sentence, indicating knowledge of the word's use, but there was no specific indication that the child knew the meaning of the word. For example, for the word assist, a score of 1 would be awarded for the response "He was going to assist him." A score of 2 was given if a child used a word correctly in a sentence where some knowledge of the word's meaning was displayed (for example "My mother assists me with my homework because I find it hard"), or if a child gave a definition demonstrating some knowledge of the word's meaning, but where the part of speech was incorrect (for example, "I gave my grandmother an assist when I helped her in the kitchen") or where the response indicated knowledge of the word's meaning but was poorly phrased or lacked detail (for example, "Helping"). A score of 3 was given for a definition displaying full knowledge of the word's meaning and utilizing the correct part of speech (for example, “If you assist someone, you help them with something.”).

Once children had given a definition, they completed a multiple choice definition recognition task in which they were asked to say which of three orally presented definitions best went with the word. This allowed us to test receptive knowledge of word meanings.

Correct word definitions for the definition recognition task were adapted from the Collins Cobuild Advanced Learner’s English Dictionary (Sinclair, 2006). Two additional distractor definitions for each word were developed - a semantically close, but incorrect, definition and a semantically distant incorrect definition. All options referred to the same part of speech as the target word. For example, for the word “Conceal”, the options were “To hide something”, “To pretend about something” or “To escape”. The test items and multiple choice responses were presented in a different random order at each testing time to reduce practice effects.

During training, word definitions were phrased in multiple different ways so that children were not solely exposed to the exact wording that appeared in the assessment task. However, children did occasionally hear the definitions phrased as they appeared in the definition recognition task, but the majority of the exposures to the words were in the context of verbal example sentences, or during games.

All definition responses were scored by a trained research assistant working from a marking rubric developed on the basis of adult responses to a pilot version of the task. Ten percent of responses were double-scored by an additional independent scorer. Agreement between scorers was 85%. There were no systematic disagreements. Scores differed, on average, by one point. In these cases, the first author made the final decision.

Test-retest reliability was high – Spearman correlations calculated on the Baseline 1 and Baseline 2 scores for the definition production and recognition components of the test were $r = 0.93$, $p < 0.01$ and $r = 0.83$, $p < 0.01$ respectively.

***Experimenter-designed reading comprehension task.***

A reading comprehension measure was developed which contained some of the training stimuli (Colenbrander, 2015). This consisted of two, one-page passages which the children read aloud. Each passage contained 5 of the trained and 5 matched untrained words[4]. Children were asked 15 questions about the first passage and 16 questions about the second passage[5]. Children were allowed to look back at the passages to answer the questions. The test contained both literal and inferential questions. The test was scored by a trained research assistant working from a marking rubric developed from pilot results with adults. Ten percent of responses were double-scored by an additional independent scorer. Agreement was 80%. The experimental comprehension task had a high level of test-retest reliability, $r = 0.83$, $p < 0.01$.

***Standardized reading comprehension task.***

The NARA Form 1 (Neale, 1999) was used as the standardized reading comprehension measure because our sample consisted of children who were from different school grades and we wanted to measure progress across the same comprehension passages for all participants. On the YARC (Snowling et al., 2012), children of different ages and grades read different test passages. However, children in Grades 4 to 6 whose reading accuracy is relatively good will generally complete all the test passages on the NARA.

As with all other measures, we calculated Spearman test-retest correlations for our sample using Baseline 1 and Baseline 2 data. Within our sample, accuracy scores were highly correlated across testing times, $r = 0.83$, $p < 0.01$. Reading comprehension scores were also highly correlated across testing times, $r = 0.94$, $p < 0.01$. These values were slightly different

[4] We attempted to ensure that the trained and untrained words were central to the text's meaning and that comprehension of the words' meanings was central to the ability to answer questions about text containing the words. However, this was challenging because the words were chosen before the story was constructed. Post-hoc analysis revealed that some questions could be answered without full knowledge of the words' meanings (see p.177). Therefore, we did not analyze questions requiring knowledge of trained words separately from questions requiring knowledge of untrained words.

[5] One question was removed from the first passage after piloting with adults, because adults found the question difficult to answer.

to the test-retest correlations published in the manual ($r = 0.95$ for accuracy and $r = 0.93$ for comprehension).

### Secondary outcome measures.

#### *Knowledge of phonological word form.*

Children's knowledge of the phonological form of the trained and untrained words was tested using an auditory lexical decision task (Colenbrander, 2015). A word or nonword was read aloud to the child. The child responded to each target by saying "word" or "not a word". Words and nonwords were presented in a different random order at each testing point. Nonwords were developed by changing one phoneme of the existing trained and untrained words – for example, the nonword counterpart for "impair" was "impore". There were 122 items. This task was always administered before the vocabulary definition production, definition recognition or reading tasks. There was moderate test-retest reliability, $r = 0.70$, $p < 0.05$.

#### *Ability to read vocabulary words.*

Although not directly trained, children were exposed to the written forms of words during the initial word-learning phase and one game per lesson. Therefore, as discussed above, it was possible that children might improve their ability to read aloud the trained words. Consequently, we asked children to read the trained and untrained words aloud (61 words in total). The words were presented in a different random order at each testing point. The test had high test-retest reliability, $r = 0.80$, $p < 0.01$.

#### *Morphology.*

We assessed children's morphological awareness as an index of possible improvements in metalinguistic skill (Colenbrander, 2015). Children were incidentally exposed to different morphological forms of trained words (e.g., prevent, prevention) during training, but no attempt was made to explicitly teach children about morphology. Thus, we hypothesized that improvements on a task of morphology could indicate improvements in broader metalinguistic knowledge.

Children were read aloud questions about the meanings of nonsense words, such as "Varp. Which one means something like 'a bit like a varp'?" Children then chose from four different multiple-choice responses, each of which consisted of a nonsense word base (e.g. varp) plus a suffix. One response contained the target suffix (e.g. varpish) and the other three responses contained distractor suffixes (e.g. varpable, varpence and varpless). Children were asked to read along with a written version of the questions to reduce load on phonological working memory. There were 27 items. Test-retest correlation was moderate, $r = 0.70$, $p < 0.05$.

***Listening comprehension.***

Listening comprehension was assessed as a measure of improvements in general oral language skills. We assessed listening comprehension by adapting passages 4, 5 and 6 of the NARA Form 2 (Neale, 1999). Instead of asking children to read passages aloud, the experimenter read the passages aloud to children. Children were then asked eight questions per passage (24 questions in total) and were required to answer these questions orally. Scoring was as for the standard (written) form of the NARA. There was moderate test-retest reliability, $r = 0.70$, $p < 0.05$).

For details of test development and copies of all tests, see Appendices 1 to 6 in Colenbrander (2015).

**Sequence generation, allocation concealment and implementation**

Group allocation was carried out immediately after Baseline 1 testing with all 11 participants. The first author carried out a simple randomization using the Random Numbers function of the GraphPad QuickCalcs website (http://www.graphpad.com/quickcalcs/; accessed 08/05/2013). This was carried out 9 times, each randomization resulting in two groups, one labeled "A" and one labeled "B". Given the small size of our sample, imbalances on key baseline measures, such as reading comprehension and vocabulary scores, had the potential to seriously bias study results. Consequently, the second and third authors examined the different randomizations and selected the one which represented the best match between groups. After this decision had been made, the authors tossed a coin in order to determine which of the two randomly generated groups would be the training group: Group A was designated the training group. This process resulted in a control group with slightly higher mean scores on all the matching variables. Although the differences between the groups were not significant (see “Baseline data” below), there was a possibility that the control group’s higher baseline scores could result in the underestimation of the training group’s results. We attempted to account for this in our analyses by examining not only the difference in mean raw scores at Post-Test but also the difference in amounts of improvement across the training period.

This method of group allocation was akin to an analogue method of minimization, which aims to reduce imbalance across participant groups, taking into account a variety of key prognostic variables (Pocock & Simon, 1975; Taves, 1974.) However, it is possible to predict the allocation of a new participant under some methods of minimization (Scott, McPherson, Ramsey & Campbell, 2002, Taves, 2010), and with a sample size of 11, the likelihood of being able to predict group membership was high. Our method therefore allowed for completely unpredictable, yet balanced, group allocation.

**Blinding**

As with many behavioural treatment trials, it was impossible to blind participants to their group membership (Boutron, Tubach, Giraudeau & Ravaud, 2004). Children were aware whether or not they were being taken out of class to participate in vocabulary training.

Due to limitations of time and resources, the first author carried out all assessment and training. However, a trained research assistant blind to group membership was employed to score all assessments. All tests (except for the auditory lexical decision and morphology tasks) were audio-recorded and were scored from these recordings.

## Results

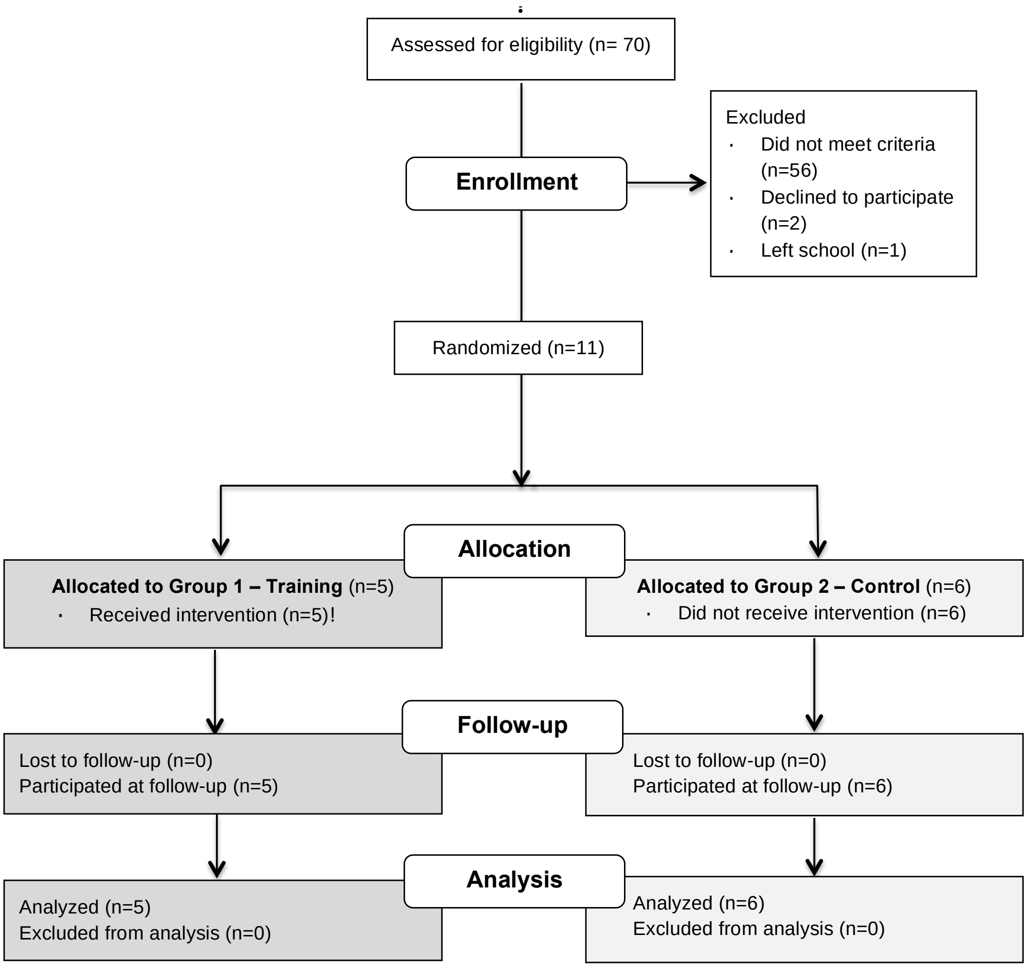


*Figure 1.* Participant flow diagram following Consolidated Standards of Reporting Trials guidelines (CONSORT, 2010).

### Numbers analyzed and participant flow

Data was analyzed for all 11 participants (5 training group participants and 6 control participants). The flow of participants through the study is shown in Figure 1.

### Baseline data

Baseline data for the trained and control group is shown in Table 3. There were no significant differences between groups across the two Baselines (see “Analysis” below). In other words, improvements over the baseline period (such as those due to maturation or practice effects, for example) were equal for both groups.

## Training fidelity

All five training group participants received training at the same time, and training was always delivered by the first author. The training was delivered according to a script and lessons followed a set format (see 'Lesson structure and content' above). Notes were taken directly onto the scripts during the lessons in order to keep track of any deviations.

A record of each child's attendance was kept. On average, children attended 86% of the lessons in the training program (the minimum attendance level was 77% and the maximum was 92%).

## Outcomes

### Analysis.

For the vocabulary definition production, definition recognition, reading and auditory lexical decision outcome measures, we wished to explore the results at the item level for both trained and untrained words. Therefore, we carried out 3 x 2 x 2 mixed ANOVAs with time of testing (Baseline 1, Baseline 2 or Post-Test) as a within-subjects variable and group (training or control) and item type (trained or untrained) as between-subjects variables. We then carried out follow-up analyses comparing trained and untrained items across time, and comparing the between-group effects for trained and untrained items separately using 3 x 2 ANOVAs with time of testing (Baseline 1, Baseline 2 or Post-Test) as a within-subjects variable and group (training or control) as a between-subjects variable.

For the remaining outcome measures, we carried out 3 x 2 mixed ANOVAs (as described above) and computed planned Helmert contrasts. The key outcome was the contrast comparing Post-Test result to aggregate baseline results. We were looking for the presence of both a main effect of time and an interaction between time and group, signaling that the improvements made over the training period were greater for the trained group than for the control group.

## Table 3

***Group means and standard deviations***

| | Baseline 1 | | Baseline 2 | | Post-Test | | Follow-Up |
|---|---|---|---|---|---|---|---|
| Measure | Training group | Control group | Training group | Control group | Training group | Control group | Training group |
| Definition production[a] | | | | | | | |
| Trained words | 0.29 (0.46) | 0.43 (0.51) | 0.36 (0.51) | 0.45 (0.59) | 2.10 (0.66) | 0.57 (0.60) | 1.68 (0.70) |
| Untrained words | 0.58 (0.62) | 0.66 (0.62) | 0.60 (0.76) | 0.58 (0.66) | 0.67 (0.73) | 0.79 (0.81) | 0.75 (0.67) |
| Definition recognition[a] | | | | | | | |
| Trained words | 0.47 (0.24) | 0.52 (0.24) | 0.51(0.25) | 0.58 (0.26) | 0.91 (0.13) | 0.61 (0.27) | 0.87 (0.17) |
| Untrained words | 0.49 (0.31) | 0.66 (0.29) | 0.62 (0.27) | 0.75 (0.25) | 0.64 (0.31) | 0.67 (0.28) | 0.57 (0.31) |
| NARA Comprehension | 14.40 (5.13) | 17.67 (5.50) | 17.00 (8.03) | 20.83 (5.27) | 20.00 (7.84) | 25.33 (4.76) | 20.40 (8.08) |
| Experimenter-designed comprehension task | 16.47 (4.07) | 19.56 (3.73) | 15.07 (3.30) | 18.39 (4.07) | 18.67 (2.15) | 19.11 (3.05) | 19.73 (2.03) |
| Auditory lexical decision task[a] | | | | | | | |
| Trained words | 0.70 (0.28) | 0.77 (0.23) | 0.75 (0.23) | 0.80 (0.22) | 0.98 (0.08) | 0.86 (0.19) | 0.96 (0.10) |
| Untrained words | 0.77 (0.22) | 0.82 (0.21) | 0.83 (0.19) | 0.87 (0.18) | 0.91 (0.21) | 0.93 (0.11) | 0.94 (0.12) |
| Vocabulary reading task[a] | | | | | | | |
| Trained words | 0.67 (0.29) | 0.79 (0.27) | 0.78 (0.27) | 0.89 (0.19) | 1.00 (0.03) | 0.92 (0.12) | 0.99 (0.06) |
| Untrained words | 0.72 (0.34) | 0.79 (0.32) | 0.83 (0.25) | 0.92 (0.15) | 0.92 (0.15) | 0.95 (0.10) | 0.95 (0.11) |
| Morphology task | 14.20 (6.76) | 16.00 (5.97) | 20.20 (2.95) | 22.67 (3.98) | 20.00 (5.24) | 22.67 (2.42) | 21.60 (4.62) |
| Listening comprehension task | 3.00 (2.00) | 4.67 (2.80) | 4.00 (4.58) | 6.17 (2.79) | 4.40 (3.51) | 6.17 (3.87) | 7.20 (5.34) |

*Note.* Standard deviations are in parentheses. NARA = Neale Analysis of Reading Third Edition. EVT-2 = Expressive Vocabulary Test Second Edition. Modified EVT-2 was not administered at Baseline 1. [a]Means calculated on item level data.

For each measure, we also calculated Cohen's *d* effect sizes for the differences between groups at Post-Test using ESCI software (Cumming, 2011).

Mean raw scores and standard deviations for each group are reported in Table 3. In addition, results are displayed graphically in Figures 2 and 3. The mean percentage correct for each group on each outcome measure is shown, as well as the 95% confidence interval for the mean.

In addition to immediate Post-Test results, we wished to know whether gains made during the training study would be maintained over time. We were unable to compare the scores of the trained group to the scores of the control group at Follow-Up because by this time point, the control group had also received training. Consequently, we used repeated measures ANOVAs to compare the Baseline 2 scores of the trained group to their own scores at Follow-Up. Results are reported below.

**Primary outcome measures.**

***Definition production task.***

For the 3 x 2 x 2 ANOVA, Mauchly's test indicated that the assumption of sphericity had been violated, $X^2(2) = 13.73$, $p = 0.001$, therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity ($\varepsilon = 0.94$). Results revealed that there was a significant overall main effect for time of testing, $F(1.87, 221.06) = 61.91$, $p < 0.001$, and item type, $F(1, 118) = 5.44$, $p = 0.02$, but no significant main effects for group, $F(1, 118) = 2.97$, $p = 0.09$. However, all two-way interactions were significant (time of test by item type, $F(1.87, 221.06) = 35.68$, $p < 0.001$; time of test by group, $F(1.87, 221.06) = 31.57$, $p < 0.001$; item type by group, $F(1, 118) = 5.44$, $p = 0.02$). The three-way interaction was also significant, $F(1.87, 221.06) = 39.64$, $p < 0.001$.

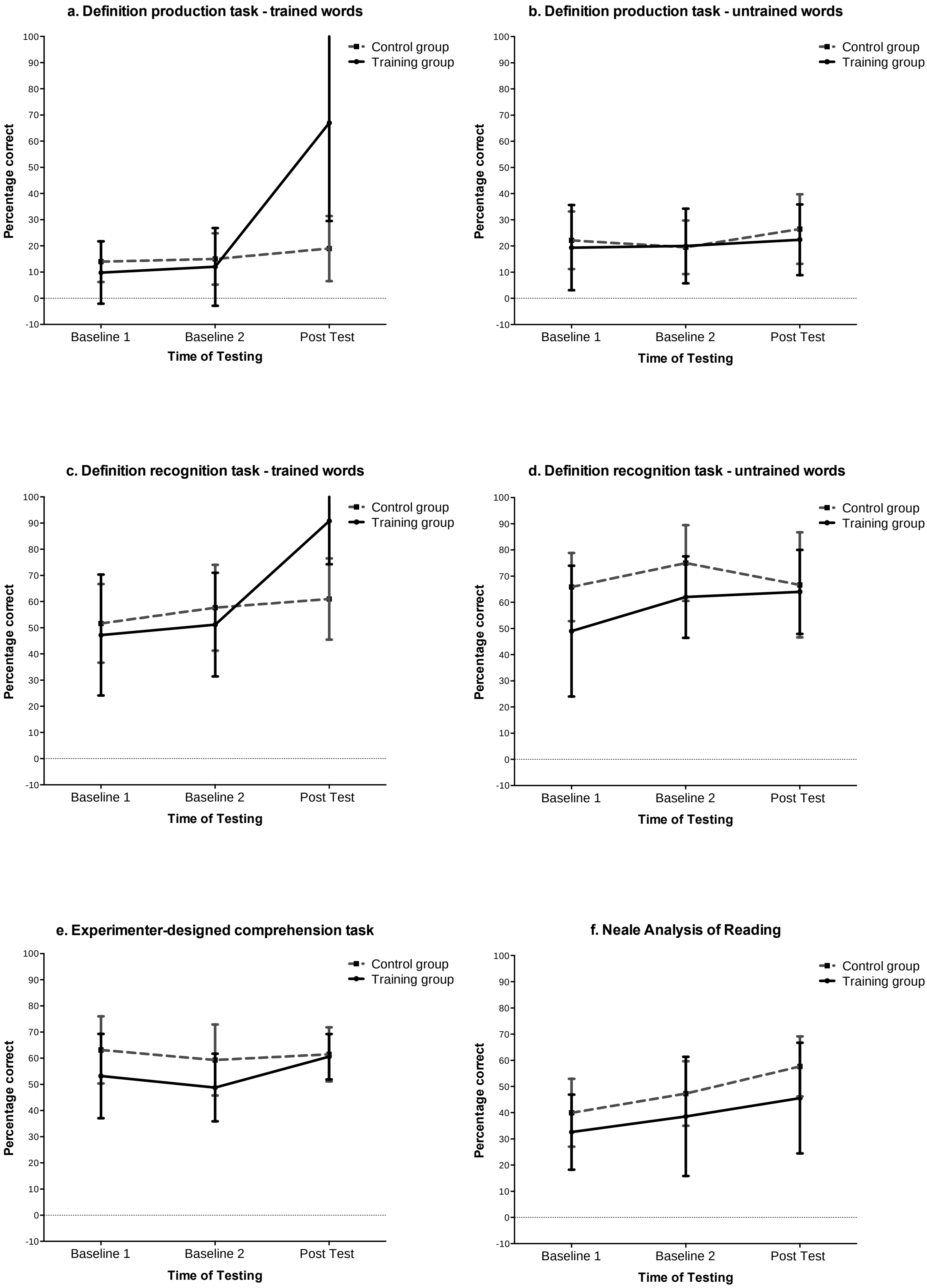


*Figure 2.* Mean percentage correct for the groups on each primary outcome measure with 95% confidence intervals. Training group results are plotted as a solid line. Control group results are plotted as a dashed line.

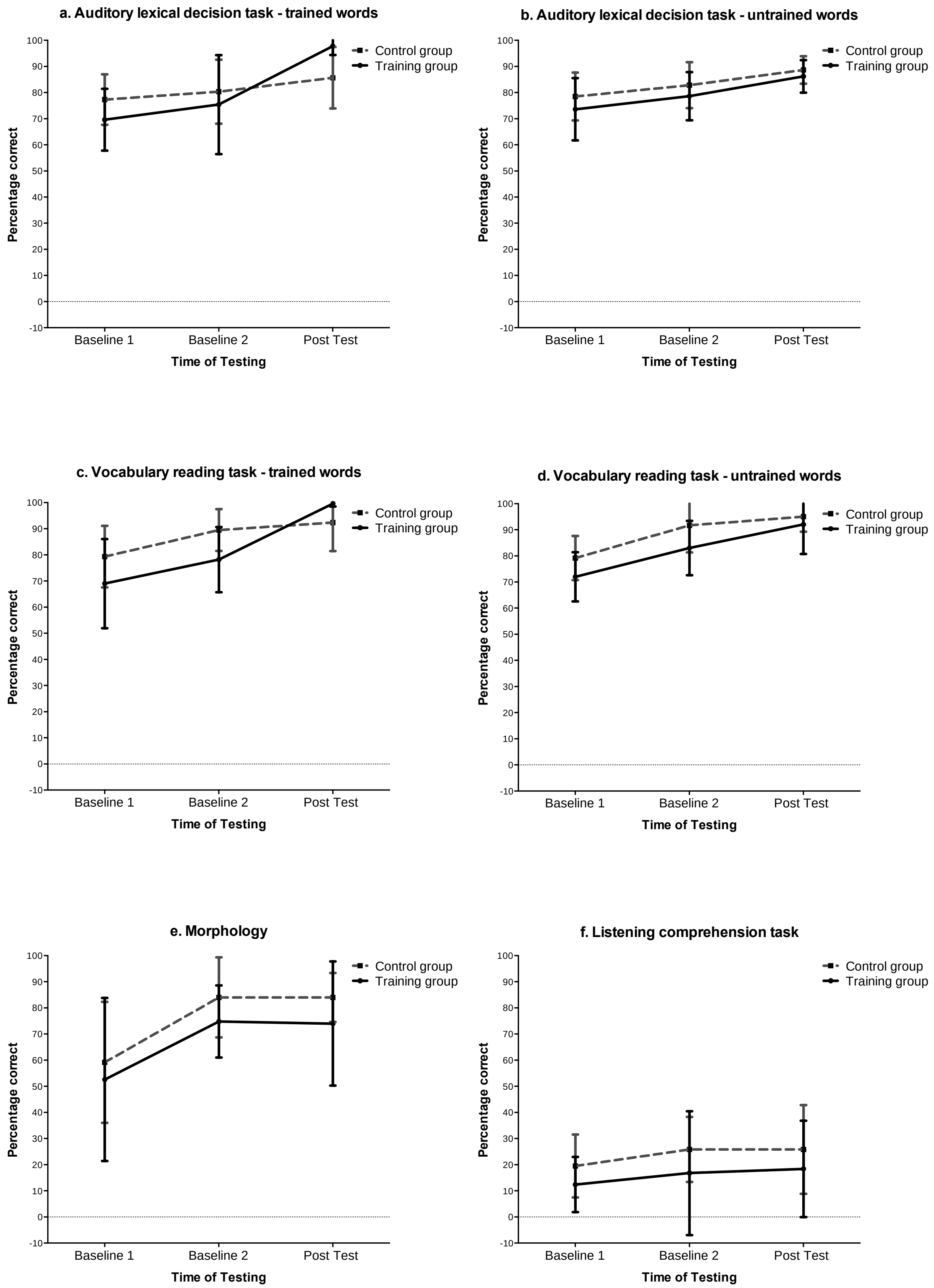


*Figure 3.* Mean percentage correct for the groups on each secondary outcome measure with 95% confidence intervals. Training group results are plotted as a solid line. Control group results are plotted as a dashed line.

To determine whether the interaction between time and item type reflected improvements in trained items after training, we examined whether there was a difference in performance on the trained items compared to the untrained items. The score for untrained items was significantly higher than the score for trained items at Baseline 1, $t(39) = -2.47$, $p = 0.02$, and there was no significant difference between the item types at Baseline 2, $t(39) = -1.63$, $p = 0.11$, while at Post-Test, the score for the trained items was significantly higher than that of the untrained items, $t(39) = 7.63$, $p < 0.001$.

We then carried out 3 x 2 ANOVAs to explore the nature of the interaction between group and time for trained and untrained items separately. For the trained items, Mauchly's test indicated that the assumption of sphericity had been violated, $X^2(2) = 14.55$, $p < 0.001$, therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity ($\varepsilon = 0.78$). There was a significant main effect for time of testing, $F(1.56, 125.01) = 227.70$, $p < 0.001$, and group, $F(1, 80) = 13.59$, $p < 0.001$, as well as a significant interaction, $F(1.56, 125.01) = 168.39$, $p < 0.001$. Planned contrasts revealed no significant difference between the baselines, $F(1, 80) = 1.85$, $p = 0.18$, nor was there an interaction, $F(1, 80) = 0.35$, $p = 0.56$, showing that there was no difference between the groups at either Baseline period. However there was a significant difference between Post-Test and Baseline scores, $F(1, 80) = 291.21$, $p < 0.001$, and a significant interaction, $F(1, 80) = 215.64$, $p < 0.001$, with the trained group scoring higher than the control group at Post-Test (see Table 3). Thus, trained children's scores on the trained items improved over the training period, but control children's scores did not.

By contrast, for the untrained items, neither the main effect of time, $F(1.71, 65.24) = 0.80$, $p = 0.44$, the main effect of group, nor the interaction, $F(1.71, 65.24) = 0.19$, $p = 0.80$, were significant (Mauchly's test was significant, $X^2(2) = 10.10$, $p = 0.006$, degrees of freedom were corrected using Huynh-Feldt estimates of sphericity, ε = 0.86). Thus, training had no effect on the untrained items.

At Follow-Up, the one-way repeated measures ANOVA on the trained items revealed that the training group's scores were significantly higher than their Baseline 2 scores, $F(1, 40) = 174.55$, $p < 0.001$. In other words, the gains made during the training period were maintained at Follow-Up.

In summary, the training program was successful in improving children's ability to define the trained words and these effects were maintained over time, but there was no transfer to matched untrained items. Cohen's *d* for the difference between groups on the trained items at Post-Test was 2.28 [95% confidence interval: 1.72, 2.84] for the trained items, a very large effect size (Cohen, 1988). By contrast, the effect size for the untrained items was -0.16 [95% confidence interval: -0.78, 0.46].

***Definition recognition task.***

As with the definition production task, the 3 x 2 x 2 ANOVA revealed that there was a significant overall main effect for time of testing, $F(2, 236) = 23.62$, $p < 0.001$, but no significant main effects of group, $F(1, 118) = 0.32$, $p = 0.57$, or item type, $F(1, 118) = 0.92$, $p = 0.34$. Again, all two-way interactions were significant (time of test by item type, $F(2, 236) = 13.08$, $p < 0.001$; time of test by group, $F(2, 236) = 15.07$, $p < 0.001$; item type by group , $F(1, 118) = 4.46$, $p = 0.03$). The three-way interaction was also significant, $F(2, 236) = 3.72$, $p = 0.03$.

Comparing trained and untrained items across time, there was no difference between trained or untrained items at Baseline 1, $t(39) = -1.65$, $p = 0.11$. At Baseline 2, the score for untrained items was significantly higher than the score for trained items, $t(39) = -3.36$, $p = 0.002$. However, at Post-Test, the trend was reversed and the score for the trained items was significantly higher than that of the untrained items, $t(39) = 5.46$, $p < 0.001$.

Examining the effects of group and time separately on trained and untrained items, the 3 x 2 ANOVA on the trained items revealed a significant main effect for time of testing, $F(2, 160) = 44.79$, $p < 0.001$. There was no significant main effect of group, $F(1, 80) = 2.54$, $p = 0.12$, but the interaction between group and time was significant, $F(2, 160) = 23.64$, $p < 0.001$. Planned comparisons indicated that there was a significant difference between Post-Test and Baseline scores, $F(1, 80) = 90.45$, $p < 0.001$, and a significant interaction, $F(1, 80) = 49.18$, $p < 0.001$, with the trained group scoring higher than the control group at Post-Test (see Table 3). In other words, trained children's scores improved after the training period, but control children's did not. There was no significant difference across the baselines, $F(1, 80) = 2.74$, $p = 0.10$, nor was there an interaction, $F(1, 80) = 0.13$, $p = 0.72$.

For the untrained items, the main effect of time was significant, $F(2, 76) = 4.56$, $p = 0.01$, but the main effect of group was not, $F(1, 38) = 1.84$, $p = 0.18$, nor was the interaction, $F(2, 76) = 1.90$, $p = 0.16$. Planned contrasts revealed that there was a significant difference between scores at Baseline 1 and Baseline 2, $F(1, 38) = 7.14$, $p = 0.01$, but no interaction, $F(1, 38) = 0.23$, $p = 0.64$. In other words, scores increased for both the trained and control group across the baseline period. There was no significant difference between groups across the training period (Post-Test compared to the baselines), $F(1, 38) = 0.67$, $p = 0.42$, but there was a significant interaction, $F(1, 38) = 4.42$, $p = 0.04$, driven by a slight increase in the control group's scores (see Table 3). Thus, training did not appear to have any effect on untrained items.

As with the definition production task, at Follow-Up, the one-way repeated measures ANOVA on the trained items revealed that the training group's scores were significantly higher than Baseline 2 scores, $F(1, 40) = 68.11$, $p < 0.001$, demonstrating that the gains made during the training period were maintained at Follow-Up.

In combination, these results indicate that the training program was successful in improving children's ability to recognize the meaning of the trained words and this effect was maintained. However, there was no transfer to matched untrained items. Cohen's *d* for the difference between groups on the trained items at Post-Test was 1.40 [95% confidence interval: 0.91, 1.88] for the trained items, a large effect size (Cohen, 1988), while the effect size for the untrained items was -0.09 [95% confidence interval: -0.71, 0.53].

***Experimenter-designed reading comprehension test.***

The 3 x 2 ANOVA showed no main effect of group, $F(1, 9) = 1.27$, $p = 0.29$, however the main effect of testing time, $F(2, 18) = 9.75$, $p = 0.001$, and the interaction, $F(2, 18) = 5.27$, $p = 0.02$, were both significant.

Comparing Baseline scores to Post-Test scores, both the main effect, $F(1, 9) = 12.17$, $p = 0.007$, and interaction, $F(1, 9) = 10.03$, $p = 0.01$, were significant. The scores of the training group increased over the training period to a greater degree than those of the control group (see Table 3). This indicates that oral vocabulary training did improve scores on the comprehension task.

Planned contrasts comparing Baseline 1 to Baseline 2 revealed a significant main effect, $F(1, 9) = 7.12$, $p = 0.03$, but no interaction, $F(1, 9) = 0.06$, $p = 0.81$. These results reflected a decrease in scores for both groups from Baseline 1 to Baseline 2. Because Baseline 1 scores were higher than Baseline 2 scores for this measure, we compared Follow-Up scores to Baseline 1 scores. Follow-Up scores were significantly higher, $F(1, 4) = 34.26$, $p = 0.03$, indicating that the effects of training were maintained over the Follow-Up period.

It is important to note that even though differences were not significant, the scores of the control group were higher than those of the training group at every time point (see Table 3). Thus, although the training group made significant gains over the training period, their scores at Post-Test were not significantly different from the control group's Post-Test scores. For this reason, a standard Cohen's *d* calculation was likely to underestimate the effects of the training program, because such calculations are based on group means and standard deviations and do not take into account within-group improvements from one testing time to another. Therefore, using the paired-data function of ESCI software (Cumming, 2011), we calculated Cohen's *d* using the training group's Baseline and Post-Test data. The effect size was 1.30, indicating a large effect of training.

***Standardized reading comprehension test: Neale Analysis of Reading.***

The main effect of testing time on reading comprehension scores was significant, $F(2, 18) = 34.65$, $p < 0.001$, but there was no main effect of group, $F(1, 9) = 1.36$, $p = 0.27$, nor an interaction, $F(2, 18) = 0.89$, $p < 0.43$. Planned contrasts revealed significant main effects for both the Baseline 1 to Baseline 2, $F(1, 9) = 8.16$, $p = 0.02$, and the Baseline to Post-Test, $F(1, 9) = 139.22$, $p <0.001$, comparisons. Neither of the interactions were significant ($F(1, 9) = 0.08$, $p = 0.79$ and $F(1, 9) = 4.11$, $p = 0.07$ respectively). This was reflective of the fact that scores for both the training and control groups increased at each testing point (see Table 3), but there was no significant difference between the groups. Cohen's *d* at Post-Test was -0.87 [95% confidence interval: -2.10, 0.40]. Thus, the training had no effect on Neale Analysis of Reading comprehension scores.

***Summary – Primary outcome measures.***

The oral vocabulary training program resulted in significant improvements in children's ability to provide and recognize verbal definitions for the trained words, but effects did not generalize to matched, untrained words. Improvements on trained words resulted in improved comprehension on the experimenter-designed task, which contained some of the trained words, but the training had no effect on a standardized task of reading comprehension (the NARA).

**Secondary outcome measures.**

***Auditory Lexical Decision Task.***

For the 3 x 2 x 2 ANOVA, Mauchly's test was significant, $X^2(2) = 6.48$, $p = 0.04$, therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity ($\varepsilon = 0.99$). Results revealed that there was a significant overall main effect for time of testing,

$F(1.98, 469.16) = 48.63$, $p < 0.001$, and item type, $F(1, 236) = 4.29$, $p = 0.04$, but no main effect of group, $F(1, 236) = 0.85$, $p = 0.36$. The two-way interaction between time of test and group, $F(1.98, 469.16) = 7.50$, $p = 0.001$, was significant, as was the three-way interaction between time of test, group and item type, $F(1.98, 469.16) = 4.49$, $p = 0.01$. The two-way interaction between time of test and item type was marginal, $F(1.98, 469.16) = 3.06$, $p = 0.05$. The two-way interaction between group and item type was not significant, $F(1, 236) = 0.67$, $p = 0.41$.

Collapsing across groups, there was a significant difference between trained and untrained items at Baseline 1, $t(75) = -3.00$, $p = 0.004$ and at Baseline 2, $t(75) = -3.09$, $p = 0.003$,with the score for untrained items being significantly higher than the score for trained items. However, at Post-Test, the trend was reversed and the score for the trained items was significantly higher than that of the untrained items, $t(39) = 2.81$, $p = 0.006$.

For the trained items, the 3 x 2 ANOVA revealed a significant main effect of testing time, $F(1.93, 312.61) = 55.65$, $p < 0.001$, and a significant interaction between time and group, $F(1.93, 312.61) = 17.61$, $p < 0.001$, but no main effect of group, $F(1, 162) = 0.01$, $p = 0.93$ (Mauchly's test was significant, $X^2(2) = 9.08$, $p = 0.01$, degrees of freedom corrected using Huynh-Feldt estimates of sphericity, $\varepsilon = 0.97$).

Turning to the planned contrasts, scores improved significantly from Baseline 1 to Baseline 2, $F(1, 162) = 94.46$, $p < 0.001$, but there was no interaction with group, $F(1, 162) = 0.51$, $p < 0.45$, indicating that scores improved for both groups across the baseline period (see Table 3). Post-Test scores were also significantly higher than baseline scores, $F(1, 162) = 6.53$, $p < 0.01$, but this time there was a significant interaction with group, $F(1, 162) = 31.07$, $p < 0.001$, with the training group making greater improvements than the control group over the training period (see Table 3). Thus, training had an effect on trained children's ability to recognize the phonological forms of the trained words.

At Follow-Up, the training group's scores on the trained items were significantly higher than Baseline 2 scores, $F(1, 81) = 57.16$, $p < 0.001$, demonstrating that the gains made during the training period were maintained at Follow-Up.

To summarize, the training program was successful in improving children's ability to recognize the phonological forms of trained words, but there was no transfer to matched untrained items. Cohen's *d* for the trained items was 0.83 [95% confidence interval: 0.51, 1.15], a relatively large effect size, while for the untrained items it was -0.15 [95% confidence interval: -0.60, 0.30].

***Vocabulary Reading Task.***

Mauchly's test was significant, $X^2(2) = 14.06$, $p = 0.001$, therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity ($\varepsilon = 0.93$). There was a significant overall main effect for time of testing, $F(1.87, 220.54) = 44.\ 31$, $p < 0.001$, but no main effect of group, $F(1, 118) = 2.47$, $p = 0.12$, or item type, $F(1, 118) = 0.07$, $p = 0.80$. The two-way interaction between time of test and group was significant, $F(1.87, 220.54) = 5.05$, $p = 0.008$, but none of the other interactions reached significance (time of test and item type: $F(1.87, 220.54) = 1.04$, $p = 0.35$, group and item type: $F(1, 118) = 0.04$, $p = 0.85$, three-way interaction: $F(1.87, 220.54) = 1.66$, $p = 0.19$).

Looking at differences between trained and untrained items (collapsing across groups), there was no difference between the item types at Baseline 1, $t(39) = -1.07$, $p = 0.29$. The difference at Baseline 2 was marginal, $t(39) = -2.04$, $p = 0.05$,with the score for untrained items being higher than the score for trained items (see Table 3). However, at Post-Test, there was a significant difference between the item types and the score for the trained items was significantly higher than that of the untrained items, $t(39) = 2.89$, $p = 0.006$.

The results of the 3 x 2 ANOVAs revealed that for trained items, there was a significant main effect for testing time, $F(2, 160) = 39.83$, $p < 0.001$, and a significant interaction, $F(2, 160) = 9.27$, $p < 0.001$, but the main effect for group was not significant, *F*(1,

80) = 1.61, $p = 0.21$. Planned contrasts indicated that scores increased from Baseline 1 to Baseline 2, $F(1, 80) = 17.34$, $p < 0.001$, but there was no interaction, $F(1, 80) = 0.03$, $p = 0.87$. Scores also increased from the baselines to Post-Test, $F(1, 80) = 58.13$, $p < 0.001$, but this time there was a significant interaction, $F(1, 80) = 16.79$, $p < 0.001$. Thus, the increase in scores was greater for the training group than the control group, indicating an effect of training.

For the untrained items, there was a significant main effect of testing time, $F(1.56, 59.26) = 14.57$, $p < 0.001$, but neither the main effect of group, $F(1, 38) = 0.95$, $p = 0.34$, nor the interaction, $F(1.56, 59.26) = 0.38$, $p = 0.63$, were significant (Mauchly's test was significant, $X^2(2) = 16.33$, $p < 0.001$, degrees of freedom were corrected using Huynh-Feldt estimates of sphericity, $\varepsilon = 0.78$). Planned contrasts indicated that scores increased from Baseline 1 to Baseline 2, $F(1, 38) = 10.24$, $p = 0.003$, and from the baseline periods to Post-Test, $F(1, 38) = 20.94$, $p < 0.001$, but there was no interaction at either time point ($F(1, 38) = 0.04$, $p = 0.83$ and $F(1, 38) = 0.88$, $p = 0.34$ respectively). In other words, training had no effect on untrained items.

Scores on the trained items were higher at Follow-Up than at Baseline 2 and the difference was significant, $F(1, 40) = 28.23$, $p < 0.001$.Thus, the effects of training were maintained over time.

In summary, the training program resulted in improvements in the ability to read the trained words, and these improvements were maintained over time, but there was no transfer to matched, untrained words. Cohen's *d* for the trained items was 0.80 [95% confidence interval: 0.35, 1.25], a relatively large effect, whereas Cohen's *d* for the untrained items was -0.24 [95% confidence interval: -0.86, 0.39].

Given that trained children made improvements in their ability to read the trained words, it was possible that such improvements may have contributed to the gains made on the experimenter-designed reading comprehension task. Therefore, we examined the training group's ability to read the 10 trained words which were included in the experimental comprehension task. At Baseline 2, six of these 10 words (60%) could be read correctly by all of the children in the training group. Of the remaining words, one was read correctly by four out of five of the children in the training group, and another was read correctly by three out of five of the children. This left only two trained words which were read incorrectly by the majority of the children in the training group. Thus, it seems unlikely that the children's comprehension improvements were a result of improved ability to read the trained words.

***Morphology Task.***

The 3 x 2 ANOVA indicated that the main effect of testing time was significant, $F(2, 18) = 14.46$, $p < 0.001$, but the main effect of group was not, $F(1, 9) = 0.90$, $p = 0.37$. The interaction between the two was also non-significant, $F(2, 18) = 0.06$, $p = 0.81$. Planned comparisons revealed that scores were significantly higher at Baseline 2 as compared to Baseline 1, $F(1, 9) = 18.46$, $p = 0.002$, and at Post-Test as compared to the Baselines, $F(1, 9) = 8.54$, $p = 0.02$. Neither interaction was significant ($F(1, 9) = 0.05$, $p = 0.83$ and $F(1, 9) = 0.07$, $p = 0.81$, respectively). Cohen's $d$ was -0.68 [95% confidence interval: -1.89, 0.57). Therefore, the training program had no effect on children's morphological abilities.

***Listening Comprehension.***

Neither the main effect of time, $F(2, 18) = 2.77$, $p = 0.09$, nor the main effect of group, $F(1, 9) = 0.99$, $p = 0.35$ nor the interaction, $F(2, 18) = 0.08$, $p = 0.93$, was significant. Cohen's $d$ was -0.48 [95% confidence interval: -1.67, 0.74]. Thus, training had no effect on children's listening comprehension abilities.

***Summary of Secondary Outcome Measures.***

Children in the training group made significant improvements in their ability to recognize the phonological forms of the trained words and in their ability to read the trained words. However, no transfer to untrained words was evident on either task. There was also no evidence of improvement on tasks of morphological skill or listening comprehension.

## Discussion

In this research, we explored the effectiveness of oral vocabulary training for children with specific reading comprehension difficulties, with the aim of examining the causal role of oral vocabulary skills in reading comprehension. Our first question related to whether vocabulary training would be effective when taught using primarily oral-language methods.

We found that the oral vocabulary training program was very successful in improving children's knowledge of trained words. Children made significant and large improvements on the primary outcome measures (definition production and recognition) – effect sizes were 2.28 and 1.40 respectively. Unlike Beck et al. (1982) and McKeown et al. (1983), we did not find any generalization to matched untrained words. The reasons for this will be discussed below. Nonetheless, our results show that poor comprehenders can learn the meanings of new words from predominantly oral language input, and without being exposed to the words in the context of discourse-level texts.

Having established that oral vocabulary training did result in improved knowledge of the meaning of trained words, we were interested in whether these improvements would result in improvements in reading comprehension. We found that children did improve on a task of reading comprehension containing some of the trained words, supporting a direct causal link between vocabulary knowledge and reading comprehension.

However, we did not find any improvements on a standardized test of reading comprehension, the Neale Analysis of Reading (NARA; Neale 1999). Given that vocabulary improvements were specific to trained words, one reason for the lack of generalization to the standardized task may be that the effects of training were very specific to trained words and to tasks containing these words. Nonetheless, it is important to note that the NARA seems to be particularly susceptible to practice effects when re-administered over relatively short periods of time (in this case, 10 weeks; see also Clarke et al., 2010). Participants in both the training and control groups made improvements on the NARA at each time of testing, with the control group making even greater improvements over the training period than the training group. Thus, it is possible that any effects of vocabulary training on NARA scores were masked by these practice effects.

Although the training focused on oral language, children were exposed to the written (orthographic) forms of the words. Exposure to the words was minimal, but it is possible that improvements in the ability to read the trained words could also have contributed to the observed gains on the experimenter-designed reading comprehension task. However, this seems unlikely because the majority of the children in the training group could read all but two of the relevant words before training began.

For this reason, improvements on the experimenter-designed reading comprehension task seem to be primarily due to improvements in oral-language based knowledge of the meanings of trained words. Nonetheless, it is possible that exposure to the written forms of the words may have helped children to learn the meanings of the words. Indeed, previous research on oral vocabulary acquisition has shown that children learn words more successfully when they are exposed to the word's written forms during training (Ricketts, Bishop & Nation, 2009; Rosenthal & Ehri, 2008).

In addition to improvements in knowledge of word meaning and written form, trained children made improvements in knowledge of the phonological forms of trained words. Therefore, the vocabulary training seems to have strengthened every aspect of children's knowledge of the trained words.

However, we found no generalization to words that were not trained. We also found no evidence of generalization to tasks of morphology or listening comprehension. It therefore seems unlikely that the observed comprehension improvements could have resulted from improvements in metalinguistic or broader oral language skills.

It is clear that our study was successful in teaching the meanings of words, and it seems reasonable that only the words that were taught should improve. However, previous studies utilizing broadly similar treatment methods have shown generalization to untrained words (e.g. Beck et al., 1982). A likely mechanism by which generalization could occur may be if the training not only teaches children the meanings of specific words, but also teaches skills that can be applied in the future for learning of new words. If this is the case then there are a number of possible reasons that could account for the lack of generalization in our study.

Firstly, our training program was much shorter in duration than previous studies. Children in our study received 24 training sessions of thirty minutes compared with 60 sessions in Clarke et al. (2010) and 75 sessions in Beck et al. (1982). It may be that the amount of training we provided was not enough to result in acquisition of the skills required to benefit the acquisition of word meanings more broadly. Secondly, our follow-up period was relatively short (10 weeks), and therefore children may not have had sufficient time to apply any new skills and learn new words. Thirdly, we assessed children on untrained words which were closely matched to the trained words. The trained words themselves were sophisticated and relatively complex words, which poor comprehenders might have difficulty learning independently, even with improved vocabulary learning skills. Thus, generalization may have occurred to simpler, more frequent words, but we were not able to detect this using our measures. Finally, it may be that exposure to words in the context of discourse-level texts during vocabulary training is important for enhancing children's ability to generalize the skills they have learnt. These are all interesting questions for future research.

Nonetheless, our results show that even over a relatively short period of training, poor comprehenders can learn the meanings of new words from predominantly oral language input, and that such gains in oral vocabulary can lead directly to reading comprehension improvements. Overall, our data provide support for a direct link between oral vocabulary skills (particularly knowledge of word meanings, or semantics) and reading comprehension. It would be interesting to further explore the causal role of oral vocabulary by comparing the effects of oral vocabulary training on poor comprehenders to the effects on a group of children with average reading comprehension. It is possible that oral vocabulary training would be effective in improving reading comprehension for children who have poor comprehension to begin with, but would have little effect on children who already have adequate comprehension skills for their age.

In addition, what remains unknown is exactly how improved knowledge of trained words led to improvements in reading comprehension. It may be that the effect of improved vocabulary knowledge was direct, with knowledge of trained words allowing children to comprehend sections of text containing those words. It could also be that improved knowledge of trained words liberated processing resources for deeper comprehension of other sections of text, leading to more general improvements in reading comprehension (Perfetti, 2007; Perfetti & Hart, 2002; Silva & Cain, 2014). These would be interesting questions for future research to explore.

**Limitations**

One limitation of this study was the small sample size. This reduces statistical power and limits the generalizations that can be made. However, smaller trials such as this can provide proof of concept for larger trials, playing a role in refining research questions and methodology and saving time and money in the implementation of future trials (Craig et al., 2008). In addition, small trials allow the exploration of individual variation across participants. This is an important consideration because poor comprehenders are a heterogeneous population (Cain & Oakhill, 2006b, 2007; Cornoldi et al., 1996), and while an intervention may have effects at the group level, improvements at the individual level may be variable (Howard, Best & Nickels, in press). Individual results from participants in this trial are reported in Colenbrander et al., (2015).

Secondly, the limited resources available for this study meant that assessment and training had to be carried out by the same person, and it was impossible to blind participants as to which group they were in. However, every effort was made to eliminate possible bias and all assessments were scored by a trained scorer who was blind to group membership.

Finally, the oral vocabulary intervention was not compared to another type of intervention as well as an untrained control group. This means that the trial could potentially be subject to Hawthorne effects, whereby participants improve simply as a result of increased attention associated with being in the trained group, rather than because of the specific intervention used (Braunholtz, Edwards & Lilford, 2001). However, the lack of transfer to untrained words or other language tasks makes this unlikely and suggests that the improvements that were observed were indeed due to the specific effects of the intervention.

**Conclusion**

Despite the limitations of this study, it is the first to show that training primarily oral vocabulary skills can lead to improvements in reading comprehension. This suggests a causal link between oral vocabulary skills and reading comprehension, and reinforces the importance of including oral vocabulary training in educational practice and remediation.

Of course, oral vocabulary skills are not the sole cause of reading comprehension difficulties. While it is necessary to know the meanings of the words in a text in order to comprehend it, knowledge of word meanings alone is not enough to guarantee successful text comprehension. Many other skills are required, such as (for example) grammatical knowledge, inferencing skills, comprehension monitoring abilities and the ability to interpret figurative language (Cain & Oakhill, 2006b; Clarke, Henderson & Truelove, 2010). Therefore, there are many factors which may play a causal role in modulating reading comprehension (Cain et al., 2006; Clarke et al., 2010; Cornoldi et al., 1996). These causal factors are tightly intertwined and it is extremely difficult to tease apart the skills involved (Cain & Oakhill, 2007). Nonetheless, experimental studies attempting to do so are necessary if we wish to refine causal models of reading comprehension and optimize the content of remediation programs.

Acknowledgements:

We wish to thank the staff, parents and students of the participating school; Marielle Smith and Yvette Kezilas for test scoring; and Serje Robidoux and Peter Petocz for assistance with statistical analysis.

## References


Adams, C., Cooke, R., Crutchley, A., Hesketh, A., & Reeves, D. (2001). *Assessment of Comprehension and Expression 6-11.* Windsor, UK: nferNelson.

Adlof, S. M., Catts, H. W., & Little, T. D. (2006). Should the Simple View of Reading include a fluency component? *Reading and Writing, 19*, 933-958. doi:10.1007/s11145-006-9024-z

Baumann, J. F., Carr Edwards, E., Font, G., Tereshinski, C. A., Kame'enui, E. J., & Olejnik, S. (2002). Teaching morphemic and contextual analysis to fifth-grade students. *Reading Research Quarterly, 37*, 150-176. doi:10.1598/RRQ.37.2.3

Beck, I. L., McKeown, M. G., & Kucan, L. (2002). *Bringing words to life: Robust vocbaulary instruction.* New York, NY: Guilford.

Beck, I. L., McKeown, M. G., & Kucan, L. (2008). *Creating robust vocabulary: Frequently asked questions and extended examples.* New York, NY: Guilford.

Beck, I. L., Perfetti, C. A., & McKeown, M. G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology, 74*, 506-521. doi:10.1037/0022-0663.74.4.506

Boutron, I., Moher, D., Altman, D. G., Schulz, K. F., & Ravaud, P. (2008a). Extending the CONSORT statement to randomized trials of nonpharmacologic treatment: Explanation and elaboration. *Annals of Internal Medicine, 148,* 295-309. doi:10.7326/0003-4819-148-4-200802190-00008

Boutron, I., Moher, D., Altman, D. G., Schulz, K. F., & Ravaud, P. (2008b). Methods and processes of the CONSORT group: Example of an extension for trials assessing nonpharmacologic treatment. *Annals of Internal Medicine, 148,* W60-W66. doi:10.7326/0003-4819-148-4-200802190-00008-w1

Boutron, I., Tubach, F., Giradeau, B., & Ravaud, P. (2004). Blinding was judged more difficult to achieve and maintain in nonpharmacologic than pharmacologic trials. *Journal of Clinical Epidemiology, 57,* 543-550. doi:10.1016/j.jclinepi.2003.12.010

Bowyer-Crane, C., & Snowling, M. J. (2005). Assessing children's inference generation: what do tests of reading comprehension measure? *British Journal of Educational Psychology, 75*, 189-201. doi:10.1348/000709904X22674

Braunholtz, D. A., Edwards, S. J. L., & Lilford, R. J. (2001). Are randomized clinical trials good for us (in the short term)? Evidence for a “trial effect”. *Journal of Clinical Epidemiology, 54*, 217-224. doi:10.1016/S0895-4356(00)00305-X

Cain, K., & Oakhill, J. (2006a). Assessment matters: issues in the measurement of reading comprehension. *British Journal of Educational Psychology, 76*, 697-708. doi:10.1348/000709905X69807

Cain, K., & Oakhill, J. (2006b). Profiles of children with specific reading comprehension difficulties. *British Journal of Educational Psychology, 76*, 683-696. doi:10.1348/000709905X67610

Carlisle, J. F. (2000). Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading and Writing, 12*, 169-190. doi:10.1023/A:1008131926604

Carlisle, J. F., & Stone, C. A. (2005). Exploring the role of morphemes in word reading. *Reading Research Quarterly, 40*, 428-449. doi:10.1598/RRQ.40.4.3

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

Catts, H. W., Adlof, S. M., & Weismer, S. E. (2006). Language deficits in poor comprehenders: A case for the Simple View of Reading. *Journal of Speech, Language and Hearing Research, 49,* 278-293. doi:10.1044/1092-4388(2006/023)

Clarke, P. J., Henderson, L. M., & Truelove, E. (2010). The poor comprehender profile: Understanding and supporting individuals who have difficulties extracting meaning from text. In J. Holmes (Ed.), *Advances in Child Development and Behaviour* (Vol. 39, pp. 79-129). New York, Boston: Academic Press.

Clarke, P. J., Snowling, M. J., Truelove, E., & Hulme, C. (2010). Ameliorating children's reading-comprehension difficulties: A randomized controlled trial. *Psychological Science, 21*, 1106-1116. doi:10.1177/0956797610375449

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Colenbrander, D. (2015). *Understanding the role of oral vocabulary in reading comprehension difficulties.* (Doctoral dissertation, Macquarie University, Sydney, Australia).

Colenbrander, D., Nickels, L. Smith-Lock, K., & Kohnen, S. (2015). *Individual differences in response to vocabulary intervention for reading comprehension difficulties.* Unpublished manuscript.

Cornoldi, C., de Beni, R., & Pazzaglia, F. (1996). Profiles of reading comprehension difficulties: An analysis of single cases. In C. Cornoldi & J. Oakhill (Eds.), *Reading Comprehension Difficulties: Processes and Intervention* (pp. 113-136). Mahwah, NJ: Lawrence Erlbaum Associates.

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2008). Developing and evaluating complex interventions: The new Medical Research Council guidance. *British Medical Journal, 337*, a1655. doi:10.1136/bmj.a1655

Cumming, G. (2011). Exploratory software for confidence intervals (ESCI). Bundoora, Australia: La Trobe University. Retrieved from http://www.latrobe.edu.au/psy/research/cognitive-and-developmental-psychology/esci

Cumming, G. (2009). Inference by eye: Reading the overlap of independent confidence intervals. *Statistics in Medicine, 28*, 205-220. doi:10.1002/sim.3471

Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist, 60*, 170-180. doi10.1037/0003-066X.60.2.170

Davis, C. J. (2005). N-Watch: A program for deriving neighbourhood size and other psycholinguistic statistics. *Behavior Research Methods, 37*, 65-70. doi:10.3758/BF03206399

Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). The impact of vocabulary instruction on passage-level comprehension of school age children: A meta-analysis. *Journal of Research on Educational Effectiveness, 2*, 1-44. doi:10.1080/19345740802539200

Garcia, J. R., & Cain, K. (2013). Decoding and Reading Comprehension: A Meta-Analysis to Identify Which Reader and Assessment Characteristics Influence the Strength of the Relationship in English. *Review of Educational Research*, *84*, 74-111. doi:10.3102/0034654313499616

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*, 6-10. doi:10.1177/074193258600700104

Hoover, W. A., & Gough, P. B. (1990). The Simple View of Reading. *Reading and Writing, 2*, 127-160. doi:10.1007/BF00401799

Howard, D., Best, W., & Nickels, L. (in press). Optimising the design of intervention studies: Critiques and ways forward. *Aphasiology*

Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading, 12*, 281-300. doi:10.1080/10888430802132279

Keenan, J. M., & Meenan, C. E. (2014). Test differences in diagnosing reading comprehension deficits. *Journal of Learning Disabilities*, *47*, 125-135. doi:10.1177/0022219412439326

Kirby, J. R., & 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

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behaviour Research Methods*, *44*, 978-990. doi:10.3758/s13428-012-0210-4

Marinus, E., Kohnen, S., & McArthur, G. (2013). Australian comparison data for the Test of Word Reading Efficiency (TOWRE). *Australian Journal of Learning Difficulties*, *18*, 199-212. doi:10.1080/19404158.2013.852981

McArthur, G., Castles, A., Kohnen, S., Larsen, L., Jones,K., Anandakumar, T., & Banales, E. (2013). Sight word and phonics training in children with dyslexia. *Journal of Learning Disabilities.* Advance online publication. doi:10.1177/0022219413504996

McKeown, M. G., Beck, I. L., Omanson, R. C., & Perfetti, C. A. (1983). The effects of long-term vocabulary instruction on reading comprehension: A replication. *Journal of Literacy Research, 15*, 3-18. doi:10.1080/10862968309547474

McKeown, M. G., Beck, I. L., Omanson, R. C., & Pople, M. T. (1985). Some effects of the nature and frequency of vocabulary instruction on the knowledge and use of words. *Reading Research Quarterly, 20*, 522-535. doi:10.2307/747940

Nash, H., & Snowling, M. (2006). Teaching new words to children with poor existing vocabulary knowledge: A controlled evaluation of the definition and context methods.

*International Journal of Language and Communication Disorders, 41*, 335-354. doi:10.1080/13682820600602295

National Reading Panel. (2000). *A report of the national reading panel: Teaching children to read.* Washington, DC: National Institute of Child Health and Human Development.

Nation, K. (2005). Comprehension. In M. J. Snowling & C. Hulme (Eds.), *The Science of Reading: A Handbook* (pp. 248-266). doi:10.1111/b.9781405114882.2005.00019.x

Nation, K., Clarke, P., Marshall, C. M., & Durand, M. (2004). Hidden language impairments in children: Parallels between poor reading comprehension and specific language impairment? *Journal of Speech, Language and Hearing Research, 47*, 199-211. doi:10.1044/1092-4388(2004/017)

Nation, K., Cocksey, J., Taylor, J. S. H., & Bishop, D. V. M. (2010). A longitudinal investigation of early reading and language skills in children with poor reading comprehension. *Journal of Child Psychology and Psychiatry, 51,* 1031-1039. doi:10.1111/j.1469-7610.2010.02254.x

Nation, K., & Snowling, M. J. (1997). Assessing reading difficulties: The validity and utility of current measures of reading skill. *British Journal of Educational Psychology, 67*, 359-370. doi: 10.1111/j.2044-8279.1997.tb01250.x

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

Neale, M. D. (1999). *The Neale Analysis of Reading Ability Third Edition*. Melbourne, Australia: ACER.

Oakhill, J. & Cain, K. (2007). Reading comprehension difficulties: Correlates, causes and consequences. In K. Cain & J. Oakhill (Eds.), *Children's comprehension problems in oral and written language: A cognitive perspective* (pp. 41-75). New York, NY: Guilford Press.

Oakhill, J. V., & Cain, K. (2011). The precursors of reading ability in young readers: evidence from a four-year longitudinal study. *Scientific Studies of Reading*, *16*, 91-121. doi:10.1080/10888438.2010.529219

Pearson, P. D., Hiebert, E. H., & Kamil, M. L. (2007). Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly, 42*, 282-296. doi:10.1598/RRQ.42.2.4

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11,* 357-383. doi:10.1080/10888430701530730

Perfetti, C., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro & P. Reitsma (Eds.), *Precursors of functional literacy* (Vol. 11, pp. 67-86). Philadelphia, PA: John Benjamins.

Ricketts, J., Bishop, D. V. M., & Nation, K. (2009). Orthographic facilitation in oral vocabulary acquisition. *Quarterly Journal of Experimental Psychology, 62*, 1948-1966. doi:10.1080/17470210802696104

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

Rosenthal, J., & Ehri, L. C. (2008). The mnemonic value of orthography for vocabulary learning. *Journal of Educational Psychology*, *100*, 175-191. doi:10.1037/0022-0663.100.1.175

Schulz, K. F., Altman, D. G., Moher, D. (2010). CONSORT 2010 Statement: updated guidelines for reporting parallel group randomized trials. *BMJ, 340*, 698-702. doi:10.1136/bmj.c332

Silva, M., & Cain, K. (2014). The relations between lower and higher level comprehension skills and their role in prediction of early reading comprehension. *Journal of Educational Psychology*. doi:10.1037/a0037769

Silverman, R. D., Speece, D. L., Harring, J. R., & Ritchey, K. D. (2013). Fluency has a role in the Simple View of Reading. *Scientific Studies of Reading*, *17*, 108-133. doi:10.1080/10888438.2011.618153

Sinclair, J. (Ed.). (2006). *Collins COBUILD advanced learner's English dictionary* (5th ed.). Glasgow, UK: HarperCollins.

Snowling, M. J., Stothard, S. E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., Nation, K., & Hulme, C. (2012). *York Assessment of Reading for Comprehension Australian Edition.* London, UK: GL Assessment.

Spooner, A. L. R., Baddeley, A. D., & Gathercole, S. E. (2004). Can reading accuracy and comprehension be separated in the Neale Analysis of Reading Ability? *British Journal of Educational Psychology, 74*, 187-204. doi:10.1348/000709904773839833

Taves, D. R. (1974). Minimization: A new method of assigning patients to treatment and control groups. *Clinical Pharmacology and Therapeutics*, *15*, 443-453. doi:10.1016/j.cct.2009.12.005

Taves, D. R. (2010). The use of minimization in clinical trials. *Contemporary Clinical Trails, 31*, 180-184. doi: 10.1016/j.cct.2009.12.005

Torgesen, J.K., Wagner, R.K., & Rashotte, C.A. (1999). *Test of Word Reading Efficiency.* Austin, TX: Pro-Ed.

Williams, K. T. (2007). *Expressive Vocabulary Test* (2nd ed.). San Antonio, TX: Pearson.

Yuill, N., & Oakhill, J. (1991). *Children's problems in text comprehension: An experimental investigation.* Cambridge, UK: Cambridge.

**Appendix A**

Structure of a Typical Lesson

1. The experimenter said a word aloud and wrote the word on the whiteboard.

   *"Prevent – the first word for today is prevent".*

2. The experimenter gave a simple, "child friendly" definition of the word in terms that the children could easily understand. Children wrote the word and its definition in their notebooks. Notebooks were then put away.

   *"Prevent means to stop something from happening."*

3. The experimenter gave three examples of the word in context.

   *"For example, you could say, 'The house was very old and would need lots of work to* prevent *it falling down.' Or you could say, 'No matter how careful you are, sometimes accidents happen, and there's nothing you can do to* prevent *them.*' *Here's another example: 'If you wear sunscreen, it will* [pause for children's replies] *sunburn'."*

4. Children then heard six sentences and had to decide whether or not the sentence was a good example of the word's meaning and explain their decision.

   *"Firemen do some backburning in the bush so that a fire does not spread."* (Good example for *prevent*)

   *"An old lady spills all her groceries on the ground. You run over to help her pick them up."* (Bad example for *prevent*)

5. Children were then asked to think of examples relating the meaning of the word to their own experience.

   *"Can you think of a time when someone prevented you from doing something? What happened?"*

6. The above steps were repeated for two further words.

7. Children then played a word association game in which they heard a series of words and had to decide which of the day's words best "went with" the words they heard. Flashcards with the written forms of the words were used during this task.

Throughout the program, the experimenter was careful to ensure that, as far as possible, all children had equal opportunities to answer questions and participate. A reward system was utilized whereby children could accumulate stickers for effort and good behaviour. These stickers could later be traded for a small prize.

## Appendix B

Trained and Untrained Words

| | |
|---|---|
| abrupt | anxious |
| abundance | conform |
| ally | conquer |
| assist | expedition |
| amateur | fatigue |
| benefactor | fury |
| brute | harsh |
| coerce | immense |
| conceal | impair |
| confide | invader |
| dubious | mingle |
| dispute | modest |
| enthusiasm | observe |
| envy | pursue |
| foe | rebel |
| frantic | regret |
| gradual | unique |
| idle | vacant |
| lunge | vandal |
| mock | vicious |
| naïve | |
| nimble | |
| offend | |
| petty | |
| ploy | |
| precise | |
| prevent | |
| request | |
| scarce | |
| scorn | |
| seize | |
| serene | |
| shrewd | |
| shudder | |
| swindle | |
| temptation | |
| tense | |
| triumph | |
| tyrant | |
| urge | |
| vague | |

# Chapter Five:
# Individual differences in response to vocabulary intervention for reading comprehension difficulties

**Abstract**

Training studies have been used to support the hypothesis that oral vocabulary skills play a role in reading comprehension. However, these training studies have not analyzed the effects of training at the individual level, and hence cannot determine for whom training is most effective, and why. Thus, in order to obtain a more nuanced understanding of the effects of training, data from an oral vocabulary training study was analyzed at the individual level. At the group level, there were significant improvements in both knowledge of the trained vocabulary and in reading comprehension. However, at an individual level, while all children made significant improvements in knowledge of the trained vocabulary words as a result of training, only four made significant reading comprehension improvements. This suggests that while poor oral vocabulary skills play a causal role in reading comprehension difficulties for some children, other children may have different underlying skill deficits which limited their reading comprehension. Results also suggest that reading comprehension remediation programs should be tailored to the skills and weaknesses of individual children in order to be most effective.

## Introduction

While many children struggle with reading comprehension tasks because of difficulties reading the words on the page, there are other children who can read aloud without difficulty, yet still have difficulty understanding what they read (Nation & Snowling, 1998; Yuill & Oakhill, 1991). These children are often referred to as poor comprehenders. Research has shown that many of these children have concurrent vocabulary difficulties (e.g., Catts, Adlof, & Weismer, 2006; Nation, Clarke, Marshall, & Durand, 2004; Nation, Snowling, & Clarke, 2007), and there is evidence to suggest that these difficulties may play a causal role in poor reading comprehension (Beck, Perfetti, & McKeown, 1982; Clarke, Snowling, Truelove, & Hulme, 2010; Colenbrander, Kohnen, Smith-Lock, & Nickels, 2015; McKeown, Beck, Omanson, & Perfetti, 1983; McKeown, Beck, Omanson, & Pople, 1985).

In particular, a large randomized controlled trial of three interventions for reading comprehension difficulties showed that oral language-based training led to significant and long-term improvements in reading comprehension (Clarke, Snowling, et al., 2010), and that these gains appeared to be mediated by improvements in oral vocabulary knowledge. We have provided further evidence for the causal role of oral vocabulary in a smaller randomized controlled trial (Colenbrander, Kohnen, et al., 2015), in which oral language-based vocabulary training led to significant improvements in reading comprehension.

However, while carefully designed randomized controlled trials are a rigorous way of determining the effectiveness of a training program (Boutron, Moher, Altman, Schulz, & Ravaud, 2008; Craig et al., 2008), results based on group means do not necessarily represent the results of all individuals within the groups (Hegde, 2007; Howard, 1986; Howard, Best, & Nickels, in press). Vocabulary training may result in comprehension gains at the group level leading to the conclusion that poor vocabulary skills are the cause of reading comprehension difficulties, but this conclusion may not apply to all participants. Indeed, if results are analyzed solely at the group level (as most training study results are), it is impossible to determine for whom the training program works, and why. This may in turn lead to recommendations for remediation that are ineffective for some participants (Cain & Oakhill, 2006; Hegde, 2007; Howard et al., in press).

The problem of ignoring individual variation is of particular concern in light of the fact that individual poor comprehenders have been shown to present with different profiles of oral language and cognitive skill (Cain & Oakhill, 2006; Colenbrander, Nickels, Smith-Lock, & Kohnen, 2015; Cornoldi, De Beni, & Pazzaglia, 1996; Nation et al., 2004). Therefore, it seems likely that there are multiple causes for specific reading comprehension difficulties (Cain & Oakhill, 2006; Clarke, Henderson, & Truelove, 2010; Oakhill & Cain, 2007). Despite this, no existing studies of vocabulary training have looked at the outcomes for poor comprehenders at the individual level.

For these reasons, the current study analyzed data from a study of oral vocabulary training at the level of the individual to determine how many participants showed improved reading comprehension skills. We also examined the individual skill profiles of each poor comprehender so that we could determine *why* the training program worked to improve reading comprehension in some participants and not in others.

## Method

### Screening

Screening took place at a Catholic primary school in a middle-class area of Sydney, Australia. In July 2012, teachers of Grade 3, 4 and 5 classes were asked to nominate children who had average word reading skills but poor reading comprehension skills, as well as children who had average word reading and reading comprehension skills for their age. Consent forms were sent out to the parents of these children. Seventy forms were returned and these participants were screened in order to determine their eligibility for the training study.

Participants were screened on the Neale Analysis of Reading Form 1 (NARA; Neale, 1999), the York Assessment of Reading Comprehension Passage Reading – Australian

Edition FORM A (YARC; Snowling et al., 2012), the nonword and irregular word subtests of the Castles and Coltheart Reading Test 2 (CC2; Castles et al., 2009) and the Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 1999).

The NARA and the YARC are similar in format. Children read short passages aloud. Any accuracy errors were prompted and reading times were recorded. Children then answered open-ended comprehension questions about the passages. On the NARA, the test was discontinued if a child made more than a specific number of accuracy errors (usually 20). On the YARC, testing was discontinued when children had read at least two passages aloud to specified accuracy levels (usually less than 20 errors), with at least two comprehension questions correct on at least one passage. Both tests returned separate scores for accuracy, fluency and reading comprehension. For our purposes, we used only accuracy and reading comprehension scores. Standard scores between 85 and 115 were considered to be within the average range[6].

The CC2 is a test of reading accuracy comprising 40 nonwords, 40 irregular words, and 40 regular words. In this study we presented only the nonwords and irregular words[7]. Items were presented interspersed and in order of difficulty. Children read the words aloud from flashcards and the test was discontinued after a child made 5 consecutive errors on any item type. Scores were reported as z scores with a mean of 0 and a standard deviation of 1. Scores between -1 and 1 were considered to be within the average range.

The TOWRE is a test of reading fluency. There are two subtests, Sight Word Efficiency (real words) and Pseudoword Decoding Efficiency (nonwords). For each subtest, children read lists of words (or nonwords) aloud. The total score was the number read correctly within 45 seconds. Rather than using the test's US norms, which tend to overestimate the performance of Australian children, we used recently collected Australian

[6] NARA norms do not provide standard scores, so percentile ranks on this test were converted to standard scores.

[7] Regular words were not administered to shorten overall assessment time and because they can be read either by sounding out or by whole word recognition, and consequently do not allow us to distinguish between different sources of reading skill.

norms (Marinus, Kohnen, & McArthur, 2013). Standard scores between 85 and 115 were considered to be within the average range.

Children were eligible to participate in the training study if their CC2, TOWRE, and NARA and YARC accuracy scores were within the normal range for their age, but at least one of their NARA and/or YARC reading comprehension scores was below the average range and at least one standard deviation below their word reading accuracy and fluency scores.

**Poor comprehender participants.**

The screening process identified 14 participants who were eligible to participate in the training study. Consent forms for training study participation were distributed to these children. Two children declined to participate and one child moved to another school. A further child attended less than 50% of the training sessions due to illness. Therefore, her data was not analyzed for the purposes of the current study. This meant that the final sample for the training study was 10 participants – 9 girls and 1 boy. All children were fluent English speakers and were born in Australia, although two participants spoke a language other than English at home with their parents. None of the children had any history of reading or language difficulties. All children had normal or corrected-to-normal vision and one child had a slight hearing loss in one ear as a result of a recent operation for cholesteatoma removal, as well as a history of glue ear.

**Controls.**

The majority of our measures were unstandardized experimenter-designed tests (see below). Therefore, in order to explore the individual skills profiles of poor comprehenders we compared their baseline scores on our assessment measures to those of controls. We collected control data from children who were in the same school grades as the poor comprehenders, but who had average word reading and reading comprehension skills for their age. Data from these control children are also reported in a study exploring the profiles of oral language skills in poor comprehenders (Colenbrander, Nickels, Smith-Lock & Kohnen, 2015)

Controls with average comprehension were recruited via the screening process described above and through the Neuronauts Brain Science Club in the ARC Centre for Excellence in Cognition and its Disorders (CCD) at Macquarie University, Sydney. In order to be eligible to participate in the study as a control participant, children had to be fluent speakers of English in Grades 4, 5 or 6, with no history of reading or language difficulties (although training study participants were screened when they were in Grades 3, 4 and 5, the training study took place the following year, when participants were in Grades 4, 5 and 6). Children had to score within the average range (standard scores between 85 and 115) on the NARA and YARC accuracy and comprehension scores and on the CC2[8]. Nine participants from the original Catholic school sample and 11 participants from the Neuronauts sample met control criteria and consented to participate in the study. Thus, the total control sample was 20 participants (7 children in Grade 4, 7 in Grade 5 and 6 in Grade 6). Average comprehenders were assessed on the screening measures and the training study outcome measures in two separate sessions. Full details of control recruitment and assessment appear in Colenbrander, Nickels, et al. (2015).

**Study design**

The training took the form of an initial parallel randomized controlled trial with two groups, a training group (Group 1) and a wait-list control group (Group 2; see Figure 1), reported in Colenbrander, Kohnen, et al. (2015). This was followed by training of the wait-list control group. The study also included design features to allow evaluation at the level of the single case (e.g. more than one pre-treatment assessment; Howard et al., in press).

Details of randomization are reported in Colenbrander, Kohnen, et al. (2015). Participants from the first training group (Group 1) were assessed at four time points, and participants from Group 2 were assessed at five time points, each 10 weeks apart. The first two testing times (Test 1 and Test 2) were baseline assessments, with no intervening training.

[8] TOWRE scores were allowed to vary as we were interested in exploring differences in fluency skills for the purposes of the profiling study (Colenbrander, Nickels, et al., 2015).

This acted as a control period for individual-level and repeated-measures analyses, allowing us to examine how children's performance changed over time without any intervention. Group 1 received training between the second and third testing times (Test 2 and Test 3). They received 7 weeks of training, followed by two weeks of school holidays and one week of revision in the following term (10 weeks in total).

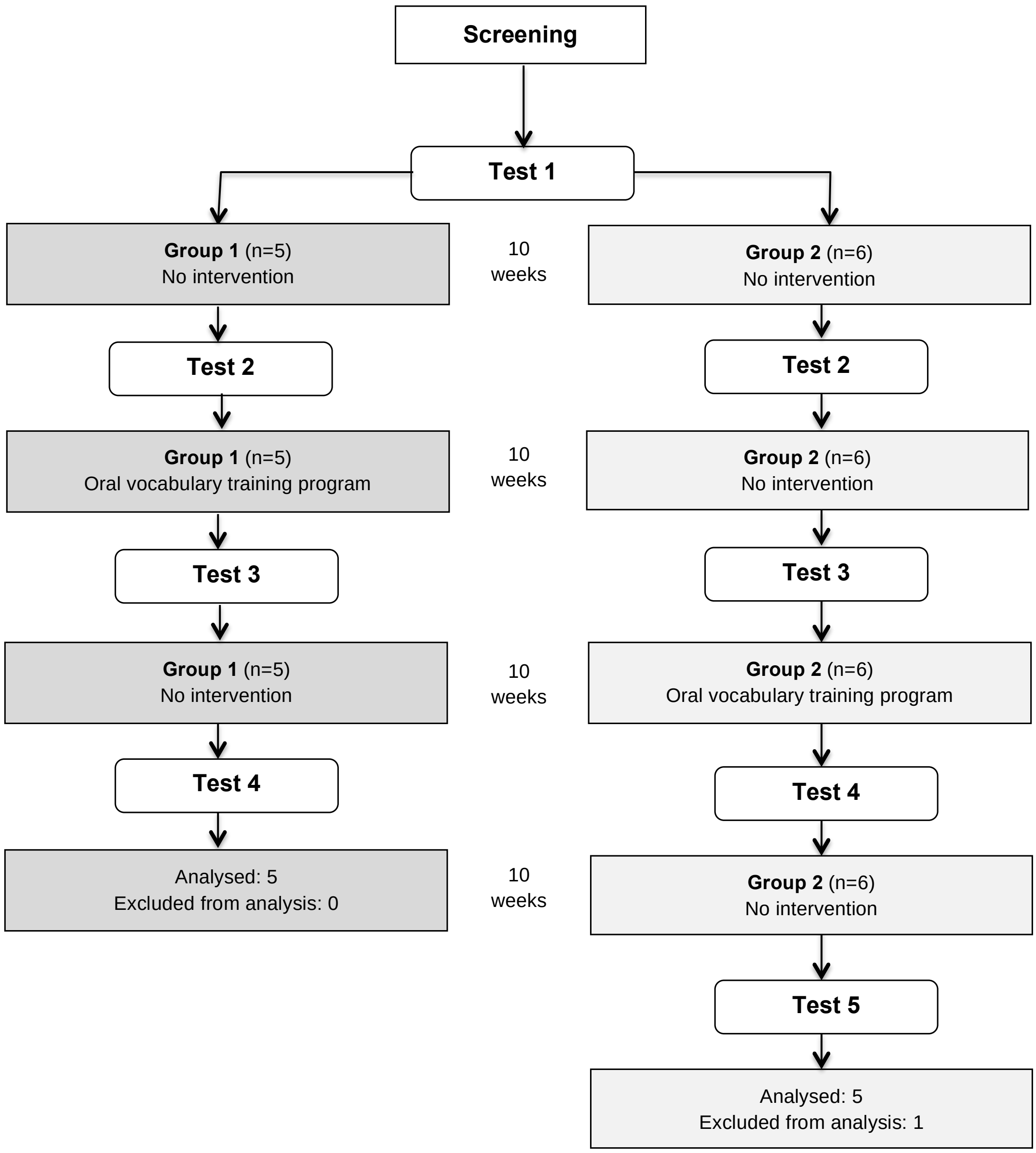


*Figure 1.* Study design.

Group 2 (wait-list control in Colenbrander, Kohnen, et al., 2015) received no training between Test 2 and Test 3, but received training between Test 3 and Test 4. Group 1 received no training during this time. Thus, Test 4 served as delayed follow-up testing for Group 1 and as immediate post-test for Group 2. Participation ended for Group 1 after Test 4, but wait-list control children were assessed again after a further 10-week period of no training, which served as their delayed follow-up assessment (Time 5).

## Outcome measures

Outcome measures are described briefly below. For more detailed descriptions of outcome measures and test development, see Colenbrander, Kohnen, et al. (2015) and Colenbrander (2015). We report the primary outcome measures from the Colenbrander, Kohnen, et al. (2015) randomized controlled trial and the results of one secondary outcome measure, the vocabulary reading task, because participants made significant improvements on this task which could potentially have accounted for the reading comprehension gains.

Assessment was carried out by the first author, who also delivered the training program (see below). However, all assessments were scored by a trained research assistant who was blind to group membership. Assessments were audio-recorded and were scored from these recordings.

### Knowledge of word meanings.

The effect of the training on children's knowledge of word meanings was assessed through tasks of definition production and recognition. Children were asked to provide the meaning of a word, for example, "Assist. What does assist mean?" If children did not know the answer they were encouraged to guess.

Regardless of whether or not children could provide a definition for the word, they completed a multiple choice definition recognition task in which they were asked to say which of three orally presented definitions best went with the word. One of the distractors was semantically close, but incorrect, while the other distractor was semantically distant.

The task assessed knowledge of 41 words which were trained during the program, and 20 words which were not trained, but which were matched to the trained words on measures of age of acquisition obtained from Kuperman, Stadthagen-Gonzalez, & Brysbaert (2012), and measures of imageability, written and spoken frequency, subjective familiarity, phonological length, number of syllables and number of letters obtained from N-Watch (Davis, 2005). A list of trained and untrained words appears in Appendix A (see Colenbrander, Kohnen, et al., 2015 for further details of word selection). Ten percent of definition responses were double-scored by an additional independent scorer and inter-rater agreement was 85%.

#### Reading comprehension.

Reading comprehension was assessed using two tasks, the NARA (Neale, 1999), described above, and an experimenter-designed reading comprehension task. On this task, children were asked to read two single-page passages aloud. After each passage, children were asked open-ended questions (15 for the first passage, 16 for the second passage). Each passage contained 5 of the trained words and 5 of the matched untrained words. Children were allowed to look back at the passages while answering the questions. Ten percent of responses were double-scored by an additional independent scorer. Agreement was 80%.

#### Ability to read vocabulary words.

As children were incidentally exposed to the written forms of the trained words during training (see “Training” below), we assessed their ability to read aloud the trained and matched untrained words.

### Training

The training was delivered to Group 1 in Terms 2 and 3 of the 2013 school year, and to Group 2 in Terms 3 and 4. Training was delivered by the first author, a trained speech-language pathologist, in a small room in the school’s library. Training was delivered in groups of five or six children. Participants received 30 minutes of training, three times a week, for 8 weeks in total.

Lessons were based on the “Rich Instruction” approach to vocabulary instruction (Beck, McKeown, & Kucan, 2002, 2008; Beck et al., 1982; McKeown et al., 1983; McKeown et al., 1985), also sometimes known as “multiple context learning” (e.g., Clarke et al, 2010). In this approach, children are taught “Tier 2” words. These are words which are used by mature language users and have a wide applicability in literary contexts, such as “persist” or “dubious”. By contrast, Tier 1 words are basic, simple words, known to the majority of speakers of a language (such as “chair” or “dog”), and Tier 3 words are domain-specific, infrequent words (such as “meninges”; Beck et al., 2002). In this approach, Tier 2 words are taught in meaning-rich contexts with a large amount of discussion about words and their meanings, so that children experience many exposures to different aspects of a word’s meaning and develop rich, detailed concepts of word meaning, rather than memorizing a definition.

In the Rich Instruction approach, children are usually taught words in the context of written texts, but in our study, words were taught in the context of spoken sentences. This was so that we could look at the effects of oral vocabulary separately from the effects of exposure to written discourse-level texts.

In weeks one to seven, children learnt five to six words per week. Week eight was a revision week. In total, 41 words were taught (see Appendix A). The structure of a typical lesson is provided in Appendix B.

**Training fidelity.**

Training for both groups was delivered by the same trainer (the first author). In order to ensure that training proceeded in the same way for both groups, lessons were taught from scripts. During Group 1 training, notes of any deviations from the script were made directly onto the script and these same scripts were used for Group 2 training.

A subset of the lessons was audio-recorded and transcribed. Appendix C contains excerpts from lesson transcripts from both groups. As demonstrated by the transcripts, lesson content was very similar for both groups.

## Results

### Group level results

Before analyzing data at the individual level, we analyzed the training study results at the group level, combining the data from Group 1 and Group 2. This allowed us to compare findings at the group and individual levels. It also allowed us to determine whether the findings of the randomized controlled trial held with a larger sample size.

When analyzing data from both groups together we used as baseline measures the two testing points immediately preceding intervention: Baseline 1 was Test 1 for Group 1 and Test 2 for Group 2; Baseline 2 was Test 2 for group 1 and Test 3 for Group 2. Post-intervention time points were immediate post-test (Group 1 Test 3; Group 2 Test 4) and delayed Follow-Up (Group 1 Test 4; Group 2 Test 5).

Means and standard deviations for each outcome measure at each time of testing are shown in Table 1. Results for each outcome measure were analyzed using repeated-measures ANOVAs with four levels of the dependent variable – Baseline 1, Baseline 2, Post-Test and Follow-Up.

**Table 1**

***Means and Standard Deviations for each Outcome Measure***

| Measure | Baseline 1 | Baseline 2 | Post-Test | Follow-Up |
|---|---|---|---|---|
| Definition production task | | | | |
| Trained words | 14.70 (11.66) | 18.30 (14.96) | 91.70 (28.15) | 81.50 (29.77) |
| Untrained words | 11.8 (6.83) | 13.30 (6.99) | 16.60 (10.44) | 18.30 (9.98) |
| Definition recognition task | | | | |
| Trained words | 21.30 (7.24) | 23.00 (6.68) | 38.60 (4.12) | 37.60 (3.37) |
| Untrained words | 12.40 (4.35) | 12.70 (3.27) | 14.30 (2.98) | 13.10 (3.45) |
| Experimental reading comprehension task | 17.33 (4.16) | 17.03 (3.77) | 20.13 (3.14) | 20.50 (2.40) |
| NARA Comprehension | 17.10 (5.63) | 20.90 (75.6) | 23.40 (6.90) | 23.40 (6.79) |
| Vocabulary reading task | | | | |
| Trained words | 32.30 (6.17) | 34.60 (4.81) | 40.90 (0.32) | 40.50 (0.71) |
| Untrained words | 16.20 (2.53) | 17.80 (1.87) | 19.10 (1.45) | 19.30 (0.82) |

*Note.* Standard deviations are in parentheses.

Pairwise comparisons were then calculated using Sidak corrections for multiple comparisons[9]. Participants' Baseline 1 and Baseline 2 scores were compared to give an indication of practice effects and levels of improvement over time without training. Baseline 2 scores were compared to Post-Test scores in order to explore the immediate effects of training. Baseline 2 scores were compared to Follow-Up scores in order to determine whether gains were maintained over a follow-up period of no training. Results of the repeated-measures ANOVAs are displayed in Table 2 and summarised below.

At the group level, the oral vocabulary training program resulted in improved ability to produce and recognize the definitions of trained words. Cohen's *d* effect sizes, calculated using ESCI software calculations for paired data (Cumming, 2011), were very large: 3.26[10] for the definition production task and 2.81[11] for the recognition task. There were also significant improvements in children's ability to read the trained words ($d$ = 1.59; 95% confidence interval: 0.55, 2.59). The gains made on the trained words were maintained at Follow-Up.

In contrast, the training had no effect on matched untrained words. There were no significant improvements for untrained words at any time point on the definition production or recognition task, although for the definition recognition task, the Baseline 2 versus Post-Test contrast for the untrained words was marginal ($p$ = 0.06). However, Tables 1 and 2 show that scores on the untrained words increased steadily at each time of testing (including the baselines). Therefore the marginal effect at Follow-Up does not seem to be a delayed effect of training – rather, it is more likely to be a result of practice effects.

[9] Sidak corrections are calculated in a similar manner to Bonferroni corrections, but result in a smaller loss of power (Field, 2009).

[10] ESCI software does not calculate confidence intervals when the unbiased *d* statistic is greater than 2. In this case, it is 2.98.

[11] Unbiased *d* was 2.57, therefore confidence intervals could not be calculated.

**Table 2**

***Repeated-Measures ANOVA Results***

| | Main effect | | | Baseline | Post-Test | Follow-Up |
|---|---|---|---|---|---|---|
| Measure | *F* | df | *p* | *p* | *p* | *p* |
| Definition production task | | | | | | |
| Trained words | 103.12 | 1.10, 9.93 | <0.001** | 0.16 | <0.001** | <0.001** |
| Untrained words | 6.82 | 3, 27 | 0.001** | 0.71 | 0.43 | 0.06 |
| Definition recognition task | | | | | | |
| Trained words | 75.21 | 1.36, 12.28 | <0.001** | 0.34 | <0.001** | <0.001** |
| Untrained words | 1.88 | 3, 27 | 0.16 | 1.00 | 0.63 | 1.00 |
| Experimental reading comprehension task | 19.91 | 3, 27 | <0.001** | 1.00 | 0.001* | 0.001* |
| NARA Comprehension | 23.18 | 3, 27 | <0.001** | 0.01* | 0.07 | 0.05 |
| Vocabulary reading task | | | | | | |
| Trained words | 17.91 | 1.24, 11.18 | 0.001* | 0.21 | 0.01* | 0.01* |
| Untrained words | 20.02 | 3, 27 | <0.001** | 0.01* | 0.01* | 0.03* |

*Note*. When estimates of sphericity were violated, the Greenhouse-Geisser correction was used when sphericity estimates were less than 0.75 and the Hyunh-Feldt correction was used when estimates were greater than 0.75 (Field, 2009). * $p < 0.05$ **p < 0.01

On the vocabulary reading task, children made significant improvements on the untrained words at every time of testing, but because scores for the untrained words improved over both the baseline and training periods, it seems likely that the improvements were a result of practice effects rather than as a result of training. The likelihood of this possibility is strengthened by the fact that the gains over the baseline period were slightly larger than those over the training period (a gain of 1.6 raw score points over the baseline period as opposed to a gain of 1.3 points over the training period), although this finding should be interpreted with caution due to the fact that four of the participants were at or near ceiling by Baseline 2. Nonetheless, the results demonstrate that the effects of training did not transfer to matched untrained words.

Despite this, the training program did lead to gains in reading comprehension skills on the experimenter-designed task containing some of the trained words ($d$ = 0.89; 95% confidence interval: 0.38, 1.39), which were maintained at Follow-Up. However, there was no improvement on the standardized reading comprehension task, which did not contain any of the trained words. These combined-group results replicated the results of the randomized controlled trial (Colenbrander, Kohnen, et al., 2015).

Because children made improvements in their ability to read the trained words, it is possible that gains on the experimenter-designed reading comprehension task may have been due to improvements in the ability to read the trained words. Therefore, we checked how many of the 10 trained words contained within the task could be read by the children before training began (in other words, how many of these words were correctly read at Baseline 2 by Group 1, and by Baseline 3 for Group 2). Six of the words were correctly read by 100% of the children before training. A further three words were correctly read by 90% of children (all but one child). The remaining word was read correctly by 80% of the children. In other words, the vast majority of the relevant words in the reading comprehension task could be read by all participants before training began. Thus, the improvement in word reading accuracy was due to an improvement in the trained words which were not included in the experimenter-designed reading comprehension task. Therefore, it is unlikely that the reading comprehension gains could be primarily due to gains in the ability to read the trained words.

**Individual level results**

Training results for individual participants were calculated using Weighted Statistics (WEST; Howard et al., in press). Weighted Statistics for evaluation of Rate of Change (WEST-ROC) was used to determine whether the amount of improvement over the training period differed from the amount of improvement over the baseline period. WEST-ROC works by multiplying the scores for each item by a weighting, summing these weightings and using these weightings to calculate a one-sample t-test (see Howard et al., in press, for details). A significant result indicated that the rate of change over the baseline period was different from that over the training period.

However, a significant result can also occur when a fall in accuracy occurs over baseline with stable performance over the training period (Howard et al., in press). Thus, in order to confirm that an effect of training is present, we also determined whether or not there was an overall trend towards improvement. The WEST-TREND procedure tests for a trend using similar methods to WEST-ROC, but different weightings (see Howard et al., in press, for details).

Therefore, to conclude that an individual child made improvements as a result of the training program, both WEST-TREND and WEST-ROC tests should be significant, indicating a significant difference between baseline and training rates of change, and a significant overall trend towards improvement.

**Outcomes.**

Raw scores for each participant are shown in Tables 3 and 4 along with each participant's WEST-ROC and WEST-TREND results for each task[12]. Raw scores for trained words are displayed in Table 3 while raw scores for untrained words are displayed in Table 4.

For the trained words on the definition production task, every child showed a significant difference in their improvements over the training period as opposed to the baseline period (WEST-ROC results), and a significant overall trend towards improvement (WEST-TREND results). In other words, training improved each and every child's ability to provide definitions for the trained words.

The majority of the participants made similar improvements on the definition recognition task. However, for two participants (Participants 1 and 6), their WEST-TREND results were significant, but their WEST-ROC results were not. This indicates that these children made improvements on the trained words across both the baseline and training periods, but the difference between their baseline and training improvements was not large enough to reach significance. Nonetheless, the results for the definition production task indicate that the training program was successful in improving vocabulary knowledge of the trained words for all participants at the individual level.

By contrast, the individual results for the vocabulary reading task show that the training had a significant effect on the ability to read the trained words for only two children (Participants 1 and 2). This is not unexpected given the fact that poor comprehenders have intact reading abilities. Thus, while significant at the group level, few individual children showed improved reading as a result of training.

Table 4 shows that no child had both significant WEST-ROC and WEST-TREND results for the untrained words on any of the vocabulary measures (definition, production, or reading). This indicates that the effects of the training program did not generalize to matched untrained words.

[12] Note: Participant numbers in this chapter do not correspond to participant numbers in other chapters.

**Table 3**

***Raw Scores and WEST Results – Comprehension Tasks and Trained Words***

| | Participant | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Task | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Definition production task<br>Trained words | | | | | | | | | | |
| Baseline 1 | 2 | 5 | 3 | 4 | 5 | 26 | 28 | 26 | 21 | 27 |
| Baseline 2 | 1 | 4 | 6 | 6 | 8 | 29 | 38 | 37 | 20 | 34 |
| Post-Test | 43 | 85 | 92 | 98 | 42 | 118 | 111 | 98 | 118 | 112 |
| Follow-Up | 33 | 74 | 80 | 80 | 29 | 111 | 94 | 94 | 109 | 111 |
| WEST-ROC | <0.001** | <0.001** | <0.001** | <0.001** | <0.001** | <0.001** | <0.001** | <0.001** | <0.001** | <0.001** |
| WEST-TREND | <0.001** | <0.001** | <0.001** | <0.001** | <0.001** | <0.001** | <0.001** | <0.001** | <0.001** | <0.001** |
| Definition recognition task<br>Trained words | | | | | | | | | | |
| Baseline 1 | 12 | 15 | 16 | 13 | 17 | 28 | 28 | 30 | 27 | 27 |
| Baseline 2 | 18 | 15 | 21 | 16 | 15 | 31 | 30 | 30 | 26 | 28 |
| Post-Test | 29 | 37 | 41 | 41 | 34 | 40 | 41 | 41 | 41 | 41 |
| Follow-Up | 32 | 38 | 38 | 39 | 31 | 40 | 39 | 40 | 38 | 41 |
| WEST-ROC | 0.21 | 0.002** | 0.02* | <0.001** | 0.003** | 0.17 | 0.02* | 0.02* | 0.01* | 0.02* |
| WEST-TREND | 0.002** | <0.001** | <0.001** | <0.001** | <0.001** | 0.003** | <0.001** | <0.001** | <0.001** | <0.001** |
| Vocabulary reading task<br>Trained words | | | | | | | | | | |
| Baseline 1 | 27 | 35 | 31 | 25 | 24 | 39 | 38 | 39 | 27 | 38 |
| Baseline 2 | 28 | 32 | 33 | 31 | 29 | 40 | 38 | 41 | 34 | 40 |
| Post-Test | 41 | 41 | 41 | 41 | 40 | 41 | 41 | 41 | 41 | 41 |
| Follow-Up | 40 | 39 | 40 | 41 | 40 | 41 | 41 | 41 | 41 | 41 |
| WEST-ROC | 0.02* | 0.01* | 0.12 | 0.23 | 0.14 | 0.50 | 0.13 | 0.92 | 0.50 | 0.71 |
| WEST-TREND | <0.001** | 0.01* | 0.002** | <0.001** | <0.001** | 0.16 | 0.04* | 1.00 | 0.003** | 0.16 |
| Experimenter-designed<br>reading comprehension task | | | | | | | | | | |
| Baseline 1 | 16.7 | 17.3 | 11.7 | 11.3 | 13.7 | 20.3 | 21.3 | 17.7 | 19.3 | 24.0 |
| Baseline 2 | 16.0 | 18.0 | 14.0 | 12.0 | 11.3 | 21.0 | 17.0 | 19.0 | 19.0 | 23.0 |
| Post-Test | 20.3 | 22.7 | 17.0 | 16.0 | 16.7 | 24.0 | 20.3 | 19.0 | 20.0 | 25.3 |
| Follow-Up | 20.7 | 21.0 | 17.0 | 16.7 | 18.7 | 22.7 | 21.3 | 21.3 | 21.3 | 24.3 |
| WEST-ROC | 0.04* | 0.05^ | 0.43 | 0.07^ | 0.01* | 0.29 | 0.02* | 0.64 | 0.34 | 0.08^ |
| WEST-TREND | 0.03* | 0.003** | 0.10 | 0.01* | 0.09^ | 0.09^ | 0.68 | 0.50 | 0.30 | 0.15 |
| Neale Analysis of Reading | | | | | | | | | | |
| Baseline 1 | 14 | 14 | 17 | 11 | 8 | 23 | 19 | 18 | 20 | 27 |
| Baseline 2 | 14 | 19 | 23 | 11 | 9 | 25 | 27 | 24 | 24 | 33 |
| Post-Test | 21 | 25 | 23 | 14 | 11 | 29 | 29 | 23 | 25 | 34 |
| Follow-Up | 16 | 24 | 23 | 15 | 14 | 27 | 33 | 25 | 24 | 33 |
| WEST-ROC | 0.07^ | 0.41 | 0.90 | 0.28 | 0.38 | 0.36 | 0.91 | 0.94 | 0.76 | 0.84 |
| WEST-TREND | 0.02* | 0.02* | 0.50 | 0.13 | 0.21 | 0.13 | 0.24 | 0.64 | 0.37 | 0.37 |

*Note*. *$p < 0.05$ **$p < 0.01$ ^Marginal or approaching significance ($0.05 < p < 0.1$)

**Table 4**

***Raw Scores and WEST Results – Untrained Words***

| | Participant | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Task | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Definition production task<br>Untrained words | | | | | | | | | | |
| Baseline 1 | 7 | 12 | 3 | 6 | 6 | 14 | 24 | 10 | 15 | 21 |
| Baseline 2 | 6 | 9 | 5 | 7 | 8 | 19 | 20 | 16 | 19 | 24 |
| Post-Test | 9 | 14 | 0 | 7 | 14 | 28 | 24 | 23 | 13 | 34 |
| Follow-Up | 14 | 14 | 8 | 9 | 6 | 28 | 30 | 24 | 16 | 34 |
| WEST-ROC | 0.35 | 0.01* | 0.96 | 0.54 | 0.36 | 0.14 | 0.44 | 0.76 | 0.25 | 0.17 |
| WEST-TREND | 0.08 | 0.29 | 0.95 | 0.50 | 0.13 | 0.50 | 0.02* | 0.78 | 0.04* | 0.01* |
| Definition recognition task<br>Untrained words | | | | | | | | | | |
| Baseline 1 | 6 | 10 | 17 | 6 | 9 | 17 | 14 | 14 | 14 | 17 |
| Baseline 2 | 11 | 8 | 9 | 11 | 10 | 15 | 16 | 17 | 14 | 16 |
| Post-Test | 9 | 12 | 17 | 12 | 13 | 19 | 16 | 14 | 14 | 17 |
| Follow-Up | 7 | 12 | 13 | 9 | 11 | 17 | 17 | 16 | 13 | 16 |
| WEST-ROC | 0.94 | 0.11 | 0.07 | 0.76 | 0.36 | 0.84 | 0.94 | 0.50 | 0.04* | 0.50 |
| WEST-TREND | 0.84 | 0.16 | 0.50 | 0.39 | 0.02* | 1.00 | 0.50 | 0.50 | 0.16 | 0.84 |
| Vocabulary reading task<br>Untrained words | | | | | | | | | | |
| Baseline 1 | 13 | 16 | 16 | 13 | 16 | 18 | 16 | 20 | 14 | 20 |
| Baseline 2 | 15 | 17 | 19 | 15 | 17 | 19 | 19 | 20 | 17 | 20 |
| Post-Test | 17 | 19 | 20 | 16 | 20 | 20 | 20 | 20 | 19 | 20 |
| Follow-Up | 19 | 18 | 20 | 18 | 19 | 20 | 20 | 20 | 19 | 20 |
| WEST-ROC | 0.50 | 0.36 | 0.75 | 0.64 | 0.29 | 0.50 | 0.84 | 1.00 | 0.63 | 1.00 |
| WEST-TREND | 0.08^ | 0.08^ | 0.16 | 0.29 | 0.04* | 0.16 | 0.16 | 1.00 | 0.08^ | 1.00 |

*Note*. *$p < 0.05$ **$p < 0.01$ ^Marginal or approaching significance ($0.05 < p < 0.1$)

Turning to the reading comprehension results, Table 3 shows that only one participant (Participant 1) had both significant WEST-ROC and WEST-TREND scores on the experimenter-designed reading comprehension task, while another (Participant 2), had a significant WEST-TREND result and a marginal WEST-ROC result. These were the same children who made significant improvements in reading aloud. It is therefore possible that the gains made by these children were a result of gains in the ability to read the trained words. This possibility is discussed in "Individual skill profiles" below.

A further participant, Participant 4, had a significant WEST-TREND result (indicating an overall trend towards improvement), but her WEST-ROC result only approached significance. This participant made a gain of 0.67[13] raw score points between Baseline 1 and Baseline 2, but gained a further 4 raw score points between Baseline 2 and Post-Test. Therefore, although the difference between the two gains does not reach significance, the direction and size of the training period gain does suggest a positive effect of training.

Participant 5 showed the opposite pattern, with a significant WEST-ROC result (a significant difference between baseline and training period improvement), but a WEST-TREND result which only approached significance. This is because Participant 5's score actually decreased by 2.34 raw score points at Baseline 2, but increased by 5.34 points from Baseline 2 to Post-Test. This represents a gain of 3 raw score points from Baseline 1 to Post-Test, which again suggests an effect of training.

Thus, at the individual level, the vocabulary training program had significant or near significant effects on reading comprehension scores for only four out of 10 participants. However, it is worth noting that the comprehension task contained half as many items as the vocabulary definition, production and reading tasks. Thus, power on the WEST-ROC and WEST-TREND tests was lower for the reading comprehension task than for the vocabulary tasks. In addition, the literal questions were very easy for the majority of participants and this meant that for some children, improvements as a result of training were only detectable on the inferential questions, further reducing power.

Turning to the NARA results, we did not expect any participants to show significant improvements at the individual level given that there was no effect of training at the group level. This is the case for 9 out of 10 participants. However, it is worth noting that Participant 1 (one of the participants who made significant gains on the experimenter-designed comprehension task) had a significant WEST-TREND result and a WEST-ROC result which approached significance. The WEST-TREND result was non-significant because the child made no improvements over the baseline period – her raw score was identical Baseline 1 and

[13] Inferential questions on the experimenter-designed comprehension test were awarded a score of 0, 1, 2 or 3 according to a marking rubric. These scores out of 3 were converted to scores out of 1 for statistical purposes. Thus, a gain of 0.67 points is possible and represents scoring 2 out of 3 for a single question.

Baseline 2. However, at Post-Test, her score increased by 7 raw score points from 14 to 21. This significant gain indicates that the oral vocabulary training did result in improved scores on the NARA comprehension task for Participant 1.

**Skill profiles associated with comprehension gains.**

***Analysis.***

Because our results showed that training was more effective for some participants than for others, it was important to understand why this was the case. We wished to determine whether children's initial levels of vocabulary and comprehension skills were related to their gains on the experimenter-designed comprehension task. Therefore, we compared children's initial performance relative to grade-matched average comprehender controls on the NARA, definition production, definition recognition and vocabulary reading tasks with the size of the gains they made on the experimenter-designed reading comprehension task.

For all of the measures of interest (NARA, definition task, multiple choice task, experimenter-designed comprehension task and vocabulary reading task) the scores of the individual poor comprehenders were compared to the scores of the grade-matched controls using a statistical method known as SinglimsES (Crawford, Garthwaite, & Porter, 2010; Crawford & Howell, 1998). SinglimsES is a modified t-test procedure implemented in a computer program which is available online (http://homepages.abdn.ac.uk/j.crawford/pages/dept/SingleCaseMethodsComputerPrograms.HTM). SinglimsES is designed to compare the results of single cases to small control populations and is accurate for samples as small as five control participants (Crawford & Howell, 1998). The test extrapolates a control population from the scores of available controls, and compares a case's score to this control population. The percentage of the control population falling below the case's score is computed – in other words, this represents the case's percentile rank within the control population. Using these Singlims-generated percentile rankings, we were able to compare the baseline scores of poor comprehenders on all outcome measures, regardless of what grade they were in.

***Outcomes.***

*Were comprehension gains related to initial vocabulary, comprehension or reading skills?*

As discussed above, the SinglimsES scores are essentially percentile rankings representing where the individual child's score falls within a control population of children in the same school grade who do not have reading comprehension difficulties.

A SinglimsES score or 10% of less (indicating that the participant scored more poorly than 90% of the estimated control population) was taken to indicate a difficulty with a particular skill. A score of 10% was significant to $\alpha < 0.10$, an acceptable alpha level given the reduced level of power which was a consequence of our small sample size (Crawford & Howell, 1998). Participants' Singlims results for the screening and outcome measures are shown in Table 5.

As shown in the table, the four participants (Participants 1, 2, 4 and 5) who made significant or near significant gains on the reading comprehension task all scored below controls on both the definition and multiple choice tasks. A further participant, Participant 3, scored below controls on both of the vocabulary tasks but his gains over the training period were not significantly different from his gains over the baseline period ($p = 0.10$). It is possible that this is due to the fact that he attended fewer training sessions than the participants who made significant gains. For example, Participant 5 attended 88% of the training sessions, (21 lessons out of 24). By contrast, Participant 3 attended 19 lessons, three lessons fewer than Participant 5. This is equivalent to missing a full week of the training program, since there were 3 lessons per week. Of the participants who did not make significant improvements, none scored below controls on both the definition and vocabulary task.

**Table 5**

***Singlims Results***

| | | | | Comprehension Improvements | | | | Singlims Results | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Participant | Grade | Age | Attendance | Raw score gain | WEST-ROC | WEST-TREND | NARA | Definition production | Definition recognition | Vocabulary reading | Comprehension task |
| 2 | 4 | 9:0 | 92% | 4.67 | 0.05^ | 0.003* | 0.92* | 9.93^ | 6.38* | 55.15 | 12.78 |
| 4 | 4 | 9:5 | 92% | 4.67 | 0.07^ | 0.01* | 0.26* | 7.22^ | 1.91* | 16.08 | 0.70* |
| 1 | 4 | 8:11 | 90% | 3.66 | 0.04* | 0.03* | 2.52* | 6.85^ | 1.62* | 22.45 | 9.99^ |
| 5 | 4 | 9:6 | 88% | 3.00 | 0.01* | 0.09^ | 1.12* | 7.62^ | 6.31^ | 19.06 | 2.74* |
| 3 | 4 | 9:1 | 81% | 3.00 | 0.43 | 0.10 | 1.36* | 8.04^ | 7.49^ | 58.80 | 3.40* |
| 6 | 5 | 9:10 | 90% | 3.00 | 0.29 | 0.09^ | 3.54* | 14.40 | 14.33 | 5.82^ | 40.87 |
| 10 | 6 | 11:3 | 65% | 1.33 | 0.08^ | 0.15 | 34.39 | 12.37 | 21.47 | 58.3 | 98.96 |
| 9 | 5 | 10:4 | 85% | 0.67 | 0.34 | 0.30 | 3.54* | 9.72^ | 18.61 | 0.20* | 13.47 |
| 8 | 5 | 10:2 | 83% | 0.00 | 0.64 | 0.50 | 0.54* | 12.66 | 23.83 | 22.16 | 8.23^ |
| 7 | 5 | 10:0 | 77% | -1 | 0.02* | 0.68 | 2.07* | 26.27 | 23.83 | 57.31 | 36.37 |

*Note.* Results are displayed in order of the size of comprehension gains. Singlims results represent the proportion of the control population who would be estimated to score below the poor comprehender's score. A Singlims score of 10% or less is equivalent to a standard score of 81 and was considered to indicate difficulties with a particular skill. Singlims scores less than 10% are highlighted in grey. $*p < 0.05$ $^p < 0.10$

We also checked how many of the 10 trained words in the experimenter-designed comprehension task were known by the children before training began (see Table 6). Mann-Whitney U tests indicated that the children who made comprehension improvements could recognize ($U$ = 1.50, z = -2.32, $p$ = 0.02) and produce ($U$ = 3.00, z = -2.01, $p$ = 0.04) definitions for significantly fewer of the relevant words than children who did not make improvements.

Together, these results suggest that the children who had poorer vocabulary skills to begin with benefited the most from training (although it is worth noting that most of the children who did not make gains also attended fewer sessions than the children who did make gains).

***Table 6***

***Participants' Knowledge of Trained Words Included in the Experimental Reading Comprehension Task at Baseline 2***

| Participant | Significant comprehension gains | Meanings recognized | Definitions provided |
|---|---|---|---|
| 1 | Yes | 4 | 0 |
| 2 | Yes | 4 | 0 |
| 3 | No | 4 | 0 |
| 4 | Yes | 1 | 1 |
| 5 | Yes | 4 | 1 |
| 6 | No | 7 | 4 |
| 7 | No | 9 | 4 |
| 8 | No | 6 | 3 |
| 9 | No | 5 | 1 |
| 10 | No | 7 | 4 |

*Note.* A total of 10 words were included in the experimental comprehension task.

*Were comprehension gains a result of improved reading abilities?*

The WEST statistics results discussed previously revealed that Participants 1 and 2 made significant gains on the vocabulary reading task and significant (or near-significant) gains on the experimenter-designed comprehension task. Therefore, we wished to explore whether there was any evidence that the comprehension gains they made were a result of improved ability to read the trained words. To do so, we looked at each child's vocabulary reading, definition recognition and definition production scores at Baseline 2 and Post-Test for the 10 trained words which were included in the reading comprehension task.

On the vocabulary reading task, Participant 1 read 9 out of the 10 words correctly at Baseline 2, but could recognize the meanings of only four of the words and could define none of the words. At Post-Test, she again read 9 of the 10 words correctly, but could recognize five of the word definitions and could define two of the words. These results represent a relatively small gain on the relevant trained words, and her reading comprehension gains were made on questions which did not contain the trained words. However, the fact that her reading aloud of the trained words was close to ceiling at Baseline indicates that improvements in word reading cannot count for her comprehension gains. Importantly, her gains on the definition production task for all the trained words (not just those included in the comprehension task) were impressive – she went from a score of 1 at Baseline 2 to a score of 43 at Post-Test. She is also the only child who made improvements on the NARA. It therefore appears that this child made broad comprehension improvements as a result of the vocabulary training.

By comparison, the comprehension gains of Participant 2 appear to be a more direct result of vocabulary gains. She correctly read all 10 of the relevant trained items at Baseline 2 and Post-Test. At Baseline 2, she could recognize the meanings of four words, and define none. However at Post-Test, she could recognize and define the meanings of 9 of the words, and her comprehension scores improved on six of the questions requiring knowledge of the trained words, implying a direct link from vocabulary gains to comprehension gains. Nevertheless, this finding should be interpreted with caution because Participant 2's scores also improved on 5 items which did not require knowledge of the trained word meanings, and prior analysis revealed that it was possible for some of the questions to be correctly answered without knowing the word meanings. However, she could correctly read all the relevant trained items before training began, so it is clear that her comprehension gains could not be due to gains in ability to read the trained words. She too made impressive gains on the vocabulary definition task (from a score of 5 at Baseline 2 to a score of 85 at Post-Test). Both Participant 1 and Participant 2 scored significantly below controls on the vocabulary definition and production tasks at Baseline.

*Correlations.*

In order to further explore the link between vocabulary skills and reading comprehension improvements, children's SinglimsES percentile rankings were correlated with the gains they made on the experimenter-designed reading comprehension task. Gains were calculated by subtracting each child's highest baseline score from their Post-Test score.

**Table 7**

***Spearman's Correlations between Comprehension Gains and Initial Vocabulary and Comprehension Scores***

| Baseline 1 scores | Comprehension gains |
|---|---|
| Definition production task total score | -0.66* |
| Definition recognition task total score | -0.89** |
| Experimental comprehension task | -0.43 |
| Vocabulary reading task | -0.08 |
| Neale Analysis of Reading | -0.35 |

*Note*. * $p < 0.05$ ** $p < 0.01$

Spearman correlation coefficients are shown in Table 7. Comprehension gains were significantly negatively correlated with both the Baseline 1 definition task score and the multiple choice score. This indicates that the largest comprehension gains were made by the children who had the lowest scores on the vocabulary tasks before training commenced. No other correlations were significant. Together, this data and the individual-level data suggest that the oral vocabulary training seems to have been most effective for children who had poorer vocabulary skills to begin with.

## Discussion

Although many studies have shown that children with specific reading comprehension difficulties have concurrent vocabulary difficulties (e.g., Catts et al., 2006; Nation et al., 2004; Nation et al., 2007) the exact nature of the role of vocabulary in reading comprehension has not been clear, especially given that not all poor comprehenders have vocabulary difficulties (Cain & Oakhill, 2006; Nation et al., 2004). Training studies are a useful way of demonstrating a causal link between two skills (Nickels, Kohnen, & Biedermann, 2011), and indeed, training studies have demonstrated a link between vocabulary improvements and reading comprehension improvements (Beck et al., 1982; Clarke, Snowling, et al., 2010; Colenbrander, Kohnen, et al., 2015; McKeown et al., 1983; McKeown et al., 1985). However, the results of training studies are usually analyzed at the group level, meaning that we cannot be sure that every child benefits (Hegde, 2007; Howard, 1986; Howard et al., in press). Therefore, in the current study, we analyzed the results of an oral vocabulary training study at both the group and individual level in order to compare the pattern of results at these levels and determine whether oral vocabulary gains resulted in reading comprehension gains for all, or only some, participants.

The group-level results showed that oral vocabulary training was successful in improving participants' knowledge of the trained words. These gains led to improvements on an experimenter-designed reading comprehension task containing some of the trained words, replicating the results of the Colenbrander, Kohnen, et al. (2015) randomized controlled trial. Therefore, the group level results provide evidence of a direct causal link between vocabulary knowledge and reading comprehension whereby improved knowledge of particular vocabulary items improves comprehension of texts including those items.

The direct and specific nature of this causal link is supported by the fact that (as in the randomized controlled trial) no improvements were made on a standardized task of reading comprehension which did not contain any trained words, the Neale Analysis of Reading (NARA; Neale, 1999). There was also no evidence of improvements on matched untrained words. However, our vocabulary training program was relatively short, and we cannot rule out the fact that a longer period of training might result in broader generalization. There is also some evidence that the NARA is subject to inflated practice effects when used over short periods of time (see Colenbrander, Kohnen, et al., 2015), which may have masked any gains children may have made[14].

The individual level results provide a more detailed picture of the nature of the links between vocabulary and reading comprehension. Although all participants made improvements in knowledge of the trained words, only four made significant or near significant improvements in reading comprehension. Analysis of the vocabulary and reading comprehension skills of the individual children revealed that the children with the poorest vocabulary skills before training were also those who made the largest comprehension gains. By contrast, vocabulary training did not result in improved comprehension scores for children who had age appropriate vocabulary skills (those whose scores on the definition production and recognition tasks at Baseline were not significantly different from controls). These children could recognize and define many of the words which were included in the reading

---

[14] Note that in our data, practice effects were observed over the Baseline period, but not over the Follow-Up period (see Table 1). It may be that practice effects plateau after a certain number of testing occasions. Attenuation of practice effects over time has been demonstrated on neuropsychological assessments (Beglinger et al., 2005), but further research is required to confirm whether this is also the case with reading comprehension assessments."

comprehension task before training began. This suggests that poor vocabulary skills play a causal role in reading comprehension difficulties for some, but not all, children.

It is possible that the children who did not show improvements in comprehension may have had other oral language difficulties which played a role in causing their reading comprehension difficulties. Therefore, even when oral vocabulary skills were improved, other oral language deficits may have limited their comprehension scores. Support for this position comes from a study exploring the profiles of oral language skill in the sample of poor comprehenders reported here (Colenbrander, Nickels, et al., 2015). The results of this study showed that although not all poor comprehenders in the sample had vocabulary deficits, all had some form of oral language deficit compared to age-matched controls.

In general, our findings point to a specific link between *oral* vocabulary skills and reading comprehension. It was true that at the group level, children's ability to read aloud the trained words improved significantly, even though children were only exposed to the written forms of words a limited number of times, and attention was never drawn to the words' written forms. It was possible, therefore, that comprehension improvements could have been due to improved reading abilities. However, we showed that this was unlikely to be the case, because before training began, the children were at ceiling at reading aloud the trained words which were included in the experimenter-designed comprehension task. Nevertheless, these results do demonstrate that poor comprehenders can learn the orthographic forms of new words relatively easily, without any direct instruction.

Support for the specific role of oral vocabulary was also found at the individual level. Two participants made improvements on both the vocabulary reading task and the experimenter designed reading comprehension task. Their ability to read the trained vocabulary words which were included in the comprehension task was at ceiling before training began, but their knowledge of the meanings of the same words was very poor. The two participants showed different patterns of generalization to improved reading comprehension and illustrate two potential mechanisms for the link between oral vocabulary and reading comprehension.

Participant 1 appears to have made broad gains in comprehension skill - she made only small improvements in knowledge of the trained words included in the experimenter-designed task, but made significant improvements on both the experimenter-designed and standardized reading comprehension tasks. This finding suggests that vocabulary training may have led to improvements in more general oral language or metalinguistic skill, which could in turn have influenced reading comprehension (Beck et al., 1982).

However, a second child, Participant 2, made significant improvements on the trained words which were included in the experimenter-designed comprehension task, and made comprehension gains on that task, but did not make any gains on the standardized task. Her results support the hypothesis of a direct link between knowledge of word meanings and comprehension of text containing those word meanings.

We hoped to further explore these two hypotheses by attempting to determine the extent to which knowledge of a word's meaning predicted ability to answer the comprehension questions, but unfortunately we found that some questions could be answered without knowing the word's meaning. After eliminating these questions, the number of items was too small to allow for appropriate statistical analyses. Future studies should explore the nature of the link between vocabulary and reading comprehension in greater detail so that questions about the exact nature of the link between vocabulary and reading comprehension can be explored. It is, of course, very likely that vocabulary could be related to reading comprehension in more than one way – for example, through specific word meanings, but also through metalinguistic skills which may be associated with the ability to learn new words effectively.

Overall, our findings do support a direct link between vocabulary and reading comprehension, and suggest that vocabulary is necessary, but not sufficient, for reading comprehension. Our findings also suggest that reading comprehension interventions should be tailored to the skill profiles of individual poor comprehenders (Cain & Oakhill, 2006). Alternatively, carefully designed multi-component interventions targeted at a number of different oral language and reading comprehension skills may be necessary, such as the programs evaluated by Clarke et al. (2010). Such multi-component interventions may be effective for a wide variety of children, though it remains to be seen whether a multi-component training program would be as effective as individually tailored programs. Our results indicate that oral vocabulary training should form a part of any multi-component intervention program, but further research at both the group and individual level is required to determine the optimum content of such programs.

In conclusion, although previous studies have shown that not all poor comprehenders have vocabulary difficulties (Cain & Oakhill, 2006; Nation et al., 2004), this study is the first to look at this individual variation in response to oral vocabulary training. Of course, many questions remain; there are numerous factors which can influence a child's improvement, or lack thereof, as a result of a training study, and it is not possible to take all of these factors into account simultaneously. In addition, the interaction between such factors may also be of importance and we were not able to explore this in the current study. Further, we were only able to look at the causal role of vocabulary skills at a single point in time.

Nonetheless, our research highlights the need to examine the results of intervention programs at the individual level. While individual variation is often seen merely as "noise" at the group level, it cannot be ignored if we wish to truly understand the multiple causes of reading comprehension difficulties, and if we wish to design intervention programs which are effective for the largest number of children.

Acknowledgements:

We wish to thank the staff, parents and students of the participating school, and Marielle Smith and Yvette Kezilas for test scoring.

## References


Beck, I., McKeown, M., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. New York, NY: Guilford.

Beck, I., McKeown, M., & Kucan, L. (2008). *Creating robust vocabulary: Frequently asked questions and extended examples*. New York, NY: Guilford.

Beck, I., Perfetti, C., & McKeown, M. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology, 74,* 506-521. doi:10.1037/0022-0663.74.4.506

Begliner, L. J., Gaydos, B., Tangphao-Daniels, O., Duff, K., Kareken, D. A., Crawford, J.,…Siemers, E. R. (2005). Practice effects and the use of alternate forms in serial neuropsychological testing. *Archives of Clinical Neuropsychology, 20*, 517-529. doi: 10.1016/j.acn.2004.12.003

Boutron, I., Moher, D., Altman, D. G., Schulz, K. F., & Ravaud, P. (2008). Methods and processes of the CONSORT group: Example of an extension for trials assessing nonpharmacologic treatment. *Annals of Internal Medicine, 148*, W60-W66. doi:10.7326/0003-4819-148-4-200802190-00008-w1

Cain, K., & Oakhill, J. (2006). Profiles of children with specific reading comprehension difficulties. *British Journal of Educational Psychology, 76,* 683-696. doi:10.1348/000709905X67610

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

Catts, H. W., Adlof, S. M., & Weismer, S. E. (2006). Language deficits in poor comprehenders: A case for the Simple View of Reading. *Journal of Language, Speech and Hearing Research, 49*, 278-293. doi:10.1044/1092-4388(2006/023)

Clarke, P. J., Henderson, L. M., & Truelove, E. (2010). The poor comprehender profile: Understanding and supporting individuals who have difficulties extracting meaning from text. In J. Holmes (Ed.), *Advances in Child Development and Behaviour* (Vol. 39, pp. 79-129). New York, Boston: Academic Press

Clarke, P. J., Snowling, M. J., Truelove, E., & Hulme, C. (2010). Ameliorating children's reading-comprehension difficulties: a randomized controlled trial. *Psychological Science, 21*, 1106-1116. doi:10.1177/0956797610375449

Colenbrander, D. (2015). *Understanding the role of oral vocabulary in reading comprehension difficulties.* (Doctoral dissertation, Macquarie University, Sydney, Australia).

Colenbrander, D., Kohnen, S., Smith-Lock, K., & Nickels, L. (2015). *Understanding the link between oral vocabulary and reading comprehension - a randomized controlled trial*. Manuscript in preparation.

Colenbrander, D., Nickels, L., Smith-Lock, K., & Kohnen, S. (2015). *Exploring the relationship between vocabulary skills and reading comprehension - a look at individual differences*. Manuscript in preparation.

Cornoldi, C., De Beni, R., & Pazzaglia, F. (1996). Profiles of reading comprehension difficulties: An analysis of single cases. In C. Cornoldi & J. Oakhill (Eds.), *Reading Comprehension Difficulties: Processes and Intervention* (pp. 113-136). Mahwah, NJ: Lawrence Erlbaum Associates.

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2008). Developing and evaluating complex interventions: the new Medical Research Council guidance. *British Medical Journal, 337*, a1655. doi:10.1136/bmj.a1655

Crawford, J. R., Garthwaite, P. H., & Porter, S. (2010). Point and interval estimates of effect sizes for the case-controls design in neuropsychology: Rationale, methods, implementations, and proposed reporting standards. *Cognitive Neuropsychology, 27*, 245-260. doi:10.1080/02643294.2010.513967

Crawford, J. R., & Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *The Clinical Neuropsychologist (Neuropsychology, Development and Cognition: Section D), 12*, 482-486. doi:10.1076/clin.12.4.482.7241

Cumming, G. (2011). ESCI, Exploratory Software for Confidence Intervals. Retrieved from http://www.latrobe.edu.au/psy/research/projects/esci

Davis, C. J. (2005). N-Watch: A program for deriving neighbourhood size and other psycholinguistic statistics. *Behavior Research Methods, 37*, 65-70. doi:10.3758/BF03206399

Hegde, M. N. (2007). A methodological review of randomized clinical trials. *Communicative Disorders Review, 1*, 17-38.

Howard, D. (1986). Beyond randomised controlled trials: The case for effective case studies of the effects of treatment in aphasia. *British Journal of Disorders of Communication, 21*, 89-102. doi:10.3109/13682828609018546

Howard, D., Best, W., & Nickels, L. (in press). Optimising the design of intervention studies: Critiques and ways forward. *Aphasiology*.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, 44*, 978-990. doi:10.3758/s13428-012-0210-4

Marinus, E., Kohnen, S., & McArthur, G. (2013). Australian comparison data for the Test of Word Reading Efficiency (TOWRE). *Australian Journal of Learning Difficulties, 18*, 199-212. doi:10.1080/19404158.2013.852981

McKeown, M., Beck, I., Omanson, R., & Perfetti, C. (1983). The effects of long-term vocabulary instruction on reading comprehension: A replication. *Journal of Literacy Research, 15*, 3-18. doi:10.1080/10862968309547474

McKeown, M., Beck, I., Omanson, R., & Pople, M. (1985). Some effects of the nature and frequency of vocabulary instruction on the knowledge and use of words. *Reading Research Quarterly, 20*, 522-535. doi:10.2307/747940

Nation, K., Clarke, P., Marshall, C. M., & Durand, M. (2004). Hidden language impairments in children: Parallels between poor reading comprehension and specific language impairment? *Journal of Speech, Language and Hearing Research, 47*, 199-211. doi:10.1044/1092-4388(2004/017)

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

Nation, K., Snowling, M. J., & 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. *Advances in Speech-Language Pathology, 9*, 131-139. doi:10.1080/14417040601145166

Neale, M. D. (1999). *The Neale Analysis of Reading Ability Third Edition*. Melbourne, Australia: ACER.

Nickels, L., Kohnen, S., & Biedermann, B. (2011). An untapped resource: treatment as a tool for revealing the nature of cognitive processes. *Cognitive Neuropsychology, 27*, 539-562. doi:10.1080/02643294.2011.609811

Oakhill, J., & Cain, K. (2007). Reading comprehension difficulties: Correlates, causes and consequences. In K. Cain & J. Oakhill (Eds.), *Children's comprehension problems in oral and written language: A cognitive perspective* (pp. 41-75). New York, NY: Guilford Press.

Snowling, M. J., Stothard, S. E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., . . . Hulme, C. (2012). *York Assessment of Reading for Comprehension: Passage Reading. Australian Edition.* London, UK: GL Assessment.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of Word Reading Efficiency*. Austin, TX: Pro-Ed.

Yuill, N., & Oakhill, J. (1991). *Children's problems in text comprehension: An experimental investigation*. Cambridge, UK: Cambridge.

## Appendix A

Trained and Untrained Words

| | |
|---|---|
| abrupt | anxious |
| abundance | conform |
| ally | conquer |
| assist | expedition |
| amateur | fatigue |
| benefactor | fury |
| brute | harsh |
| coerce | immense |
| conceal | impair |
| confide | invader |
| dubious | mingle |
| dispute | modest |
| enthusiasm | observe |
| envy | pursue |
| foe | rebel |
| frantic | regret |
| gradual | unique |
| idle | vacant |
| lunge | vandal |
| mock | vicious |
| naïve | |
| nimble | |
| offend | |
| petty | |
| ploy | |
| precise | |
| prevent | |
| request | |
| scarce | |
| scorn | |
| seize | |
| serene | |
| shrewd | |
| shudder | |
| swindle | |
| temptation | |
| tense | |
| triumph | |
| tyrant | |
| urge | |
| vague | |

## Appendix B

Structure of a Typical Lesson

1. The experimenter said a word aloud and wrote the word on the whiteboard.

   *“Prevent – the first word for today is prevent”.*

2. The experimenter gave a simple, “child friendly” definition of the word in terms that the children could easily understand. Children wrote the word and its definition in their notebooks. Notebooks were then put away.

   *“Prevent means to stop something from happening.”*

3. The experimenter gave 3 examples of the word in context.

   *“For example, you could say, ‘The house was very old and would need lots of work to* prevent *it falling down.’ Or you could say, ‘No matter how careful you are, sometimes accidents happen, and there's nothing you can do to* prevent *them.*’ *Here’s another example: ‘If you wear sunscreen, it will* [pause for children’s replies] *sunburn’.”*

4. Children then heard six sentences and had to decide whether or not the sentence was a good example of the word’s meaning and explain their decision.

   *“Firemen do some backburning in the bush so that a fire does not spread.”* (Good example for *prevent*)

   *“An old lady spills all her groceries on the ground. You run over to help her pick them up.”* (Bad example for *prevent*)

5. Children were then asked to think of examples relating the meaning of the word to their own experience.

   *“Can you think of a time when someone prevented you from doing something? What happened?”*

6. The above steps were repeated for two further words.

7. Children then played a word association game in which they heard a series of words and had to decide which of the day's words best "went with" the words they heard. Flashcards with the written forms of the words were used during this task.

Throughout the program, the experimenter was careful to ensure that, as far as possible, all children had equal opportunities to answer questions and participate. A reward system was utilized whereby children could accumulate stickers for effort and good behaviour. These stickers could later be traded for a small prize.

## Appendix C

Lesson Transcripts

Excerpts come from Week 7, Lesson 2.

T = Trained

P1, P2, P3, P4, P5, P6, P7, P8 = Participants (two participants were absent from Group 2 during this lesson)

**Group 1**

T: The first word for today is gradual.

P1: Is that where it's slow, but you eventually get there?

T: Yeah, very good. It's when something happens slowly over a long time. So for example, you might say global warming causes a *gradual* rise in sea levels. It happens very slowly over a long period of time. Or you could say the piece of ice melted *gradually*. Or you could say, in autumn, the leaves change colour…(pauses for group to respond) gradually. OK, now we will play a game. If you think what I'm saying is gradual, then you say "gradual". If you think it's not gradual, then you don't say anything. OK, so…do you think this is gradual or not? Trying to lose weight by doing more exercise…

P1: I think it's gradual because it doesn't just happen, it takes a long time.

T: Exactly. So you can't do more exercise and then the next day you're 30 kilograms lighter, that doesn't happen. If it's gradual it happens slowly over time. Very good. OK…learning a foreign language…

P2: That is not gradual because it'll take up…well…oh, it is gradual because it'll take a long time and it won't just happen like that.

T: Yes, right. It won't just happen overnight, it happens over a long time, so it's gradual. Alright…what about breaking a glass?

P3: It isn't gradual because it doesn't happen slowly…it happens like…(pauses)

T: Yes, it doesn't happen slowly, it happens…(waits for response but child does not give one)…quickly. That's right. OK, next one…a plant growing.

P4: It's gradual because it takes a long time to grow.

T: Yes, because plants don't just go "boom" and you've got a fully grown tree, it takes a long time. OK, what about switching off the light?

P1: I don't think it's gradual, because you can just switch it off and it's gone.

T: Exactly, it's not gradual, it's pretty sudden. So, here's a demonstration (switches room light on and off). Alright, how about saving money.

P3: Yeah, it is gradual, because it takes a long time.

T: Very good, exactly right. OK, so I want you to have a think about something else you can think of that happens gradually.

P1: A snail in slow motion

T: Yes, a slow snail moves gradually, good one.

P2: Your fingernails.

T: Yes, your fingernails grow gradually.

P3: You could grow!

T: Yes, you grow gradually. Were you going to say that too P4? (She nods). Yes, very good. So any growing process is actually gradual. Good work everyone.

**Group 2**

T: So the first word is gradual…gradual. Yes, P5?

P5: To um…to slowly and not very quickly…

T: Yes, that's part of it

P6: To slowly and eventually…something

T: Yes, you're both on the right track. Eventually is very similar to gradual…but if we want a meaning that covers all the meanings of gradual, we would say that something is gradual if it happens slowly over a long time. So for example, you might say, global warming will cause a *gradual* rise in water levels. Or you might say the piece of ice melted *gradually*, or you could say, in autumn, the leaves change colour *gradually*. These are all examples.

P6: So, in tap we gradually did our dance…we gradually…we gradually perfected our dance.

T: Yes, good. It’s hard to do a dance gradually, unless it’s a very slow dance…but learning a dance, that can happen gradually.

P5: Um if someone’s sick, you’d usually gradually get better.

T: Yes, excellent example, very good!

P6: Yeah…my brother is gradually getting better…his arm.

T: Yes, good. So your brother is gradually healing. And we’ve got some more examples. Some of these things are gradual and some of these things are not gradual. So I just need you to put your hands on your head and say “gradual” if you think it’s gradual, otherwise you keep your hands on the table, OK? Here we go – trying to lose weight by doing more exercise.

All: Gradual!

P5: Because you can’t just go for a run and then lose 6 kilograms.

T: Yeah, I think people wish you could lose weight that quickly, but it’s a gradual process, it happens slowly over time. Or if it happens quickly then it’s usually very unhealthy.

OK…what about learning a foreign language?

All: Gradual!

P6: Because you can’t just say “Ah, now I can speak French”, you’ve got to gradually get used to it.

T: Yeah, you’ve got to learn it slowly over time. It would be nice if you could just go “poof” now I’m speaking Russian, but it doesn’t work like that, it’s a gradual process. OK, next one…dropping a glass and breaking it.

P7: Not gradual, because normally when a glass drops, it doesn’t go slowly, it normally just goes “crash” to the ground.

T: Yes, that's right.

P5: It's kind of like gravity.

T: Yeah, it just kind of falls to the ground and there's not much you can do about it, it happens really fast.

P5: Unless you lunge to seize it![15]

T: Yes…but you have to be careful seizing glass, because you don't want to crack it.

P6: But what if the glass doesn't break, if you drop it a few times, is that gradual?

T: Well the moment of breaking is always sudden, so if you drop it a few times and it doesn't break, the last time you drop it is when it finally cracks, that's sudden. So what about a plant growing?

All: Gradual!

P5: Because you can't just get a plant and it's fully grown.

T: Yes, that's very good.

P6: But what if you give it some plant food and it makes it grow faster?

T: Well even if it grows faster, it's still gradual – it doesn't just go "whoosh" and there's a fully-grown tree. When things grow, it's always gradual. OK, next one…switching off the light.

All: Not gradual!

P7: Because when you're switching off the light it doesn't take a slow amount of time, because it's like flicking it off.

T: Yes

P6: But it could be gradual!

T: Why do you think that?

P6: Because some lights, you can turn them off gradually.

[15] Lunge and seize were words trained in a previous week

T: Well that's a dimmer switch, when you gradually turn down the light slowly, but if I do this (switches off the light), it's off and it happens straight away. OK, so last one. Saving money…is that gradual?

P5: Gradual…because like although people can win the Lotto and stuff, you have to…people have to work to do these types of things and you don't get paid everything in one day.

T: Yeah, you get paid usually once every two weeks, or once a month, and then you've got to put little bits of money away over time, so it's a gradual process.

P6: For holidays, me and my mum are going to Disneyland, and we have to gradually save up money to go.

P8: My dog moves gradually.

T: Well that's kind of different, so slow and gradual are not always the same thing. When you're talking about people moving, you don't usually say they're moving gradually. So you might say your dog grew gradually, but you wouldn't usually say that a dog moves gradually, when something is gradual it's more like a slow process that's happening.

# Chapter Six:
# Discussion and conclusions

# Discussion

While the role of oral vocabulary in reading comprehension has been the focus of much research (Cain & Oakhill, in press; Nation & Snowling, 1998, 1999; Protopapas, Mouzaki, Sideridis, Kotsolakou, & Simos, 2013; Ricketts, Bishop, & Nation, 2008), relatively few studies have specifically explored the causal role of oral vocabulary skills, and still fewer have considered the importance of individual differences in clarifying the nature of the links between the two skills. The studies presented in this thesis aimed to address this by exploring issues related to the assessment of comprehension and vocabulary, determining the effectiveness of oral vocabulary training on the comprehension skills of poor comprehenders, and exploring individual differences in vocabulary skills and response to vocabulary intervention.

## Measurement of comprehension and the poor comprehender profile

As discussed in Chapter One, the poor comprehender profile has been operationalized in many different ways, and many different assessments of reading comprehension have been used across studies. However, reading comprehension assessments are not interchangeable – it is possible for a child to be diagnosed with comprehension difficulties on one test but not another (Keenan & Meenan, 2014). This has implications for the comparability of results across studies and for the diagnosis of reading comprehension difficulties in clinical and educational contexts (Cutting & Scarborough, 2006; Eason, Goldberg, Young, Geist, & Cutting, 2012; Keenan, Betjemann, & Olson, 2008; Keenan & Meenan, 2014). Therefore, in Chapter Two, differences between a widely used assessment of reading comprehension, the Neale Analysis of Reading (NARA; Neale, 1999) and a more recently developed assessment, the York Assessment of Reading for Comprehension (YARC; Snowling et al., 2012), were analyzed.

We found that approximately one third of a sample of primary school children received different diagnoses on the two tests, even though both assessments are relatively similar in format (children read passages and answer questions about the passages). In particular, the NARA tended to diagnose more children as having below average reading comprehension than the YARC. This is an important finding given that in some educational and research contexts, the YARC is being considered as a replacement for the NARA (GL Assessment, 2014; Howe, 2013; Ricketts, 2014).

Our results indicated that the NARA was somewhat more reliant on children's reading abilities, and was more difficult for good readers, than the YARC. This was because good readers answered a greater number of questions on the NARA. The CohMetrix (McNamara, Louwerse, Cai, & Graesser, 2011) analysis also demonstrated that the NARA passages were more challenging, having lower levels of referential cohesion and containing more difficult vocabulary items. This meant that children answered questions from a wider range of difficulty levels on the NARA, which might have made the test more sensitive to subtle reading comprehension deficits.

Further research exploring the predictive abilities of the NARA and the YARC in educational contexts is required in order to determine whether the YARC is under-diagnosing comprehension difficulties, or whether the NARA is over-diagnosing. However, many more of our participants were classified as poor comprehenders on the NARA than on the YARC (nine participants were classified as poor comprehenders on the NARA but not on the YARC, two on the YARC but not on the NARA, and two on both tests), and all of these participants had some level of oral language deficit (see below). This makes it more likely that the NARA was indeed more sensitive (and not oversensitive) and was picking up on oral language impairments causing subtle comprehension difficulties.

Previous studies have recommended that the diagnosis of reading comprehension difficulties be made on the basis of more than one assessment (Bowyer-Crane & Snowling, 2005; Cain & Oakhill, 2006a; Keenan & Meenan, 2014). Our results support this recommendation. However, these results do not suggest that children should only be considered poor comprehenders when they are diagnosed as such on two or more comprehension assessments. Rather, our findings suggest that the use of multiple assessments should be considered as additional sources of information about a child's skill reading comprehension skill, and these assessments should always be interpreted in relation to knowledge about the content and format of the test. The results of the study that examined the skill profiles of the sample of poor comprehenders (Chapter 3) suggest that reading comprehension assessment should be followed up by oral language assessment when a reading comprehension deficit is suspected. The frequency of vocabulary difficulties within our sample suggests that assessment of vocabulary skills should form a central part of any such oral language assessment.

**The relationship between oral vocabulary and reading comprehension**

This thesis tested the predictions of two widely cited theoretical views. The first of these views, the Simple View of Reading (Gough & Tunmer, 1986; Hoover & Gough, 1990), states that reading comprehension is made up of two main areas of skill, reading and oral language. This view predicts that poor comprehenders' reading comprehension difficulties are due to poor oral language skills, given that they have intact word and text reading skills. Under this view, poor oral vocabulary skills may be a sole or contributory factor at the root of the poor oral language skills of poor comprehenders.

Another theoretical view, the Lexical Quality Hypothesis (Perfetti, 2007; Perfetti & Hart, 2001, 2002), proposes that vocabulary skill is itself made up of multiple components. In this hypothesis, word knowledge is made up of three main constituents, phonological knowledge, orthographic knowledge, and semantic knowledge. Therefore, if it can be demonstrated that poor comprehenders have deficits in the oral language (semantic or phonological) components of vocabulary, but do not have deficits in the written (orthographic) components of vocabulary, this would support the proposed distinction between reading and oral language put forward by the Simple View.

Previous research has shown that at the group level, poor comprehenders tend to have irregular word reading deficits (Nation & Snowling, 1998; Ricketts, Nation, & Bishop, 2007). Such deficits have been claimed to be due to poor comprehenders' semantic weaknesses (Nation & Snowling, 1998; Ricketts et al., 2007). This implies that semantic knowledge is necessary for correct reading of irregular words, consistent with the Triangle Model of reading (Plaut, McClelland, Seidenberg, & Patterson, 1996). Under this view, semantic knowledge forms part of word reading *and* oral language skills, which challenges the separation between reading and oral language skills proposed by the Simple View of reading (Protopapas et al., 2013). Consistent with this, some studies of the role of vocabulary in the Simple View have found that vocabulary contributes variance to both the reading and oral language components (Ouelette & Beers, 2010; Protopapas et al., 2013).

The Dual Route Model of reading (Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001) sees the process of reading differently. This model predicts that reading takes place via two routes, a sub-lexical route which converts letters into sounds using of grapheme-phoneme correspondence rules, and a lexical route in which words are recognized by sight. An irregularly spelled word can be read via the lexical route, either by accessing the word via semantics, or by accessing it directly via orthography.

However, in fact, both the Triangle Model and the Dual Route Model predict that it is possible to read irregular words correctly when semantics is compromised. While the Triangle Model argues that irregular word reading in the presence of poor semantic knowledge can be achieved by relying to a greater degree on connections between orthographic and phonological knowledge (Harm & Seidenberg, 2004), the Dual Route model states that reading of irregular words can be achieved by directly accessing orthographic forms of the words (Coltheart et al., 1993; Coltheart et al., 2001). Therefore, a poor comprehender with poor semantic skills could nonetheless have age-appropriate irregular word reading skills. If such poor comprehenders exist, this would provide further evidence that reading abilities and oral language abilities are separable (though it is very likely that they are interconnected in most children).

**The nature of vocabulary difficulties in poor comprehenders.**

The results of the studies reported in this thesis supported the predictions of both the Simple View and the Lexical Quality Hypothesis. Firstly, in terms of reading and phonological skills, not one single poor comprehender participant scored below controls on a task of nonword reading. Poor comprehenders also had strengths in phonological skills as assessed by a task of nonword recall, and scored as well as controls on tasks of real word and nonword reading fluency. However, even though all poor comprehenders scored within the normal range on irregular word reading, two children performed more poorly than controls on this task. Both of these children also had semantic deficits.

Nonetheless, strikingly, there were 8 participants who scored poorly on tasks tapping semantics, but as well as controls on the irregular word reading task. This provides evidence that poor semantic skills do not necessarily result in poor irregular word reading. This finding is consistent with both the Dual Route Model of reading (Coltheart et al., 1993; Coltheart et al., 2001) and the Triangle model (Plaut et al., 1996), as described above. Intact irregular word reading in the context of poor semantic skills is also consistent with the Simple View of Reading (Gough & Tunmer, 1986; Hoover & Gough, 1990), suggesting that reading-specific skills (i.e. orthographic knowledge) and oral language skills are separable even in the case of irregular word reading.

The finding that poor comprehenders (as a group) had relatively good irregular word reading skills contrasts with the findings of previous research (Nation & Snowling, 1998, 1999; Ricketts et al., 2007). However, unlike previous research, we selected poor comprehenders who had age appropriate irregular word reading skills as well as age-appropriate nonword and text reading skills. The assumption of a close link between irregular word reading difficulties and poor semantic skills leads to the prediction that selecting poor comprehenders in this way should exclude children who have vocabulary difficulties, but this was not the case.

In fact, while irregular word reading deficits were rare within the sample, vocabulary deficits were common, and these primarily took the form of semantic difficulties. More than two thirds of the sample (77%) had some level of semantic deficit. Poor comprehenders' difficulties with semantics were not just restricted to tasks of lexical semantics – nearly half of the sample performed more slowly or less accurately than controls on a task of conceptual semantics (picture-picture matching).

An interesting finding was the fact that, as well as knowing fewer meanings of experimental words than controls, five children (38% of the sample) also had difficulty recognizing the phonological forms of the same words in an auditory lexical decision task. Although we cannot rule out the fact that semantic skills may play a role in success on this task (see Betjemann & Keenan, 2008), this finding suggests that some poor comprehenders may have difficulties with lexical phonology, even though they do not have difficulties with broader phonological skills.

Further support for the Simple View of Reading comes from the fact that although there were three participants who did not display any semantic difficulties, each and every participant in the study had deficits on at least one oral language task. The three children who did not have semantics deficits scored below controls on tasks of syntax and listening comprehension. Thus, overall, poor comprehenders had subtle oral language weaknesses but intact reading abilities.

As well as being consistent with the Simple View, our findings are consistent with The Lexical Quality Hypothesis (Perfetti, 2007; Perfetti & Hart, 2001, 2002). The Lexical Quality Hypothesis does not suggest that *all* reading comprehension problems stem from lexical weaknesses, but rather claims that when children have lexical weaknesses, these should result in reading comprehension problems. In line with this, the majority of poor comprehenders had lexical level deficits, and these deficits were primarily semantic in nature – in other words, these deficits were present in the oral language aspects of vocabulary.

**The causal role of oral vocabulary skills in reading comprehension.**

Crucially, the studies in this thesis not only explored the nature of oral vocabulary deficits in poor comprehenders, but also explored the causal role played by these deficits. The results of a randomized controlled trial of oral vocabulary training for reading comprehension difficulties (reported in Chapter Four) showed that the training was successful in improving all three aspects of word knowledge of the trained words – semantic, orthographic and phonological – and resulted in improvements on an experimenter-designed task of reading comprehension. Reading comprehension gains could not be accounted for by gains in ability to read the trained words, given that children were at or near ceiling in their ability to read the trained words in the experimenter-designed reading comprehension task before training began. Consistent with both the Simple View and the Lexical Quality Hypothesis, this provided evidence for a direct link between *oral* vocabulary skills and reading comprehension.

We attempted to provide further evidence for a direct, causal, link by determining whether improvements in the knowledge of word meanings led directly to improvements on reading comprehension questions requiring knowledge of the trained words. Unfortunately, however, this was not possible because post-hoc analysis revealed that some of the comprehension questions could be answered without knowledge of the relevant word meanings.

However, analysis of the results of the 10 trained participants (from both the initial training and wait-list control group) allowed a more detailed exploration of the relationship between reading comprehension and oral vocabulary. Each and every child in the training group showed significant improvements in their ability to provide verbal definitions for the trained words. In other words, the oral vocabulary training program was very effective in improving children's semantic knowledge of the trained words.

Nevertheless, although all children made improvements in semantic knowledge, only four children made significant or near-significant gains on the experimenter-designed reading comprehension task. A further analysis demonstrated that these four children all scored more poorly than controls at baseline on the definition production and definition recognition tasks. Overall, the size of comprehension gains was negatively correlated with baseline scores on the definition recognition task. In other words, results showed that vocabulary training was of most benefit to children who had poor vocabulary skills before training began. Together, these results imply that poor vocabulary skills were playing a causal role in the reading comprehension difficulties of some, but not all, of the poor comprehender participants.

The individual-level analyses allowed further exploration of the mechanisms by which vocabulary training led to comprehension improvements. The results of the two participants who made significant improvements on the experimenter-designed comprehension task were explored in detail. For one child, results appeared to indicate general improvements in reading comprehension as a result of the training program. She made large gains in her ability to define the trained words and made improvements not only on the experimenter-designed comprehension task but also on the standardized reading comprehension task.

The second child also made large gains on the definition production task, but her comprehension gains were specific to the task containing the trained words. Therefore, for this participant, comprehension gains appeared to be a direct result of improved knowledge of the word meanings in the comprehension task.

More research into the mechanisms of the link between vocabulary and reading comprehension gains is required, including training studies using more sensitive measures of the effects of item-level improvements in word knowledge on comprehension of text containing the words. Nonetheless, taking both group and individual-level findings into account, the results provide evidence for the existence of a direct causal link between oral vocabulary and reading comprehension.

However, the variability of comprehension improvements at the individual level demonstrate that vocabulary skills are necessary, but not sufficient, for reading comprehension success. Several children in the sample did not make gains on reading comprehension despite making large and significant vocabulary gains. Given the results of the profiling study in Chapter Three, it is possible that the comprehension difficulties of these children were caused by oral language difficulties above the word level, although further research would be required to determine this.

**Practical, theoretical and methodological implications**

The results of the studies in this thesis point towards a number of practical and theoretical implications. Firstly, the findings reported in Chapter Two reinforce the importance of carefully selecting and interpreting reading comprehension assessments when attempting to identify children with specific reading comprehension difficulties.

These findings also imply that reading comprehension research should be interpreted in light of the comprehension assessment used to identify participants. Our results suggest that for poor comprehenders in particular, the NARA may be a more sensitive test, but future research into differences in the content of different tests is warranted, as well as research exploring the characteristics of samples diagnosed with poor comprehension on different tests. Such research would allow more accurate interpretation of existing reading comprehension tests and may provide guidance for the development of better assessments in the future.

Secondly, the findings reported in this thesis indicate that children with poor reading comprehension and good irregular word reading do exist. In fact, these children made up nearly two thirds of the sample. Given the findings of previous studies, (e.g., Keenan & Betjemann, 2007; Nation & Snowling, 1998; Ricketts et al., 2007), we do not deny that semantics may be accessed during the word reading process, and it is very likely that there are children for whom semantic deficits cause irregular word reading deficits (Nation & Snowling, 1997, 1998; Ricketts et al., 2007). However, our results show that poor irregular word reading skills should not be considered a feature of the poor comprehender profile, and irregular word reading deficits should not be equated with semantic deficits. Rather, if the intention of a research study is to identify children who have reading comprehension problems in the absence of any reading deficits, then that study should exclude participants with poor irregular word reading as well as participants with poor nonword or regular word reading.

Future studies should determine the incidence of irregular word reading difficulties in a large, randomly selected sample of children in order to determine how frequently irregular word reading difficulties are associated with semantic difficulties. In addition, future research should look more closely into the orthographic and semantic skills of poor comprehenders. In particular, research could use online paradigms (such as eye-tracking) to explore how poor comprehenders use orthographic and semantic information during the process of reading comprehension.

Thirdly, this thesis was the first to systematically assess the effects of vocabulary training on the semantic, orthographic and phonological aspects of word knowledge. Consequently, this research has provided support for the hypothesis that poor comprehenders' vocabulary difficulties are primarily oral-language based, and in particular, that the semantic aspects of vocabulary knowledge play a causal role in the comprehension difficulties of some poor comprehenders. Along with the finding that poor comprehenders' orthographic skills are generally intact, these results provide support for both the Simple View of Reading and the Lexical Quality Hypothesis. Further support for the Simple View comes from the fact that the whole sample of poor comprehenders had some degree of oral language difficulty at the individual level.

Finally, the findings of the training studies indicate that although many poor comprehenders have vocabulary difficulties, not all do, and vocabulary training was most effective for the children who did have vocabulary difficulties. This indicates that remediation programs for reading comprehension should be tailored to the skill profiles of individual children (Cain & Oakhill, 2006b; Pimperton & Nation, 2014). Some children may have multiple oral language deficits, and for these children, multi-component oral language programs such as that administered by Clarke et al. (2010) in their large randomized controlled trial, may be the best solution.

More research is required in order to establish and compare the effectiveness of both individually-tailored and multi-component training programs. Research should focus on not only whether these programs work, but why and how they work, using outcome measures designed to target particular theoretical questions. Crucially, research should not only assess outcomes at the group level, but should also track improvements at the individual level.

Given the complex nature of reading comprehension, it is important that a variety of research methods are utilized when attempting to identify the causes of reading comprehension difficulties. For example, while training studies can be used to explore the causal role of a specific skill at one point in time, longitudinal studies can be used to determine the influence of a putative causal factor over time. In fact, training and longitudinal methods could be combined to powerful effect. For example, the causal role of oral vocabulary skills in reading comprehension could be assessed by following an unselected sample of children across a period of several years, in order to determine whether classroom-based oral language teaching could reduce the incidence of later reading comprehension difficulties. As well as these behavioural methods, computational methods could be used to model complex relationships between causal factors, and the results of such modeling could in turn provide guidance for questions to be explored in future behavioural research. Thus, the combination of information from different research designs should allow us to converge on a more detailed explanation of the causes of reading comprehension difficulties.

**Conclusion**

Reading comprehension is an incredibly complex construct, made up of many sub-skills and processes. Not only does it comprise multiple components, but these components are likely to interact with each other, causing even greater complexity. In addition, different skills may play greater or smaller roles in the reading comprehension process over time (Oakhill & Cain, 2012; Ouelette & Beers, 2010), and a skill which initially causes the poor development of reading comprehension may not be the same skill which is limiting a child's reading comprehension at a later stage (Castles, Kohnen, Nickels, & Brock, 2014; Oakhill & Cain, 2007). However, if we wish to understand the reading comprehension process, we must attempt to disentangle the contributions of these component skills and determine the roles they are playing in different individuals and at different stages of development.

The results of the studies in this thesis suggest that vocabulary knowledge is necessary, but not sufficient, for reading comprehension. If an individual does not know the meanings of the words in a text, he or she cannot understand it. However, even if a person does know the meanings of the words in a text, there may be other reasons why reading comprehension fails. A person may have oral language difficulties beyond the word level, such as syntactic difficulties, or difficulties with discourse-level skills, such as inferencing or comprehension monitoring (Cain & Oakhill, 1999; Oakhill & Cain, 2007; Silva & Cain, 2014). Nonetheless, the results presented here have provided evidence that a direct causal relationship between reading comprehension and vocabulary does exist, and specifically, that some poor comprehenders have semantic difficulties which appear to play a role in causing their comprehension difficulties. It is hoped that these findings will lead to the development of more effective remediation programs for reading comprehension difficulties, and will stimulate further research into the multiple causes of reading comprehension difficulties.

## References

GL Assessment. (2014). York Assessment of Reading for Comprehension: Improving literacy outcomes (City of Edinburgh Council). Retrieved 25 August 2014, from http://www.gl-assessment.co.uk/research-and-articles/york-assessment-reading-comprehension-improving-literacy-outcomes-city

Betjemann, R. S., & Keenan, J. M. (2008). Phonological and semantic priming in children with reading disability. *Child Development, 79*, 1086-1102. doi:10.1111/j.1467-8624.2008.01177.x

Bowyer-Crane, C., & Snowling, M. J. (2005). Assessing children's inference generation: what do tests of reading comprehension measure? *British Journal of Educational Psychology, 75,* 189-201. doi:10.1348/000709904X22674

Cain, K., & Oakhill, J. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing, 11,* 489-503. doi:10.1023/A:1008084120205

Cain, K., & Oakhill, J. (2006a). Assessment matters: issues in the measurement of reading comprehension. *British Journal of Educational Psychology, 76*, 697-708. doi:10.1348/000709905X69807

Cain, K., & Oakhill, J. (2006b). Profiles of children with specific reading comprehension difficulties. *British Journal of Educational Psychology, 76*, 683-696. doi:10.1348/000709905X67610

Cain, K., & Oakhill, J. (in press). Reading comprehension and vocabulary: Is vocabulary more important for some aspects of comprehension? *L'Annee Psychologique*.

Castles, A., Kohnen, S., Nickels, L., & Brock, J. (2014). Developmental disorders: what can be learned from cognitive neuropsychology? *Philosophical Transactions of the Royal Society B, 369*, 20130407. doi:10.1098/rstb.2013.0407

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed processing approaches. *Psychological Review, 100,* 589-608. doi:10.1037/0033-295X.100.4.589

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

Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading, 10*, 277-299. doi:10.1207/s1532799xssr1003_5

Eason, S. H., Goldberg, L. F., Young, K. M., Geist, M. C., & Cutting, L. E. (2012). Reader-text interactions: How differential text and question types influence cognitive skills needed for reading comprehension. *Journal of Educational Psychology, 104*, 515-528. doi:10.1037/a0027182

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*, 6-10. doi:10.1177/074193258600700104

Harm, M., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. P*sychological Review, 111,* 662–720. doi: 10.1037/0033-295X.111.3.662

Hoover, W. A., & Gough, P. B. (1990). The Simple View of Reading. *Reading and Writing, 2*, 127-160. doi:10.1007/BF00401799

Howe, W. (2013). *The use of assessments in the inclusive classroom*. Paper presented at the Pedagogy in Practice, Brisbane, Australia.

Keenan, J. M., & Betjemann, R. S. (2007). Comprehension of single words: The role of semantics in word identification and reading disability. In E. L. Grigorenko & A. J.

Naples (Eds.), *Single-word reading: Behavioural and biological perspectives* (pp. 191-210). New York, NY: Taylor & Francis.

Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*, *12*, 281-300. doi:10.1080/10888430802132279

Keenan, J. M., & Meenan, C. E. (2014). Test differences in diagnosing reading comprehension deficits. *Journal of Learning Disabilities, 47*, 125-135. doi:10.1177/0022219412439326

McNamara, D. S., Louwerse, M. M., Cai, Z., & Graesser, A. C. (2011). Coh-Metrix version 3.0 [Computer software]. Retrieved 20 December 2012 and 22 August 2014, from http://cohmetrix.com

Nation, K., & Snowling, M. J. (1997). Assessing reading difficulties: The validity and utility of current measures of reading skill. *British Journal of Educational Psychology, 67*, 359-370. doi:10.1111/j.2044-8279.1997.tb01250.x

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

Nation, K., & Snowling, M. J. (1999). Developmental differences in sensitivity to semantic relations among good and poor comprehenders: Evidence from semantic priming. *Cognition, 70*, B1-B13. doi:10.1016/S0010-0277(99)00004-9

Neale, M. D. (1999). *The Neale Analysis of Reading Ability Third Edition*. Melbourne, Australia: ACER.

Oakhill, J., & Cain, K. (2007). Reading comprehension difficulties: Correlates, causes and consequences. In K. Cain & J. Oakhill (Eds.), *Children's comprehension problems in oral and written language: A cognitive perspective* (pp. 41-75). New York, NY: Guilford Press.

Oakhill, J., & Cain, K. (2012). The precursors of reading ability in young readers: evidence from a four-year longitudinal study. *Scientific Studies of Reading, 19*, 91-121. doi:10.1080/10888438.2010.529219

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

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*, 357-383. doi:10.1080/10888430701530730

Perfetti, C., & Hart, L. (2001). The lexical basis of comprehension skill. In D. S. Gorfien (Ed.), *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity* (pp. 67-86). Washington, DC: American Psychological Association.

Perfetti, C., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of funtional literacy* (pp. 189-213). Amsterdam/Philadelphia: John Benjamins.

Pimperton, H., & Nation, K. (2014). Poor comprehenders in the classroom: Teacher ratings of behavior in children with poor reading comprehension and its relationship with individual differences in working memory. *Journal of Learning Disabilities, 47*, 199-207. doi:10.1177/0022219412454172

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*, 56-115. doi:10.1037/0033-295X.103.1.56

Protopapas, A., Mouzaki, A., Sideridis, G. D., Kotsolakou, A., & Simos, P. G. (2013). The role of vocabulary in the context of the Simple View of Reading. *Reading and Writing Quarterly, 29*, 168-202. doi:10.1080/10573569.2013.758569

Ricketts, J. (2014). *Reading: Research and assessment*. Retrieved 4 September 2014, from http://www.gl-assessment.co.uk/research-and-articles/reading-research-and-assessment-jessie-ricketts

Ricketts, J., Bishop, D. V., & Nation, K. (2008). Investigating orthographic and semantic aspects of word learning in poor comprehenders. *Journal of Research in Reading, 31*, 117-135. doi:10.1111/j.1467-9817.2007.00365.x

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

Silva, M., & Cain, K. (2014). The relations between lower and higher level comprehension skills and their role in prediction of early reading comprehension. *Journal of Educational Psychology*. Advance online publication. doi:10.1037/a0037769

Snowling, M. J., Stothard, S. E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., . . . Hulme, C. (2012). *York Assessment of Reading for Comprehension: Passage Reading. Australian Edition.* London, UK: GL Assessment.

# Appendices

# Appendix 1

## Development of Experimental Language and Reading Comprehension Assessments

## Training study outcome measures

### Word selection for training and assessment

Words with an age of acquisition greater than 8 were initially selected from an age of acquisition database, as our participants were all over 8 years of age (Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012). We obtained frequency values for each word using N-Watch (Davis, 2005). Any words for which a frequency value could not be obtained were eliminated. Words were then selected based on how well they fitted the definition of a "Tier 2" word (sophisticated, "literary" words which are used by mature language users, such as "expedition" or "swindle"; Beck, McKeown, & Kucan, 2002, 2008). This procedure resulted in a shortlist of 151 words.

Words were grouped into semantic categories, as per previous studies of the Rich Instruction method of vocabulary instruction (Beck, Perfetti, & McKeown, 1982; McKeown, Beck, Omanson, & Perfetti, 1983; McKeown, Beck, Omanson, & Pople, 1985). Words which could not be grouped into semantic categories were excluded. Definitions for the words were then developed using the Collins Cobuild Advanced Learner's English Dictionary dictionary (Sinclair, 2006). Words for which a simple, clear, "child friendly" definition (Beck et al., 2002, 2008) could not be constructed were excluded. This resulted in a shortlist of 82 words.

These words were then randomly allocated into two sets. Words were then moved across sets until the two resulting groups of words were matched on measures of age of acquisition (Kuperman et al., 2012), as well as measures of imageability, written and spoken frequency, subjective familiarity, number of phonemes, number of syllables and number of letters obtained from N-Watch (Davis, 2005). Both groups of words contained the same proportion of verbs, nouns and adjectives.

One set of words was designated to be the training words. The other set was reduced to 20 words in order to shorten overall assessment time. These 20 words were randomly selected and exchanged as necessary so as to be matched to the 41 trained words on the aforementioned variables. Thus, 61 words in total were assessed (see Chapter 5 Appendix A).

### Vocabulary assessments

#### Definition production.

The ability to define the experimental words was assessed. Children were asked the meaning of a given word, for example “Assist. What does assist mean?”. The task was piloted on two adult participants, and a marking rubric was developed on the basis of these adult pilot responses. Responses were scored by a research assistant who was blind to group membership. Ten percent of responses were double-scored by an additional independent scorer, also blind to group membership. Agreement between scorers was 85%.

Definitions were awarded a score between 0 and 3. Completely incorrect responses were scored 0. A score of 1 was awarded if the child was able to use the target experimental word correctly in a sentence, but there was no indication that the child knew the meaning of the word. This indicated that the child knew the syntactic function of the word, but could not provide its meaning. For example, for the word “assist”, a score of 1 would be awarded for the response “He is going to assist him”. A child was awarded a score of 2 if he or she used a word correctly in a sentence and it was apparent that the child knew the meaning of the word (for example “My mother assists me with my homework because I find it hard”), if the child demonstrated some knowledge of the word’s meaning but used the incorrect part of speech (for example, “If you assist someone you are a helpful person”) or if the meaning was correct, but was poorly phrased or lacked detail (for example, “Helpful”). A score of 3 was given for a well-phrased definition displaying full knowledge of the word’s meaning (for example, “If you assist someone, you help them with something”, “to assist means to help”).

If children did not know the meaning of the word, or took a long time to come up with a response, they were asked to guess the meaning of the word. No other prompts were given.

**Definition recognition.**

At the same time as definition production, children's ability to recognize the definitions of the words was assessed. Once children had been asked to define each word, and after they had provided their response, they completed a multiple choice task in which they were asked to say which of three given definitions best went with the target word.

Correct word definitions for the multiple choice task were adapted from the Collins Cobuild Advanced Learner's English Dictionary (Sinclair, 2006). Children were presented with the definition of the target word as well as two distractor definitions. These distractors were a semantically close, but incorrect, definition and a semantically distant incorrect definition. All definitions were the same part of speech as the target word. For example, for the word "Conceal", the options were "To hide something", "To pretend about something" or "To escape".

After completing the pilot versions of the definition production and recognition tasks, the two adult pilot participants were asked to go through the distractors for each item on the definition recognition task and comment on any they thought were too close to the meaning of the target word. The distractors were then refined based on adult pilot responses. The order of test items and multiple choice responses was randomized at each testing point to reduce practice effects.

Test-retest correlations for both the definition and multiple choice components of the test were calculated by correlating the poor comprehenders' Baseline 1 and Baseline 2 results. Test-retest reliability was high: $r = 0.97$, $p < 0.01$ and $r = 0.90$, $p < 0.05$ respectively. Cronbach's alpha was high for both the definition component, $\alpha = 0.88$, and the multiple choice component, $\alpha = 0.87$, indicating high levels of internal consistency.

### Knowledge of phonological word form.

An auditory lexical decision task was developed to assess children's ability to recognize the phonological form of the words, and distinguish them from other words. The experimental words were presented to children orally, intermixed with nonwords which were developed by changing one phoneme of the core 61 item stimulus set. Children responded by saying "word" when they thought the target was a word, or "not a word" when they thought the target was not a word. Words and nonwords were presented in random order and the order was re-randomised at each testing point. A list of the nonwords and their pronunciations can be found in Appendix 2. Test scores at Baseline 1 and Baseline 2 were moderately correlated, $r = 0.70$, $p < 0.05$. Internal consistency was high, $\alpha = 0.88$.

### Ability to read vocabulary words.

Children were asked to read the stimulus set aloud. The order of words was re-randomized at each testing point. Scores at Baseline 1 and Baseline 2 were highly correlated, $r = 0.89$, $p < 0.01$, and Cronbach's alpha was high, $\alpha = 0.85$.

## Experimenter-designed reading comprehension task

The experimenter-designed reading comprehension measure consisted of two page-long passages which the children read aloud. Passages were written by Danielle Colenbrander. Five words from the set of 41 training words, and five words from the set of 20 control words, were included in each passage. These sets of 5 words were matched to each other on all the matching variables used for the vocabulary task. In order to ensure that the passages were of a level of difficulty suitable for the participants who were children in Grades 3 to 6, several sources of grade-appropriate material were consulted and used as models for passage structure and content. These included National Assessment Program – Literacy and Numeracy (Australian Curriculum, Assessment and Reporting Authority, 2013) sample assessments (NAPLAN assessments are tests of basic reading and writing skills which are administered annually to all Australian children in Grades 3 and 5); sample assessments from the International Competitions and Assessments for Schools (ICAS) English test (Educational Assessment Australia, 2013), undertaken annually on a voluntary basis by children in all grades of Australian schools; fictional texts from the Premier's Reading Challenge book lists (New South Wales Department of Education and Communities, 2013), which are selected to be appropriate for each grade level; and texts from The School Magazine (New South Wales Department of Education and Communities, 2013), a magazine containing original fiction and non-fiction texts designed to be read by children in different stages of primary education.

Children were asked 15 oral questions about the first passage and 16 questions about the second passage. Children were allowed to look back at the passages to answer the questions. The test contained both literal and inferential questions. Literal questions could be answered directly from information present in the text. Inferential questions could be answered by combining two pieces of information from within the text or by combining general knowledge with information from the text. During development, the test was piloted on 7 adults. Any questions that the adults found ambiguous or could not answer were re-worded or replaced.

Literal questions were assigned a score of either 0 or 1. Inferential items were assigned a score of 0, 1, 2 or 3 by a trained research assistant working from a marking rubric developed from the adult pilot results. The research assistant was blind to group membership. A score of 0 was awarded for a completely incorrect answer or a "don't know" response. A score of 1 was awarded for an answer which was close to the correct response but which was incomplete or partially incorrect. A score of 2 was awarded for a response which was correct, but poorly expressed or lacked detail. A score of 3 was awarded for a correct, detailed and well-expressed response. Ten percent of responses were double-scored by an independent scorer who was also blind to group membership. Agreement was 80% percent.

For the purpose of statistical analysis, the scores for the inferential items were converted back to scores out of 1. In other words, for an inferential question, a participant could score 0, 0.3, 0.6, or 1.

**Table A1**

***Coh-Metrix Results for Experimenter-Designed Comprehension Test Passages***

| Coh-Metrix variable | Dog story | Zoo story |
|---|---|---|
| Number of paragraphs | 6 | 10 |
| Number of sentences | 26 | 47 |
| Number of words | 318 | 409 |
| Mean sentences per paragraph | 53 | 40.90 |
| Mean sentence length in words | 12.23 | 8.70 |
| Mean word length in syllables | 1.43 | 1.20 |
| Mean word length in letters | 4.17 | 3.69 |
| Lexical diversity (type/token ratio) | 0.50 | 0.46 |
| CELEX mean word frequency for content words | 2.19 | 2.29 |
| CELEX mean log frequency for all words | 2.88 | 3.03 |
| Mean age of acquisition for content words[a] | 246.03 | 269.25 |
| Mean familiarity rating for content words[a] | 580.59 | 575.13 |
| Mean concreteness of content words[a] | 390.79 | 404.62 |
| Mean imageability of content words[a] | 421.45 | 431.93 |
| Flesch-Kincaid Grade level[b] | 6.10 | 1.96 |

*Note.* [a]Age of acquisition, familiarity, concreteness and imageability are based on adult ratings converted to a scale between 100 and 700. A score of 700 represents the highest possible age of acquisition, level of familiarity, concreteness and imageability. [b] Calculated as follows: (0.39 x average sentence length) + (11.8 x average number of syllables per word) – 15.59

The final versions of the passages were analyzed using Coh-Metrix 3.0 (McNamara, Louwerse, Cai & Graesser, 2011). Selected results from this analysis are presented in Table A1[16]. As shown in Table A1, the passages were not equal in difficulty, however as all participants completed both passages, and as equal numbers of training and transfer words were present in each passage, this was not seen as a problem.

Comprehension scores on the experimental comprehension test at Baseline 1 and Baseline 2 were highly correlated, $r = 0.928$, $p < 0.01$. Cronbach's alpha for the comprehension scores was also high, $\alpha = 0.84$, indicating that the test had a high level of internal consistency. Experimental comprehension test scores at Baseline 1 correlated highly with NARA comprehension scores at Baseline 1, $r = 0.82$, $p < 0.01$.

The final experimental reading comprehension test appears in Appendix 3.

**Morphological awareness**

Morphological awareness was assessed as an outcome measure in the randomized controlled trial because we wished to assess whether there was any generalization to tasks of broader oral language or meta-linguistic skill.

The majority of existing tasks of morphological awareness used in the literature (e.g., Carlisle, 2000; Kirby et al., 2011; Tong, Deacon, & Cain, 2014) use real words as the stems of morphologically complex words. However, we wished to use nonsense word bases because we were interested in obtaining a relatively pure measure of a child's ability to recognize morphemes and determine their meaning without the influence of word frequency, or phonological or orthographic transparency (whether the spelling or pronunciation of the base word is intact within the morphologically complex word), factors which have been shown to affect children's performance on morphological tasks (Carlisle, 2000; Carlisle & Stone, 2005). In addition, we wished to develop a task which closely paralleled the process a child might go through when encountering an unfamiliar morphologically complex word in text for

[16] Cohmetrix 3.0 produces over 100 output variables. For reasons of clarity and parsimony, we have chosen to present only 15 of these variables.

the first time. With these goals in mind, we developed our own task of morphological awareness.

We assessed children's knowledge of the meaning of derivational suffixes, because morphological difficulties in poor comprehenders seem to occur primarily in the domain of derivational morphology (Tong et al., 2014; Tong, Deacon, Kirby, Cain, & Parrila, 2011). Children were read aloud questions about the meanings of nonsense words, such as "Varp. Which one means something like 'a bit like a varp'?" Children then chose from four different multiple-choice responses, each of which consisted of a nonsense word base (eg. varp) plus a suffix. One response contained the target suffix (eg. varpish) and the other three responses contained distracter suffixes (eg.  varpable, varpence and varpless). To avoid placing a heavy load on phonological working memory, children were asked to read along with a written version of the questions.

During test development, we selected the derivational morphemes from a list of English suffixes and their frequencies developed from the CELEX database (Baayen, Piepenbrock & van Rijn, 1993; Beyersmann, Castles & Coltheart, 2012). We then developed definitions for each derivational suffix. We rejected any suffixes with very abstract meanings or which could not be defined (ie. suffixes which performed the function of changing the base word's part of speech, but did not have a clear definition, such as "–ness" as in "happiness"), because we thought these would be too difficult for children. This resulted in a shortlist of 12 suffixes. We then developed questions about the meaning of each suffix, of the form "Which word means something like [suffix definition]".

We developed nonsense words to act as base words by changing one or two letters of real words which were commonly associated with each of the shortlisted suffixes. In some cases this resulted in nonwords which were not orthographically or phonologically legal in English. For these items, we used nonwords from the ARC nonword database (Rastle, Harrington & Coltheart, 2002). We continued this process until there were at least 4 nonsense words for each suffix. This meant that there would be at least four multiple choice questions for each shortlisted suffix.

We developed three distractors for each multiple choice question by combining the nonsense word base for each question with other suffixes from the shortlist.

This version of the test was piloted on ten adults. Based on the results of this pilot, six items that the adults found ambiguous or overly difficult were removed or altered and the number of suffixes assessed was reduced to nine. This version of the test was then piloted on eight children between the ages of 8 and 11. The piloted version of the test was then shortened by selecting only 3 nonword items for each suffix. Items were rejected if they were much easier or much harder than the other items for a particular suffix. In some cases there were no obvious targets for elimination and items to be removed were chosen at random. This resulted in 27 items. Test-retest correlation was moderate, $r = 0.65$, $p < 0.05$. Internal consistency was high, $\alpha = 0.87$. This test is provided in Appendix 4.

## Other oral language tasks

A number of additional oral language measures were developed for the study in Chapter 3 which examined the vocabulary and oral language skills of individual poor comprehenders. The development of these assessments is outlined below.

### Conceptual semantics task.

We were interested in whether children's difficulties with lexical semantics would also be visible in a task of non-linguistic concept knowledge (conceptual semantics). Therefore, we developed a picture-picture association task by combining items from the Squirrel Nut Test (Pitchford, Funnell, Ellis, Green, & Chapman, 1997) and a semantic association test originally developed for use with Hebrew speakers (Biran & Friedmann, 2007). There were 38 items in total. During administration, three pictures were displayed on a computer screen. Pictures came from the Biran et al. (2007) semantic association test and from Microsoft Word Clip Art. The target picture was displayed at the top of the screen (for example, a raincloud) and two stimulus pictures were displayed at the bottom of the screen (for example, an umbrella and a broom). The child's task was to select which of the two stimulus pictures were semantically linked to the target picture by pressing an appropriate key. Pictures were presented on a laptop computer using E-Prime 2.0 software (Psychology Software Tools, 2012). Both accuracy and reaction time were recorded. Test-retest reliability information is not available for this test because children were only assessed on this test at a single point in time, and Cronbach's alpha could not be calculated as the majority of participants were at or near ceiling on this task. A list of the test items appears in Appendix 5.

**Syntax task.**

As with morphology, we wished to assess syntax in a task which was closely aligned to the type of task children perform when comprehending speech or text. Therefore, we developed a task of syntactic comprehension. This was adapted from a task designed for use with Hebrew speakers (Friedmann & Novogrodsky, 2002). Work by Friedmann and colleagues (Friedmann & Levy, 2009; Friedmann & Novogrodsky, 2004, 2011) has shown that tasks involving the comprehension of object wh-questions and relative clauses are particularly difficult for children with syntactic impairments. Similar results have been found in English (van der Lely, Jones, & Marshall, 2011). Therefore, we reasoned that such a task might be sensitive to subtle syntactic impairments in poor comprehenders.

During the task, pictures depicting three characters (for example, two women and a girl) were displayed on a computer screen using Microsoft Powerpoint. The tester read aloud a series of sentence relating to the pictures. The child was asked to point to the correct referent for each sentence. Ten of the sentences were subject wh- questions ("Which lady is pinching the girl?"), ten were object wh- questions ("Which lady is the girl pinching?"), ten were subject relatives ("Point to the lady that is pinching the girl") and ten were object relatives ("Point to the lady that the girl is pinching"). Sentences were presented intermixed in a random order. Subject wh- and subject relative sentences were expected to be relatively easy for all children, but they were included so that low scores on the object questions could be attributed to syntactic difficulties rather than processing factors unrelated to syntax. A list of the test items appears in Appendix 6.

## References

Australian Curriculum, Assessment and Reporting Authority. (2013). *National Assessment Program (NAP) Sample Assessments*. Retrieved from http://www.nap.edu.au/nap-sample-assessments/nap-sample-assessments.html.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical database* [CD-ROM]. Philadelphia, PA: University of Pennsylvania, Linguistic Data Consortium.

Beck, I., McKeown, M., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. New York, NY: Guilford.

Beck, I., McKeown, M., & Kucan, L. (2008). *Creating robust vocabulary: Frequently asked questions and extended examples*. New York, NY: Guilford.

Beck, I., Perfetti, C., & McKeown, M. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology, 74,* 506-521. doi:10.1037/0022-0663.74.4.506

Beyersmann, E., Castles, A., & Coltheart, M. (2012). Morphological processing during visual word recognition in developing readers: Evidence from masked priming. *The Quarterly Journal of Experimental Psychology, 65*, 1306-1326. doi:10.1080/17470218.2012.656661

Biran, M., & Friedmann, N. (2007). *'Shmita vehar Sinai': picture and word association test.* Tel Aviv, Israel: Tel Aviv University.

Carlisle, J. F. (2000). Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading and Writing, 12*, 169-190. doi:10.1023/A:1008131926604

Carlisle, J. F., & Stone, C. A. (2005). Exploring the role of morphemes in word reading. *Reading Research Quarterly, 40*, 428-449. doi:10.1598/RRQ.40.4.3

Sinclair, J. (Ed.). (2006). *Collins COBUILD advanced learner's English dictionary* (5th ed). Glasgow, UK: HarperCollins.

Davis, C. J. (2005). N-Watch: A program for deriving neighbourhood size and other psycholinguistic statistics. *Behavior Research Methods, 37*, 65-70. doi:10.3758/BF03206399

Educational Assessment Australia. (2013). *International Competitions and Assessments for Schools (ICAS) Practice Questions*. Retrieved from https://www.eaa.unsw.edu.au/icas/download-practice-questions

Friedmann, N., & Levy, H. (2009). Treatment of syntactic movement in syntactic SLI: A case study. *First Language*, *29*, 15-50. doi:10.1177/0142723708097815

Friedmann, N., & Novogrodsky, R. (2002). *BAMBI: Battery for assessment of syntactic abilities in children*. Tel Aviv, Israel: Tel Aviv University.

Friedmann, N., & Novogrodsky, R. (2004). The acquisition of relative clause comprehension in Hebrew: A study of SLI and normal development. *Journal of Child Lanuageg, 31*, 661-681. doi:10.1017/S0305000904006269

Friedmann, N., & Novogrodsky, R. (2011). Which questions are most difficult to understand?: The comprehension of Wh questions in three subtypes of SLI. *Lingua, 121*, 367-382. doi:10.1016/j.lingua.2010.10.004

Kirby, J. R., Deacon, S. H., Bowers, P. N., Izenberg, L., Wade-Woolley, L., & Parrila, R. (2011). Children's morphological awareness and reading ability. *Reading and Writing, 25*, 389-410. doi:10.1007/s11145-010-9276-5

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, 44*, 978-990. doi:10.3758/s13428-012-0210-4

McKeown, M., Beck, I., Omanson, R., & Perfetti, C. (1983). The effects of long-term vocabulary instruction on reading comprehension: A replication. *Journal of Literacy Research, 15*, 3-18. doi:10.1080/10862968309547474

McKeown, M., Beck, I., Omanson, R., & Pople, M. (1985). Some effects of the nature and frequency of vocabulary instruction on the knowledge and use of words. *Reading Research Quarterly, 20*, 522-535. doi:10.2307/747940

New South Wales Department of Education and Communities. (2013). New South Wales Premier's Reading Challenge Booklists. Retrieved from https://online.det.nsw.edu.au/prc/booklist/home.html

Pitchford, N. J., Funnell, E., Ellis, A. W., Green, S. H., & Chapman, S. (1997). Recovery of spoken language processing in a 6-year-old child following a left hemisphere stroke: A longitudinal study. *Aphasiology, 11*, 83-102. doi:10.1080/02687039708248457

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC nonword database. *Quarterly Journal of Experimental Psychology, 55A*, 1339-1362. doi:10.1080/02724980244000099

Tong, X., Deacon, S. H., & Cain, K. (2014). Morphological and syntactic awareness in poor comprehenders: another piece of the puzzle. *Journal of Learning Disabilities, 47*, 22-33. doi:10.1177/0022219413509971

Tong, X., Deacon, S. H., Kirby, J. R., Cain, K., & Parrila, R. (2011). Morphological awareness: A key to understanding poor reading comprehension in English. *Journal of Educational Psychology, 103*, 523-534. doi:10.1037/a0023495

Psychology Software Tools. (2012). E-Prime 2.0. Pittsburgh, PA: Psychology Software Tools.

van der Lely, H. K., Jones, M., & Marshall, C. R. (2011). Who did Buzz see someone? Grammaticality judgement of wh-questions in typically developing children and children with Grammatical-SLI. *Lingua, 121*, 408-422. doi:10.1016/j.lingua.2010.10.007

## Appendix 2

### Auditory Lexical Decision Task Items

| Words | | Nonwords | |
|---|---|---|---|
| abrupt | modest | abendance | naiorv |
| abundance | naive | abrept | nemble |
| ally | nimble | arge | obsoove |
| amateur | observe | assost | offond |
| anxious | offend | boonefactor | omateur |
| assist | petty | breet | petto |
| benefactor | ploy | canquer | plor |
| brute | precise | confarm | preserce |
| coerce | prevent | confede | prevoant |
| conceal | pursue | consorl | pursyor |
| confide | rebel | darbeous | regroot |
| conform | regret | dispote | requost |
| conquer | request | eedle | ribble |
| dispute | scarce | enthaysiasm | sarze |
| dubious | scorn | expedotion | scarn |
| enthusiasm | seize | fatoogue | scorce |
| envy | serene | foo | serain |
| expedition | shrewd | forey | sheeder |
| fatigue | shudder | frintic | shrord |
| foe | swindle | groodual | swendle |
| frantic | temptation | heesh | tance |
| fury | tense | illigh | teerant |
| gradual | triumph | immonse | temptooshen |
| harsh | tyrant | impore | trayumph |
| idle | unique | invooder | unorque |
| immense | urge | inxious | unvy |
| impair | vacant | keyerce | vashious |
| invader | vague | lenge | vendle |
| lunge | vandal | meck | vorgue |
| mingle | vicious | middest | vuccent |
| mock | | mungle | |

*Note.* Items appeared intermixed in random order during testing. The order of items was changed at each time of testing.

## Appendix 3

## Experimental Reading Comprehension Test

### Passage 1

*Note: Trained and matched untrained words included in the passage are underlined*

My name is Jess and I have a dog called Pirate. We call him Pirate because he has a big scar across his nose. As well as his scar, Pirate has only three legs. Although this looks a bit funny, it doesn’t impair him at all.

One day, our neighbour invited Pirate over to play with her dog. I wasn’t so sure that this was a good idea. The neighbour’s dog, Tinkerbell, was very small and slow and always had a vacant look in her eyes. I didn’t think she and Pirate would get along.

Mum and I walked Pirate into our neighbour’s house. Pirate’s paws were muddy and he left big pawprints on the tiles. We went out into the back garden. Tinkerbell was chewing on a small bone and looked up at Pirate dubiously.

Pirate ran over as best he could on three legs, and started sniffing at Tinkerbell. She began to growl and tried to conceal her bone. Clearly she did not like the look of this invader! Tinkerbell’s growling didn’t bother Pirate at all and he bounced around her enthusiastically.

“Now Tinkerbell, be nice!” said our neighbour, bending over to pat her. Unfortunately Pirate did not see our neighbour and ran right into her, knocking her over. Her phone fell out of her pocket and, in one nimble movement, Pirate had grabbed it!

“No Pirate!” Mum yelled, “Drop it!”

Pirate thought it was a wonderful game as we pursued him around and around the garden trying to get the phone. As we ran, our neighbour became more and more frantic. Eventually I managed to catch Pirate and Mum pulled the soaking-wet phone from his slobbery jaws.

Needless to say, Pirate was never invited over to play with Tinkerbell again. Mum and I were very sorry, but we couldn’t help thinking it was a little bit funny! Pirate may be a troublemaker, but he’s totally <u>unique</u>, and we love him.

**Questions.**

1. Why is the dog in the story called Pirate?
2. How many legs does Pirate have?
3. What is the neighbour’s dog called?
4. Why do you think Tinkerbell looked at Pirate dubiously?
5. How do we know that Tinkerbell thinks Pirate is an invader?
6. "Tinkerbell's growling didn't bother Pirate at all and he bounced around her enthusiastically". What does this tell us about Pirate's personality?
7. Why did Tinkerbell try to conceal her bone?
8. How did Pirate get the phone?
9. Why did everyone pursue Pirate around the garden?
10. Why do you think the neighbor became more and more frantic?
11. Who got the phone back from Pirate?
12. How do we know that Pirate isn’t impaired by his missing leg?
13. Why did Jess and her mum think that it was funny when Pirate stole the phone?
14. Why was Pirate never invited over to play with Tinkerbell again?
15. What makes Pirate unique?

**Passage 2**

Today we went on a school outing to the zoo. I like animals, but I am very scared of birds. I didn't want to tell anyone, and was hoping nobody would notice.

Unfortunately, the first thing we did at the zoo was to go and see the birds! First we saw a huge hawk in a cage. "This type of hawk is very <u>scarce</u> in the wild. Actually most of the birds in the zoo are <u>scarce</u> in the wild," our teacher said.

Suddenly the hawk screeched. I jumped!

"What a noise it makes" she said, "and look at its sharp beak! Hawks can be quite <u>vicious</u>!" I shivered.

Then we went and stood outside a huge cage full of birds.

"Now we have a special treat," the teacher said. "We can actually go inside this cage, so you can <u>mingle</u> with some of these beautiful birds!"

"I'm really scared of birds," I <u>confided</u> in my friend Alex. "I don't want to go in!"

We got to the entrance of the cage. Everybody was walking in happily. But as soon as I got to the entrance, I stopped <u>abruptly</u>. A boy from my class, Sam, had been standing right behind me, and he bumped straight into me.

"Why did you stop, Tim?" he asked, with a look of <u>scorn</u>. "Are you scared of birds?"

"N…n..no…" I said, and walked nervously into the cage.

Once we walked inside, I saw Sam starting to pull something out of his pocket.

"Can you hold something for me?" he asked.

"Um…OK.." I said, and held out my hand. I shouldn't have been so <u>naïve</u>! Sam had pulled some bits of bread off his sandwich and put them in my hand. Before I knew it, birds were swooping down all around me. It was an <u>immense</u> effort not to scream! I threw the pieces of bread onto the ground.

"What's going on?" my teacher said, rushing over. She saw the bits of bread on the ground. Then she took one look at Sam's face and knew what had happened. "You're in big trouble, Sam!"

"Come on Tim," she said, leading me out of the cage. "I'm very proud of you for being so brave."

"I wasn't brave at all," I said, feeling <u>fatigue</u> come over me.

"Brave and <u>modest</u> too!" said my teacher. "I think you deserve an ice-cream!"

**Questions.**

1. Where did the class go on the school outing?
2. What kind of bird did they see first ?
3. Why does Tim shiver when he hears the hawk is vicious?
4. Why might scarce birds be kept in the zoo?
5. What is the special treat for the class?
6. Why did Tim stop abruptly when he got to the entrance of the cage?
7. Why does Sam give Tim a look of scorn?
8. Does Sam believe Tim when he says he is not afraid of birds? How do you know?
9. What did Sam put into Tim's hand?
10. Why was it naïve for Tim to put out his hand?
11. Why did the birds swoop around Tim?
12. Why did Tim throw the pieces of bread onto the ground?
13. "Then she took one look at Sam's face and knew what was going on". Tell me how the teacher knows what is going on.
14. Why do you think Tim feels fatigue after he comes out of the birdcage?
15. Why is Sam in big trouble?
16. Why does the teacher say that Tim is "brave and modest too"?

## Appendix 4

## Morphological Awareness Test

Practice item 1. Fat. Which word means something like “to be full of fat”?

a. fatless b. fatty

c. fatly d. fattest

Practice item 2. Toy. Which one means something like “more than one toy”?

a. toying b. toyless

c. toys d. toy

Practice item 3. Gamp. Which one means something like “the one who is the most gamp”?

a. gamping b. gampless

c. gampest d. gamply

1. Pild. Which one means something like “to be without pild”?

a. pildous b. pilder

c. pildible d. pildless

2. Florp. Which one means something like “a person who florps”?

a. florpness b. florpence

c. florpist d. florpish

3. Snold. Which one means something like “a bit like a snold”?

a. snoldable b. snoldify

c. snoldish d. snoldless

4. Snim. Which one means something like “to make snim”?

a. snimmify b. snimming

c. snimmible d. snimmed

5. Slont. Which one means something like “could be slonted”?

a. slontable b. slontify

c. slontish d. slontful

6. Gress. Which one means something like “can be gressed”?

a. gressible b. gressify

c. gressless d. gressish

7. Smair. Which one means something like “to be without smair”?

a. smairable b. smairish

c. smairify d. smairless

8. Lafe. Which one means something like “the act of lafing”?

a. lafeness b. lafer

c. lafeish d. lafement

9. Blean. Which one means something like “a person who bleans”?

a. bleanify b. bleaner

c. bleanity d. bleanance

10. Spresh. Which one means something like “the act of spreshing”?

a. spresher b. spreshible

c. spreshness d. spreshment

11. Grat. Which one means something like “to be without grat”?

a. gratter b. grattish

c. gratless d. grattable

12. Neech. Which one means something like “could be neeched”?

a. neechable b. neechful

c. neechify d. neechive

13. Stilch. Which one means something like “to be full of stilch”?

a. stilchous b. stilchable

c. stilchify d. stilchish

14. Tarb. Which one means something like “a person who tarbs”?

a. tarbation b. tarbist

c. tarbence d. tarbish

15. Spetch. Which one means something like “to make spetch”?

a. spetching b. spetched

c. spetchify d. spetchous

16. Jerch. Which one means something like “a bit like a jerch”?

a. jerchable b. jercher

c. jerchous d. jerchish

17. Trab. Which one means something like “a person who trabs”?

a. trabbish b. trabness

c. trabber d. trabment

18. Frimple. Which one means something like “to make frimple”?

a. frimpling b. frimplify

c. frimplish d. frimpled

19. Pleeb. Which one means something like “to be full of pleeb”?

a. pleeber b. pleebable

c. pleebless d. pleebous

20. Plert. Which one means something like “can be plerted”?

a. plertible b. plertish

c. plertist d. plertful

21. Blerth. Which one means something like “to be full of blerth”?

a. blerthous b. blerthist

c. blerthish d. blerthable

22. Plink. Which one means something like “a person who plinks”?

a. plinking b. plinkness

c. plinker d. plinkance

23.Varp. Which one means something like “a bit like a varp”?

a. varpable b. varpence

c. varpless d. varpish

24. Triss. Which one means “could be trissed”?

a. trissable b. trissful

c. trissify d. trissive

25. Plart. Which one means “a person who plarts”?

a. plartence b. plartish

c. plartist d. plartness

26. Toove. Which one means “the act of tooving”?

a. toover b. toovity

c. toovement d. toovible

27. Drex. Which one means something like “can be drexed”?

a. drexible b. drexish

c. drexist d. drexful

## Appendix 5

## Semantic Association Test Items

| | Display item | Choice 1 | Choice 2 | Semantic link |
|---|---|---|---|---|
| P1 | Raincloud | Umbrella | Broom | Use umbrella when it rains |
| P2 | Ear | Necklace | Earring | Earring worn in the ear |
| 1 | Soup | Spoon | Fork | Soup eaten with spoon |
| 2 | Goal | Soccer ball | Basketball | Soccer ball kicked into goal |
| 3 | Umbrella | Sandals | Gumboots | Both used when it rains |
| 4 | Arm | Leg | Eye | Limbs |
| 5 | Glove | Foot | Hand | Glove worn on hand |
| 6 | Keyboard | Monitor | Radio | Used together |
| 7 | Pot | Cake | Stew | Pot used to make stew, not cake |
| 8 | Anchor | Boat | Plane | Anchor keeps boat in place |
| 9 | Bread | Knife | Scissors | Cut bread with knife, not scissors |
| 10 | Tie | Necklace | Ring | Worn around the neck |
| 11 | Egg | Dog | Hen | Hen lays eggs |
| 12 | Lipstick | Eye | Lips | Lipstick goes on lips |
| 13 | Nest | Bird | Fish | Bird lays eggs in nest |
| 14 | Glasses | Lips | Eyes | Glasses worn on eyes |
| 15 | Bed | Pillow | Backpack | Pillow goes on bed |
| 16 | Saddle | Giraffe | Horse | Saddle goes on horse |
| 17 | Juice | Glass | Plate | Juice goes in glass |
| 18 | Lips | Piano | Trumpet | Trumpet blown |
| 19 | Giraffe | Kangaroo | Elephant | African animals |
| 20 | Curtains | Door | Window | Curtains go on a window |
| 21 | Chef | Cookpot | Plant pot | Both to do with cooking |
| 22 | Lake | Duck | Cat | Ducks swim in lakes |
| 23 | Wool | Lion | Sheep | Wool comes from a sheep |
| 24 | Baby | Bottle | Glass | Babies drink from a bottle |
| 25 | Wine | Tomato | Grapes | Wine made from grapes |
| 26 | Rollerskate | Glove | Boot | Worn on the feet |
| 27 | Cow | Milk | Coke | Milk comes from a cow |
| 28 | Sunglasses | Night | Day | Sunglasses worn in daytime |
| 29 | Tie | Jacket | Swimsuit | Tie worn with jacket |
| 30 | Phone | Nose | Ear | Listen to the phone |
| 31 | Thermometer | Pencil | Stethoscope | Things doctors use |
| 32 | Car | Steering wheel | Helm | Steering wheel in car |
| 33 | Tea | Shopping bag | Teabag | Teabag goes in tea |
| 34 | Swimring | Swimsuit | Jumper | Things you swim with |
| 35 | Hair | Comb | Toothbrush | Brush hair with hairbrush |
| 36 | Shore | Seal | Dog | Seal lives on the shore |
| 37 | Gift | Bread | Cake | Things you have at birthday parties |
| 38 | Bear | Penguin | Swan | Live in cold habitats |

## Appendix 6

### Syntax Task Items

| Sentence | Item type |
|---|---|
| 1. Point to the boy who is pushing the man. | Practice |
| 2. Which boy is pouring water on the man? | Practice |
| 3. Which lady is the girl pinching? | Object question |
| 4. Point to the bear that is pushing the girl. | Subject relative clause |
| 5. Which man is the boy pulling? | Object question |
| 6. Point to the giraffe that the girl is measuring. | Object relative clause |
| 7. Which octopus is grabbing the man? | Subject question |
| 8. Point to the giraffe that is measuring the girl. | Subject relative clause |
| 9. Point to the elephant that the boy is spraying | Object relative clause |
| 10. Point to the angel that the man is patting. | Object relative clause |
| 11. Which man is pulling the boy? | Subject question |
| 12. Point to the angel that is patting the man. | Subject relative clause |
| 13. Point to the octopus that is grabbing the man. | Subject relative clause |
| 14. Which bear is pushing the girl? | Subject question |
| 15. Which octopus is the man grabbing? | Object question |
| 16. Point to the lady that is pinching the girl. | Subject relative clause |
| 17. Which angel is the man patting? | Object question |
| 18. Point to the bear that the girl is pushing. | Object relative clause |
| 19. Which dolphin is drying the boy? | Subject question |
| 20. Which lady is pinching the girl? | Subject question |
| 21. Point to the elephant that is spraying the boy | Subject relative clause |
| 22. Point to the man that is pulling the boy | Subject relative clause |
| 23. Point to the dolphin that the boy is drying. | Object relative clause |
| 24. Point to the boy who is pushing the man. | Subject question |
| 25. Which boy is pouring water on the man? | Object relative clause |
| 26. Which lady is the girl pinching? | Object question |
| 27. Point to the bear that is pushing the girl. | Subject relative clause |
| 28. Which man is the boy pulling? | Subject question |
| 29. Point to the giraffe that the girl is measuring. | Subject question |
| 30. Which octopus is grabbing the man? | Object relative clause |
| 31. Point to the giraffe that is measuring the girl. | Object question |
| 32. Point to the elephant that the boy is spraying | Object question |
| 33. Point to the angel that the man is patting. | Object question |
| 34. Which man is pulling the boy? | Subject question |
| 35. Point to the angel that is patting the man. | Subject question |
| 36. Point to the octopus that is grabbing the man. | Object relative clause |
| 37. Which bear is pushing the girl? | Subject relative clause |
| 38. Which octopus is the man grabbing? | Object relative clause |
| 39. Point to the lady that is pinching the girl. | Object relative clause |
| 40. Which angel is the man patting? | Object question |
| 41. Point to the bear that the girl is pushing. | Subject relative clause |
| 42. Which dolphin is drying the boy? | Object question |

# Appendix 7

## Final Ethics Approval

Gmail

Danielle Colenbrander <danielle.colenbrander@gmail.com>

**Approved- Ethics application Kohnen (5201200038)**

1 message

**Ethics Secretariat <ethics.secretariat@mq.edu.au>** **Fri, Mar 16, 2012 at 10:25 AM**
To: Dr Saskia Kohnen <saskia.kohnen@mq.edu.au>
Cc: Dr Karen Smith-Lock <karen.smith-lock@mq.edu.au>, Prof Lyndsey Nickels <lyndsey.nickels@mq.edu.au>, Miss Danielle Colenbrander <Danielle.Colenbrander@mq.edu.au>

Dear Dr Kohnen

Re: "Understanding and treating the causes of reaching comprehension problems" (Ethics Ref: 5201200038)

Thank you for your recent correspondence. Your response has addressed the issues raised by the Human Research Ethics Committee and you may now commence your research.

The following personnel are authorised to conduct this research:

Chief Investigator- Dr Saskia Kohnen

Co-Investigators- Dr Karen Smith-Lock, Miss Danielle Colenbrander and Prof Lyndsey Nickels

NB. STUDENTS: IT IS YOUR RESPONSIBILITY TO KEEP A COPY OF THIS APPROVAL EMAIL TO SUBMIT WITH YOUR THESIS.

Please note the following standard requirements of approval:

1. The approval of this project is conditional upon your continuing compliance with the National Statement on Ethical Conduct in Human Research (2007).

2. Approval will be for a period of five (5) years subject to the provision of annual reports. Your first progress report is due on 16 March 2013.

If you complete the work earlier than you had planned you must submit a Final Report as soon as the work is completed. If the project has been discontinued or not commenced for any reason, you are also required to submit a Final Report for the project.

Progress reports and Final Reports are available at the following website:

http://www.research.mq.edu.au/for/researchers/how_to_obtain_ethics_approval/human_research_ethics/forms

3. If the project has run for more than five (5) years you cannot renew approval for the project. You will need to complete and submit a Final Report and submit a new application for the project. (The five year limit on renewal of approvals allows the Committee to fully re-review research in an environment where legislation, guidelines and requirements are continually changing, for example, new child protection and privacy laws).

4. All amendments to the project must be reviewed and approved by the Committee before implementation. Please complete and submit a Request for Amendment Form available at the following website:

http://www.research.mq.edu.au/for/researchers/how_to_obtain_ethics_approval/human_research_ethics/forms

5. Please notify the Committee immediately in the event of any adverse effects on participants or of any unforeseen events that affect the continued ethical acceptability of the project.

6. At all times you are responsible for the ethical conduct of your research in accordance with the guidelines established by the University. This information is available at the following websites:

http://www.mq.edu.au/policy/

http://www.research.mq.edu.au/for/researchers/how_to_obtain_ethics_approval/human_research_ethics/policy

If you will be applying for or have applied for internal or external funding for the above project it is your responsibility to provide the Macquarie University's Research Grants Management Assistant with a copy of this email as soon as possible. Internal and External funding agencies will not be informed that you have final approval for your project and funds will not be released until the Research Grants Management Assistant has received a copy of this email.

If you need to provide a hard copy letter of Final Approval to an external organisation as evidence that you have Final Approval, please do not hesitate to contact the Ethics Secretariat at the address below.

Please retain a copy of this email as this is your official notification of final ethics approval.

Yours sincerely
Dr Karolyn White
Director of Research Ethics
Chair, Human Research Ethics Committee