Teaching high frequency words to poor readers using flashcards: Its effects on novel word acquisition, skill transfer to in-text word reading, and passage reading competencies

A thesis in partial fulfillment of the requirements for the degree of

Master of Educational Psychology

Massey University, Albany, New Zealand

Butul Samera Iqbal

2016

Abstract

Several literacy reports published in the last decade have emphasised the large gap in the reading attainment of children in New Zealand. A common barrier that prevents poor readers to catch up to their peers is difficulty in reading fluency, which is theorised to represent underlying difficulty in rapid and automatic word recognition. The ability to rapidly recognise a few common words, also known as *high frequency words* (HFWs), may increase the fluency of reading the majority of novel text. As such, the National Standards for literacy achievement outline the development of basic HFW vocabulary by the end of the first few years at school. However, past research that has investigated single word training has rarely used HFWs and those that have used HFWs have scarcely investigated its transfer to in-text reading. Therefore, the aims of the current research were to provide an investigation of HFW training and its influence on word reading accuracy, in-text word reading, and passage reading accuracy, speed, and comprehension. Experiment 1 was a single case design carried out with one 8 year old participant and was largely used to inform the second experiment. Experiment 2 was a multiple baseline design carried out with five 8-9 year old participants using a modified training procedure. Experiment 1 utilised visual analysis and Cohen's d effect size analysis whereas Experiment 2 also used statistical analysis, made possible through the Wampold-Worsham method of randomisation incorporated into the experimental design. The results of both experiments indicated that training facilitated word reading accuracy but the successful transfer of target words to in-text reading was only observed in Experiment 2. Post-training increases to passage reading accuracy, speed, and comprehension scores were not apparent in either experiment. The main contribution of the current research is its applicability to classroom practice. Another important contribution of the study to research practice is the rare application of the Wampold-Worsham method of randomisation

Acknowledgements

I would like to sincerely thank everyone that has supported me in completing this research project. Thank you to the wonderful children that took part in the study, watching you all grow in confidence and enjoy the training with me confirmed to me that all the hard work I was putting in was meaningful and worthwhile. Thank you also to the amazing teachers at the wonderful west Auckland school for welcoming me into your school community.

I would like to thank my supervisor Tom for your constant support and guidance. I really appreciate the effort you put in to listen to all my complicated (and sometimes convoluted) ideas. It made a huge difference to my confidence in communicating my ideas. I would also like to thank my second supervisors Jane for contributing her amazing talent for critical thinking and attention to detail, and Alison for contributing her valuable subject expertise at the end stages of the project.

My family has been my rock throughout this journey and I would like to express my immense gratitude for the unyielding support, love, and enthusiasm of my amazing mama, siblings, and husband study buddy.

Finally, I would like to thank the Almighty for giving me the opportunity to know and meet such amazing people. I will (God willing) continue to seek my path to success as one that is defined by my intention to please God.

Table of Contents

List of Illustrations

List of Tables

Chapter 1 - Introduction to the Study

Reading of text is a multifaceted skill. It involves a combination of factors including oral language ability (Share, 2004), understanding and application of letter-sound associations (Ainsworth, Evmenova, Behrmann, & Jerome, 2016), short and long-term memory (Kornell, 2009), cognitive capacity (Kuhn & Stahl, 2003), and motivation (Walczyk & Griffith-Ross, 2007).

International survey data indicates that New Zealand has one of the highest levels of disparity between good and poor readers in the OECD (Chamberlain, 2013; Ogle et al., 2003). Tunmer, Chapman, and Prochnow (2004) attribute this to the fact that the majority of children who start school with below average literacy skills continue to have reading difficulty in their subsequent years at school. This can be likened to Stanovich's (1986) concept of *Matthew Effects in Reading* which describes the difficulties children with poor reading skills may have in catching up to their peers. These children may continue to fall behind their peer group without reading instruction catered to their specific difficulty.

Difficulty in reading fluency is a common problem faced by poor readers and is described to be the primary cause of reading fatigue and decreased motivation to engage in reading tasks (Nathan & Stanovich, 1991; Richards, 2000; Schwanenflugel, Hamilton, Kuhn, Wisenbaker, & Stahl, 2004; Walczyk & Griffith-Ross, 2007). Fluent reading is also thought to be the key to reading comprehension (Ehri, 2014, Tan & Nicholson, 1977). Although there is some variance in the literature regarding the definition of fluent reading, three characteristic properties are

highlighted by the majority of authors; the accuracy of word identification, the speed of reading, and the application of appropriate prosodic rules (Kuhn & Stahl, 2003; Pikulski & Chard, 2005).

The main theoretical construct associated with fluent reading is automaticity (Nathan & Stanovich, 1991; Richards, 2000; Tønnessen & Uppstad, 2015). Automaticity in simple terms means the ability to perform a task quickly and effortlessly as a result of repeated practice (Samuels & Flor, 1997; Tønnessen & Uppstad, 2015). The inability to achieve automaticity is described by Perfetti and Lesgold's (1979) early bottleneck hypothesis as the inability to move beyond the slow decoding of sounds in words toward identifying words quickly by sight, leading to laborious and non-fluent reading. Without instruction, children that experience this bottleneck may fall behind their peers in reading attainment due to decreased motivation to gain reading experience (Stanovich, 1986; Walczyk & Griffith-Ross, 2007).

Teaching strategies aimed at improving fluency have used *sight word* identification training (Martin-Chang, Levy & O'Neil, 2007; McArthur et al., 2013; Watts & Gardner, 2013). *Sight words* are words that can be identified rapidly and automatically without conscious effort (Ehri, 2014; Fry, 1980, 1997). Single sight word training has been widely investigated in the literature as a way to boost reading fluency without forcing children to first laboriously read extended text, with the expectation that, when they do go back to reading those same words in a passage, it is more fluent and gratifying (Stanovich, 1986; Tønnessen & Uppstad, 2015; Walczyk & Griffith-Ross, 2007).

Much research has attempted to conceptualise the relationship between sight word development and passage reading competencies (Ehri, 2014; Fuchs, Fuchs, Hosp, & Jenkins,

2001; Nathan & Stanovich, 1991; Tønnessen & Uppstad, 2015). Some research has investigated it through observational research designs (Clemens, Shapiro, Wu, Taylor & Caskie, 2014) and some have trained the rapid recall of words using flashcards and measured its influence on passage reading (Martin-Chang et al., 2007; McArthur et al., 2013; Nist & Joseph, 2008; Tan and Nicholson, 1997; Watts & Gardner, 2013). The majority of research that has sought to train sight words has used content words from the passages they presented in their experiment (Martin-Chang et al., 2007; Nist & Joseph, 2008; Tan and Nicholson, 1997). However, the majority of classrooms in New Zealand train High Frequency Words (HFW) as part of the National Standards for Literacy Achievement which require that children are able to recognise the first most common HFWs in their first year at school (Ministry of Education, 2009). HFWs are those words that occur most commonly in written material and are viewed to be different to the less commonly encountered content words that are trained by the majority of past research (Fry, 1980; Watts & Gardner, 2013). In order to fill this gap, the current study will seek to investigate the effect HFW training has on novel word reading accuracy, in-text word reading transfer, and passage reading accuracy, speed, and comprehension.

A Statement of the Problem

Past research has used content words to investigate the utility of sight word training, making it difficult to generalise to HFW training in schools. The current study also aims to investigate whether sight word training using flashcards increases word and passage reading competencies. However, the current research is unique in its implementation of HFW training rather than content word training and its use of single case research design. Both modifications to past research were aimed at increasing applicability of the study to the classroom context.

Nature of the Study

Two experiments were carried out to investigate the influence of sight word training using HFW flashcards. Single case research methodology was applied for both experiments. Experiment 1 was a single case design with one participant and Experiment 2 was a single case multiple baseline experimental design with five participants.

Single case research methodology is widely used in the educational and behavioural sciences as a way to carry out research that informs evidence-based teaching and learning strategies (Gast & Ledford, 2014; Horner et al., 2005). Single case design is a holistic category of research designs that encompasses different types of withdrawal and reversal designs as well as multiple probe and multiple baseline designs (Gast & Ledford, 2014; Smith, 2012). It was used in the current research to determine whether manipulation of one variable (HFW identification) would cause changes to the other variables of interest (HFW reading accuracy, word reading transfer, and passage reading accuracy, speed, and comprehension).

The differentiation between single case design and case study design is noteworthy. Case study designs are largely in-depth observations of a single participant whereas single case designs have predetermined research objectives within an experimental research design and may involve more than one participant (Gast & Ledford, 2014; Smith, 2012).

The small samples used in single case designs are appropriate for investigating the effectiveness of teaching strategies as they allow detection of individual differences in receptiveness to the intervention (Gast & Ledford, 2014; Horner et al., 2005). Experiments with larger samples, aimed at making inferences about the general effect of the intervention, usually

miss the individual differences in receptiveness to the intervention (Horner et al., 2005). This is counterintuitive to the current direction of educational practices in the classroom that emphasises individualised education plans for struggling learners (Gast & Ledford, 2014; Horner et al., 2005). Additionally, single case research designs are easily replicated in the classroom due to their small sample sizes and simple data analysis procedures such as visual analysis (Kratochwill et al., 2012). The current research has applied single case designs due to these advantages that increase the applicability of the research to the classroom context.

A multiple baseline research design was used for Experiment 2 to control for the contingencies of *carryover* and *demand effects* that may be present in single case designs with one participant. A *carryover effect* is the influence of one experimental treatment on subsequent experimental phases (Marks et al., 2012). A *demand effect* is the possibility that participants guess the experimenter's intentions, which leads to a conscious or unconscious change in behaviour (Marks et al., 2012). These contingencies were controlled for with the multiple baseline design by staggering the onset of the intervention for each participant (Marks et al., 2012; Smith, 2012). This means each participant started the intervention at a different time, strengthening any inferences made regarding the effects of the intervention condition by differentiating it from other contingent variables. Contingent variables are variables other than the manipulated variables (i.e. flashcard drill training) that may inadvertently influence the outcomes (Marks et al., 2012). These may be factors such as the day of week the intervention was initiated, the content of the story at which the intervention was initiated, and/or the difficulty level of the initial story. An intervention that manages to display a significant effect through this

method would have superior *internal validity*, which refers to the extent to which contingent factors are ruled out (Bulté & Onghena, 2008; Maggin & Odom, 2014).

The Purpose of the Study

The primary purpose of Experiment 1 and 2 was to investigate how HFW training effects single word recognition and passage reading competencies. The second purpose of Experiment 1 and Experiment 2 was to gain an understanding of the social validity of single word training as rated by the students. The third purpose of experiment 1 was to give insight into the research practicality of applying a similar within-subject design with a larger number of participants in Experiment 2. This included whether the drill training procedure could be completed within the allocated time period of 20-30 minutes and whether it was suitable for students 8-10 years old at reading levels 12-16.

Primary Research Questions

1. Will the intervention using flashcard drill training enable children reading below their peers to improve their accurate identification of HFWs both in isolation and when reading those same words in the text?
2. Will the intervention also lead to significant improvements in passage reading accuracy, fluency, and comprehension?

Implications for Teaching in New Zealand Classrooms

In New Zealand, the National Standards for Reading prescribed by the Ministry of Education emphasises the ability to recognize the most common HFWs in the first year of school (Ministry of Education, 2009). This requirement is usually facilitated by classroom teaching practices through training using HFW lists or single word cards. The extent to which this training transfers to passage reading and to reading comprehension is unclear in the literature. The current research will, therefore, contribute insight into the effects of HFW training on single word reading accuracy, word reading transfer, and overall passage reading accuracy, speed, and comprehension.

Summary

Chapter 1 was the introduction to the current research study. It has given the general background of the problem, the nature of the study, and the purpose of the study. The primary research questions and implications for New Zealand have also been outlined. The two experimental designs used to answer the primary research questions were described as Experiment 1, a single case design with one participant and Experiment 2, a multiple baseline design with five participants.

The chapter made explicit the need for additional evidence to fill the gaps in the current research regarding HFW training and the application of this training in the classroom context. It described training in word recognition as pivotal for preventing the difficulties of reading fluency described by Perfetti and Lesgold's (1979) widely mentioned bottleneck hypothesis.

Fluency is thought to influence overall reading comprehension (Ehri, 2014; Tan & Nicolson, 1997) as well as indirectly increase the frequency in which reading is independently attempted by children by making reading a more enjoyable experience (Walczyk & Griffith-Ross, 2007; Stanovich, 1986). The chapter also highlights the role of the current research in adding to evidence-based practice in the classroom.

The following Chapter 2 will provide a review of literature related to the conceptual underpinnings of reading difficulty and experimental literature that has applied single word training to increase reading skills. It also mentions literature related to the use of single case designs to give context to the research approach chosen for the current study.

Chapter 3 will outline the methodology, results, and discussion for Experiment 1. Chapter 4 will outline the methodology, results, and discussion for Experiment 2. The final Chapter 5, will conclude with the aims, contributions, and implications of both experiments.

Chapter 2 - Literature Review

The purpose of the current research is to provide an investigation of HFW training, commonly applied in classrooms throughout New Zealand. This chapter will support this research by outlining the underlying theoretical understandings of reading and how these theories explain the contribution of single word identification on reading fluency and comprehension. It will also mention theories that outline the utility of single word training and will review past literature that has attempted to train single words. The arguments pertaining to the transfer of this training to reading capabilities will also be discussed.

Acquisition of Word Reading Skill

One of the main theoretical foundations for the current research will be Ehri's (2014) sequential phases of reading development. Ehri's (2014) describes four phases of reading acquisition; *pre-alphabetic*, *partial alphabetic*, *full alphabetic*, and *consolidated alphabetic*. The central relevance of this theoretical framework to the current research is its contextualisation of *sight word* knowledge as one part of the reading acquisition process. *Sight words* are words that can be identified rapidly by sight and are conceptualised to develop as early as the pre-alphabetic stage but the capacity to build sight word vocabulary is limited by phonological awareness, developed in the full alphabetic stage.

Pre-alphabetic reading ability. The pre-alphabetic phase in Ehri's (2014) theory describes the skill young children may have to recognise familiar words in their environment, such as street signs or shop names, prior to instruction towards alphabetic rules. Consistent with

this, a literature review carried out by Neumann, Hood, Ford, and Neumann (2011) found that the majority of studies that have investigated the skills displayed by pre-readers have identified that participants could recognize a number of common words in their environment before alphabetic/phonological instruction.

An early study conducted by Ehri and Wilce (1985) gives specific insight into whether this visually based identification is void of any phonological processing. In their investigation of early reading skills, Ehri and Wilce (1985) compared the responses of prereaders, novel readers, and experienced readers after word training. The first phase of this research involved training the association between orally known words to novel groups of letters (i.e. qDJK was trained to represent the word *pencil*). Half of these novel groups of letters were phonetically irregular but visually distinct (e.g. qDJK for pencil) and the other half were words with phonetic spellings (e.g. PNSL for pencil). The authors found that prereaders identified visually distinct but phonetically irregular words more accurately than words with phonetic spellings, whereas novel and experienced readers identified words with phonologically regular spellings more accurately than visually distinct words (Ehri & Wilce, 1985). This suggests that the ability to identify words prior to alphabetic instruction may not involve phonological processing. It also suggests that in the next phases of word recognition called the *partial* and *full alphabetic* phases, children increasingly rely on phonological processing for word recognition.

Phonological awareness and decoding ability. Phonological awareness is the awareness of letter sounds in speech whereas decoding is the ability to apply this knowledge to reading practice (Ehri, 2005). Specifically, the former represents to ability to distinguish different sounds within words used in oral language and the latter represents the application of letter-sound (i.e.

grapheme-phoneme) correspondence to reading and then searching memory for familiar words with similar letter combinations that also fit into the given context (Binamé & Poncelet, 2016; Lai, Benjamin, Schwanenflugel & Kuhn, 2014; Ehri, 2014).

Ehri (2005, 2014) suggests that phonological processing ability enables children to use more efficient mnemonics to learn sight words. Instead of using non-alphabetic visual cues to remember and identify words, readers use phonological information within the word. For example, the word *look* may be memorized by a pre-reader by making the connection between the shape of the word to its meaning, as observed from the pre-readers mentioned in the research by Ehri and Wilce (1985) above. However, this same process would not be useful for memorizing and identifying thousands of words in a text without error, in which letter-sound knowledge serves as a much more effective mnemonic for remembering sight words (Ehri, 2005). This knowledge of letter-sound connections leads to *partial, full alphabetic, and consolidation* phases of reading. Sight words developed in these phases are termed *cipher sight words. Cipher sight words* are described as words that can be identified automatically as whole units based on a reader's knowledge of letter-sound connections rather than guessed based on context (Ehri, 1992, 2005).

Unitisation. Unitisation describes the rapid recognition of whole words (Ehri; 2014; van den Boer & de Jong, 2015). It describes the processing of whole words as synonymous to processing single digit numbers or pictures (Ehri, 2005). Some literature refers to this as *parallel processing*, where each letter in a word is processed simultaneously as opposed to *serial processing* that describes letter-by-letter processing (van den Boer & de Jong, 2015). A study conducted by van den Boer and de Jong (2015) investigated this concept by comparing the

recognition of single number digits, common words, and non-words. The study recruited 314 Dutch students from 2nd to 5th grade and found that younger students processed common words with the same speed as non-words whereas older students recognized common words as quickly as single digit numbers. Similar reaction times of word identification to single digit identification is theorised to represent the presence of unitisation/parallel-processing as a process that represented rapid and automatic word recognition (Ehri; 2014; van den Boer & de Jong, 2015). The study suggests that younger students processed words through serial phonological processing whereas older students processed words as whole units. This suggests a shift in cognitive strategies used for word recognition.

Ehri (2014) describes four different routes of word identification: *phonological recoding*, *analogizing*, *prediction,* and *memory* routes. The first *phonological recoding* route refers to the application of phoneme-grapheme rules. The second *analogizing* route refers to the use of familiar letter patterns to identify novel words. The third *prediction* route refers to the use of contextual information to guess novel words. Finally, the fourth *memory* route is the process that describes the identification of familiar words based on repeated exposure. Words identified automatically as whole units are indicative of the use of memory routes for word retrieval.

Experienced readers are said to have a number of these sight words in their mental lexicon which they can identify automatically from memory as whole units rather than through decoding, analogy, or guessing (Ehri, 2014; Tønnessen & Uppstad, 2015). Automaticity of word identification is suggested to allow readers to make maximum use of cognitive capacity to comprehend text (Lai et al., 2014; LaBerge & Samuels, 1974). The association between automaticity, cognitive capacity, and comprehension are elaborated on in further sections.

Orthographic Mapping

The orthographic structure of a word refers to its exact spelling. The theory of *orthographic mapping* highlights the role phonological awareness, decoding, and unitisation play in the development of sight words (Ouellette, 2010). The word *mapping* describes the simultaneous connection between the spellings, pronunciations, and meanings of words in memory (Ouellette, 2010; Ehri, 2014). Once these connections are established, they allow the reader to identify and retrieve the meanings of words rapidly through memory without going through the other routes of identification such as *phonological recoding*, *analogising*, and *prediction* mentioned above. Proficient readers are said to have detailed orthographic mappings of a number of common words in memory (Binamé & Poncelet, 2016; Bosse, Chaves, Largy, & Valdois, 2015; Tønnessen & Uppstad, 2015).

Research that looks into the role of phonological awareness and decoding in developing orthographic mapping may give insight into the vital components of sight word development (Apel, Thomas-Taite, Wilson-Fowler, & Brimo, 2012; McArthur et al. 2013; Ouellette & Fraser, 2009). For instance, a study by Ouellette and Fraser (2009) highlights the role of phonological awareness in developing detailed orthographic representations of sight words. These authors gave 35 4th grade children novel non-word training to determine how this would influence the development of orthographic representations. Their multiple regression analysis revealed that phonological processing was a significant predictor of the development of accurate orthographic representations. Similarly, Apel et al. (2012) measured the development of orthographic representations of 46 kindergarten children after they were trained on a number of novel non-words. Spelling and word recognition tasks showed that children higher in initial word

recognition and decoding ability developed more detailed orthographic representations of trained words. Therefore, phonological awareness and the ability to decode the smaller units within a word seems to mediate the quality of orthographic mapping developed.

Further research from the branch of literature focused on orthographic mapping highlights the role of unitisation in orthographic mapping and may give further insight into the theoretical understanding of how detailed sights words are formed. One such study was conducted by Nation and Cocksey (2009) who implemented an experiment involving rapid word recognition with a group of 7-year-old students. In this experiment, students were asked to categorise words such as *ship*, with the embedded word *hip* in it, and *crown*, with the embedded word *crow* in it. Students were asked to either accept of decline the categorisation of each target words as a *body part*, *animal*, or *other object*. It was found that embedded words that were related to one of the other categories elicited slower reaction times. For instance, students were slower in rejecting the proposition that the word *crown* fitted in the 'animal' category than they were in rejecting it as within the 'body part' category. This suggests the whole embedded words were recognised rapidly, signifying the presence of unitisation, and the meanings of each word was also retrieved rapidly, signifying the presence of orthographic mapping. This, therefore, highlights the role of unitisation in the development of sights words with detailed orthographic mapping.

Single Word Automaticity for Reading Fluency and Comprehension

Reading difficulty is associated with a breakdown in one or more processes that are theorised to facilitate reading (Ehri, 2014; Tønnessen & Uppstad, 2015). One such breakdown is the inability

to identify words quickly and effortlessly, signifying the absence of automaticity (Samuels & Flor, 1997; Tønnessen & Uppstad, 2015). Automaticity describes the rapid, unconscious, and effortless processing of information (LaBerge & Samuels, 1974; Sprenger-Charolles & Siegel, 2013). Words that can be identified automatically are termed *sight words* characterised by their rapid identification. Automatic word identification is required for fluent and comprehensive reading (Pikulski & Chard, 2005; Kuhn & Stahl, 2003).

LaBerge and Samuels (1974) were the first to suggest that automatisation of word reading could be the bridge between decoding and reading comprehension due to the efficient use of attention and energy put into reading. This theory suggests that a readers' comprehension of text can be affected by how automatically the text is read. The majority of modern theorists agree with this and hypothesise that, with practice, word identification will require less and less mental capacity until it requires very little capacity at all (Sprenger-Charolles & Siegel, 2013; Therrien, 2004; Tønnessen & Uppstad, 2015). When this happens, automaticity of word identification has been achieved. The theoretical discourse around word automaticity, therefore, is built on the theory that short-term memory is involved in reading and this processor has a limited capacity (Muter & Snowling, 1998; Tønnessen & Uppstad, 2015). If word identification puts too much stress on this capacity, then it will be less available for reading comprehension.

One way to examine the automaticity of word identification is through the *Stroop effect*. This effect describes the phenomenon in which a reader cannot help but process a text within their visual field (Sprenger-Charolles & Siegel, 2013). The images below (Figure 1 from Ehri, 1987, p. 8) exemplify a *Stroop Task* in which children were asked to name images and ignore the

printed words on top of each image. These printed words were termed *interference stimuli*. The example illustrates how difficult it is for a skilled reader to avoid reading the text.

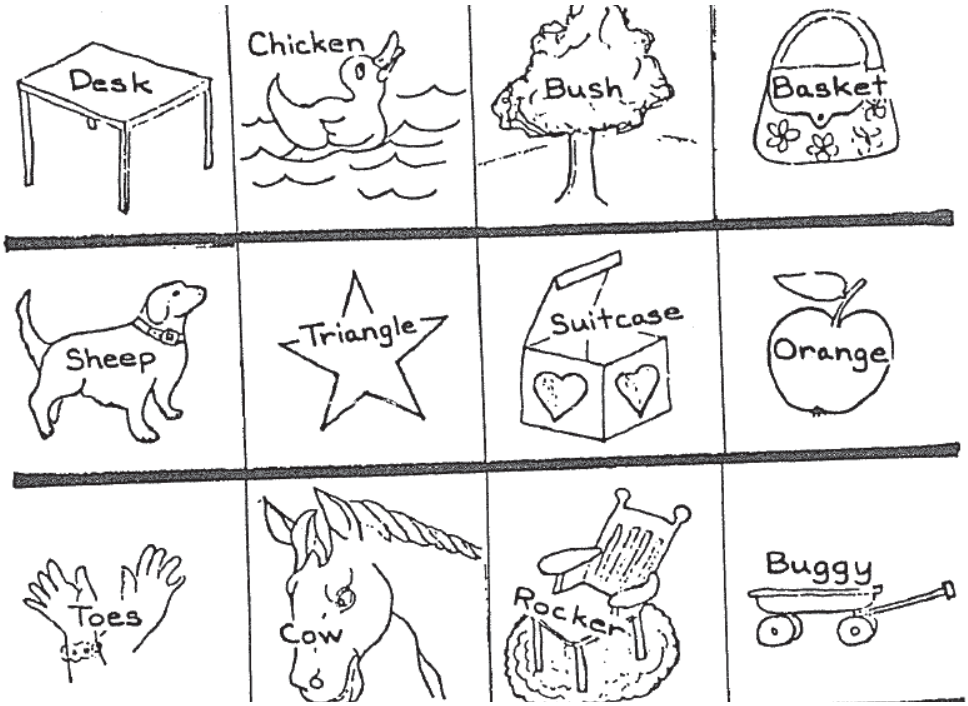


Figure 1. An example of a Stroop Effect task from Ehri (1987, p. 8).

In a study conducted by Guttentag and Haith (1978), 60 children and university students were given a Stroop Task in which they had to categorise images on a screen and ignore the words in the centre of the image. The three types of categories used were *animals*, *furniture*, and *modes of transportation*. One task involved choosing what categories the images on a screen belonged to while the word used as *interference stimuli* were related or unrelated. Figure 2 gives an example *of intra-categorical and inter-categorical* stimuli presented as a Stroop task by Guttentag and Haith (1978, pp. 710).

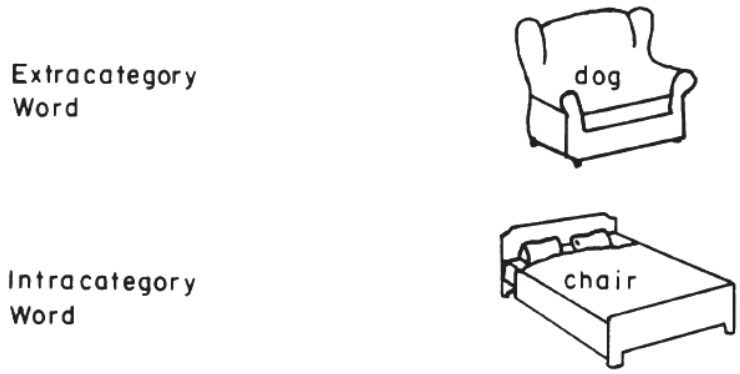


Figure 2. An example of Stroop task in Guttentag and Haith (1978, pp. 710).

The authors found that the reaction times for both poor and good readers, who had had at least 9 months of explicit instruction in letter-sound connections, were slowest under the unrelated word condition. For instance, the above image of a chair with the word *dog* printed on it resulted in a slower reaction time to if the word *bed* was printed on it. Differing reaction times suggest that the interference condition was successful in inducing unconscious processing of words. The results suggested that words and their meanings were processed automatically without conscious effort.

The mechanisms through which word recognition automaticity leads to reading fluency and comprehension may be made clear through an example of automatic problem solving in mathematics. A study conducted by Dennis, Sorrells, and Falcomata (2015) compared the effectiveness of two different interventions aimed at assisting children with mathematics learning disabilities to solve problems. One was an automatic memory based mathematics intervention using drill training and another was a number sense intervention using an analytic problem-solving approach. The participants were randomly allocated into two groups and the order in

which they received the two interventions was varied. The findings were that the number sense intervention initially increased the scores and extensive practice provided by the drill training intervention further added to problem-solving competency. The authors suggest that the initial number sense intervention allowed students to solve basic problems and automatisation of these basic processes through training allowed the students to progress on to more complex problems. Therefore, in using this study as an analogy for automaticity in reading, it can be said that automatic memory systems allowed students to effortlessly achieve basic cognitive tasks, freeing up cognitive resources for more novel or complex tasks.

Perfetti and Lesgold (1979) describe the inability to achieve automatisation through their bottleneck hypothesis. This early and widely referred to theory explains how the inability to move beyond slow decoding of sounds in words toward identifying words quickly by sight may lead to laborious and non-fluent reading. Without instruction, children who experience this bottleneck may fall behind in developing reading skills due to decreased motivation and excessive effort required to read (Stanovich, 1986; Walczyk & Griffith-Ross, 2007). For instance, in Walczyk and Griffith-Ross's (2007) review of fluency and comprehension research, they highlight the role of fluency in developing reading comprehension and the role of comprehension in increasing reading motivation. Based on this, the ability to progress past letter-by-letter decoding skills onto rapid word recognition seems to be pivotal for reading comprehension and readers' self-motivation to engage in further reading exercises. A number of prevention and intervention instructional strategies have been researched to allow readers to progress past letter-by-letter decoding skills onto rapid word recognition (McArthur et al. 2013; Marulis & Neuman, 2010; Tan & Nicholson, 1997; Therrien, 2004; Watts & Gardner, 2013).

Some of these have involved training students to read words faster through the use of flashcards (McArthur et al. 2013; Tan & Nicholson, 1997; Watts & Gardner, 2013).

Review of Word Training Studies

Bosse et al. (2015) found evidence for the effectiveness of training whole words for the development of detailed orthographic mapping in memory. In their study, 88 grade 5 and 6 students were trained on a set of non-words presented either as whole words or in parts. They found that training whole words led to more efficient orthographic learning in that it led to the most detailed recall of letters strings with less one-on-one training. It was suggested that the link between whole words and their meanings facilitated word acquisition.

Watts and Gardner (2013) implemented an intensive HFW intervention in order to compare its effectiveness with a pre-existing phonics intervention. In this study, the authors monitored the progress of eight Year 1 children over five weeks in a school in the UK. The intervention involved training 113 HFWs using flashcards. Baseline measurement using the Salford Sentence Reading Test (SSRT) showed that only three out of eight children met the reading standard for their age prior to the intervention. After five weeks of HFW training, only one pupil remained below the standard for their age. The study did not look at transfer to passage reading comprehension or speed but it did establish the effectiveness of word training to improve passage reading accuracy. However, the effect of the HFW intervention could not be differentiated with the pre-existing phonics program.

McArthur et al. (2013) also compared the effectiveness of a phonics program and a HFW intervention but the order in which each intervention was administered was varied. Participants

were 104 children with dyslexia between the ages of 7 and 12, divided into three groups. Group 1 received HFW training first, group 2 received phonics training first, and group 3 received both interventions at the same time. HFW training took five training sessions per week over eight weeks. Each session trained approximately 24 words using flashcards. Pre and post-training recognition of the 24 target words was probed using flashcards in the beginning and end 5-10 minutes of each training session. In the first phase, in which group 1 and 2 was exposed to either sight word training or phonics training, sight word intervention training led to the higher passage reading accuracy scores than phonics alone. In the second phase of the experiment, when both groups had received both interventions, reading ability was increased more when phonics training was followed by sight word training rather than the other way around. Therefore, this suggests that HFW training led to better results than phonics training alone when only one intervention was implemented. When both were implemented, phonics training followed by HFW training had the most impact on reading ability.

Tan and Nicholson (1997) investigated the influence of single word training on reading accuracy, speed, and comprehension. A randomized block design was used with matched pairs of 42 children aged 7-10 years. Each child was allocated into one of three groups: single word flashcards, sentence training, and a control group. The participants who received training using single word flashcards were trained to rapidly identify 10-25 words orally up until they were taking no longer than one second to respond to the visual representation. For participants in the sentence training group, each word was presented on a short sentence card such as *I like raspberry jam on bread*, with *raspberry* as the target word. Both training conditions involved repeated drill training where each card (containing either a single word or sentence) was

presented repeatedly in rotation. The control group was given the word in isolation without the drill training procedure. The meaning of each word was expressed through a short phrase contextualising the word and was given to each participant regardless of their conditional grouping. Participants were then asked to read a passage containing the target words, after which their accuracy, speed, and comprehension were measured. Accuracy represented the number of words identified correctly over the total number of words present in the passage and speed represented the number of words correctly identified per second. Performance on five open ended questions made up the comprehension scores. It was found that participants who received either form of drill training had significantly higher speed, accuracy of list reading, and performance on comprehension questions than the control group.

Word Training and Transfer to Passage Reading

An alternative to sight word instruction is self-teaching through reading experience (Share, 1994, 2004). The *self-teaching hypothesis* theorises that learners acquire new words and consolidate known words through decoding words when first encountered and eventually developing the orthographic mapping of those words through repeated exposure in text (Nation & Snowling, 1998; Share, 2004). This is based on evidence that readers with both high and low ability identify words more accurately when they encounter them in context when compared with encountering them in isolation (Nation & Snowling, 1998). Based on this, self-teaching and contextual learning are viewed to be more efficient for skill retention than sight word training using flashcards (Archer & Bryant, 2001).

A study by Martin-Chang et al. (2007) highlights the contextual learning benefits of self-teaching. This study found no difference in the acquisition and maintenance of words trained in isolation versus in context but did find that words trained in context lead to higher transfer scores. This study used a within-subject design with 69 participants who received word training in isolation as well as in context. The isolated words were presented on flashcards and the words in context were presented in short stories made up of 200-300 words in which each target word was repeated at least twice. Maintenance and transfer of isolated word training and contextual training were assessed both in list form and application to reading. They found that learning the target words in context led to higher transfer compared to learning them in isolation on flashcards but there was no difference in word acquisition and maintenance scores. This research supports the notion that contextual learning may lead to better transfer to in-text reading, which is an advantage of self-teaching but does not give evidence for the other aspects of self-teaching such as the independent acquisition of these words.

Stanovich's (1986) concept of *Matthew Effects in Reading* suggests that the self-teaching hypothesis may be difficult to apply in practice for poor readers because it requires the independent acquisition of words. *Matthew Effects in Reading* describes the reciprocal relationship between the degree of effort put into reading and the development of reading experience. This suggests that, since children with poor reading skills may find the reading of text cumbersome and strenuous, this may perpetuate the avoidance of reading and increased reading deficit. This is supported by the study conducted by Leinonen et al. (2001) who compared the reading habits of adults who read slowly but accurately and those who read fast but inaccurately. The authors found that readers with relatively faster speed had better reading

habits, regardless of the numerous errors made in the text, seemingly because faster reading led to more rewarding reading compared to slow accurate reading. This study gives evidence for Stanovich's (1986) *Matthew Effects in Reading* by describing the association between reading difficulty and the motivation to build reading experience. It also specifically highlights the role of fluent reading in this association.

Sight Word Training and the Irregularity of English Spelling

One rationale for training sight words is that sight word vocabulary is necessary due to the irregularity of words that do not lend themselves to phonological analysis (Ehri, 2005; Ellis et al., 2004). Studies of orthographic depth have outlined a number of language systems that employ purely visual memory based word recognition (Ellis et al., 2004; Seymour, Aro, & Erskine, 2003). At one extreme, logographic orthographic systems such as Chinese and Japanese kanji directly link symbols to meanings and allow rapid recognition of whole symbols without analysis of letter-sound principles (Ellis et al., 2004). This postulates a purely visual memory system for word recognition.

At the other extreme are the language systems employed by transparent alphabetic orthographies such as Spanish, Dutch, and German with regular letter-sound correspondence which largely utilise auditory memory of oral language to comprehend text (Ellis et al., 2004). In between these two extremes is English with the highest number of irregularities among European languages. This suggests that the relative irregularity of English may justify the use of visually based training strategies such as sight word training with flashcards to increase reading competency (Seymour et al., 2003). With this in mind, the current research is based on Ehri's

(2014) stages of reading acquisition, which acknowledges the role of phonological awareness and decoding skills in the development of sight words.

Single Case Research Designs

This section on single case research accompanies the *Nature of the Study* section in Chapter 1 and contextualises the methodology used in the current study with previous literature. The literature mentioned in the above review includes both within-subject (Martin-Chang et al., 2007; Watts & Gardner, 2013) and between-subject experimental research designs (McArthur et al. 2013; Tan & Nicholson, 1997). Both have their advantages and disadvantages. A potential advantage of a within-subject design is its elimination of contingencies caused by differences between participants. This can be contrasted with between subject experimental studies such as that conducted by Tan and Nicholson (1997) which used randomly assigned matched pairing to reduce the chances of individual differences confounding their results. However, a within-subject design may accomplish this without compromising the statistical power of its findings. It does this by exposing each participant to both experimental conditions, where the participant acts as his/her own control, doubling the number of data points per participant available for analysis (Marks, Doctorow, & Wittrock, 2012). Therefore, a within-subject design may be best at making the most of a small sample size. Single case designs, as a form of within-subject research, also has these advantages.

According to Kratochwill et al. (2012), single case research designs and specifically, multiple baseline design, are useful for evaluating interventions that do not allow reversal and in instances in which it may be unethical to have a control group of participants. Single case

designs allow each participant to serve as their own control as the differences between their individual scores before and after the intervention are compared (Gast & Ledford, 2014;Horner et al., 2005; Kratochwill et al., 2012). Through this, individual responsiveness to the intervention can also be taken into account when deciding the effectiveness of the intervention under research (Horner et al., 2005; Kratochwill et al., 2012). As mentioned in Chapter 1, these factors were taken into consideration when deciding to use a single-case research design for the current research.

A disadvantage of the multiple baseline procedure is its lack of statistical power relative to randomised experimental designs with a control group (Kratochwill et al., 2012). In order to strengthen the statistical power of a multiple baseline design with a small sample size, a number of authors have suggested the use of regulated randomisation procedures (Ferron & Sentovich, 2002; Koehler & Levin, 1998; Wampold & Worsham, 1986). These procedures involve random assignment of each participant to a staggered point at which the intervention is initiated. According to Koehler and Levin (1998), this simple change to the experimental methodology can allow multiple baseline designs to meet the assumption of randomisation required for statistical credibility despite a small sample size. Therefore, Experiment 2 of the current study will aim to employ a multiple baseline design with a randomised staggered allocation of intervention onsets.

A disadvantage of randomisation methods is their difficulty of analysis (Bulté & Onghena, 2008). Researchers such as Bulté and Onghena (2008) and Ferron and Sentovich (2002) comment on the inappropriateness of widely used parametric analysis such as *t* and *f* tests for within-subject designs due to extreme violations to normality, homogeneity of variance, and especially, serial independence. These are the foundational assumptions of parametric analysis

and only research with a large sample and a control group may satisfy these assumptions. Widely applied nonparametric analyses are also problematic for within-subject designs such as rank tests and time series analysis as rank tests lack sensitivity when applied to small sample sizes and time series analysis requires a larger number of observations than what is usually possible with a multiple baseline design (Bulté & Onghena, 2008). Bulté and Onghena (2008) and Ferron and Sentovich (2002) outline a number of alternatives to analysis which involve both a manual computation of randomized test distributions and the use of the SCRT package specifically made for single case randomisation tests in R. R a is a programming language that allows statistical analysis and graphic display. The current research will make use of these strategies for analysis.

Summary

Chapter 2 highlighted the theoretical rationale for the present study. Ehri's (2014) phases of reading development, the concept of orthographic mapping, and automaticity theory were emphasised. A review of single word training strategies researched previously and the transfer of this training to in-text reading was also highlighted. The orthographic structure of the English language was also mentioned due to its relevance to single word training.

The current research seeks to add to the current body of literature by filling the gaps in the methodologies used previously. A gap in the past literature seems to be the lack of direct applicability to the classroom. The majority of the previous literature failed to measure the effect of HFW training on word reading transfer and comprehension. Studies that have trained HFWs have not measured transfer to passage reading accuracy and comprehension (McArthur et al. 2013; Watts & Gardner, 2013) whereas those that have measured comprehension (Tan and

Nicholson, 1997) and transfer (Martin-Chang et al., 2007) have used content words instead of HFWs. An analysis of the influence of HFW training on word reading accuracy, word reading transfer, and passage reading accuracy, speed, and comprehension would fill this gap.

Additionally, previous studies have used experimental and observational designs with a large number of participants, aimed to determine an average effect. A departure from this is the use of single case design. Single case designs are commonly used in behavioural and educational research to investigate individual responsiveness to an intervention. Single case multiple baseline designs have been used in the past to control for potential *carry-over* and *demand* effects that may arise with the use of other forms of single case designs. Authors such as Koehler and Levin (1998) have proposed a randomisation procedure that will enhance the statistical power of multiple baseline designs. They describe the application of a randomised staggered design as the cornerstone of statistical credibility for research with smaller sample sizes.

Chapter 3 - Experiment 1

Experiment 1 was a single case design carried out with one 8-year-old participant at Reading Recovery level 12. Its purpose was to answer the primary research questions:

1. Will the intervention using flashcard drill training enable the participant reading below his/her peers to improve in accurate identification of HFWs both in isolation and when reading those same words in the text?
2. Will the intervention also lead to significant improvements in passage reading accuracy, fluency, and comprehension?

The experiment's secondary purpose was to give insight into the social validity of the drill training process. The third purpose was the research practicality of applying a similar within-subject design with a larger number of participants in Experiment 2. The method mentioned below was the same as was conducted for Experiment 2 apart from the alterations mentioned in Chapter 4. A measure of social validity was also taken to give the student an avenue to express any distress caused by the intervention as well as to gain an understanding of the overall student acceptance of single word drill training.

Method

A single word drill training procedure was used within a single case research design in order to fulfil the aims of the research. One child was recruited for participation. He attended three baseline sessions and four training sessions. Single word training involved the use of flashcards

to aid rapid recognition of words. The rapid recognition of 10-15 novel target words was the aim of each 20-30 minute training session.

The effects of the training were examined by measuring four dependent variables associated with the primary research questions. The dependent variables associated with the first research question were termed *word reading accuracy* and *word reading transfer* while the dependent variables associated with the second research question were termed *passage reading accuracy*, *passage reading speed*, and *passage reading comprehension*.

Participant. The participant attended a decile two primary school in West Auckland, New Zealand (Ministry of Education, 2016). Decile levels represent the economic advantage of the students attending the school in which decile 10 represents the highest advantage and decile 1 represents the lowest advantage (Ministry of Education, 2015). In this school, 46.4% of children were originally from South Pacific countries, 41.9% were Maori, and 4.3% European/P**ā**keh**ā** (Ministry of Education, 2016).

Children suggested by the school were aged between 8 and 10 years at Reading Recovery (RR) levels between 12 and 16. According to the National Standards for reading published by the Ministry of Education (2009), level 12 is the benchmark for the end of year 1 (or 40 weeks of schooling), equivalent to children aged 6 years 0 months (summarised in Appendix B). Based on this, the majority of children nominated were reading two or more years below average for their age. The chosen participant for experiment 1 was an 8 years and 9 months old male reading at a Reading Recovery level 12.

Ethics. Low-risk ethics approval was granted by the Massey University Human Ethics Committee. Participation consent forms were sent out to the school principal and the parents of the nominated children. Only children who had school and parent consent as well as gave their own assent were included in the study (see documentation in Appendix A).

Design. Experiment 1 was a within-subject design with 3 baseline and 7 intervention sessions. The main measures of interest were word reading accuracy, word reading transfer and passage reading accuracy, speed, and comprehension.

Materials Used. The main material used for the training program were flashcards with HFWs printed on them. Additionally, several passages from the *Ready to Read* book series published by the Ministry of Education were used to measurement changes to passage reading competencies as a result of the training.

Flashcards. Fifteen to 25 HFWs were printed with a large font size on a card with the dimensions 9cm by 5.5cm. The HFWs used in this experiment were from a generic HFW list used at the participant's school. HFW lists are usually compiled by statistically analysing the most common words that occur in text (Watts & Gardner, 2013). A wide range of texts from fiction and fictional stories, newspapers, and articles are usually included in the analysis to form a list of most commonly occurring words (Dolch, 1936; Fry, 1980). The HFWs used in the current study were modified to match the form in which the word was presented in the story (i.e. the HFW *start* in Figure 3 was modified to *started* to resemble the form of the word that was present in the story).

Figure 3. Front of the flashcard.

A unique set of target cards were chosen for each session based on the HFWs present in the assessment passage and the HFWs noted as unknown prior to the intervention. Specifically, these target words were identified as unknown prior to the intervention by probing a list of words from the generic HFW list that were also present in the content of the assessment passage.

The novel set of 10-15 words were selected rather than having a mix of novel and known words due to research that has shown that training a complete set of novel words led to high levels of word acquisition per minute of instructional time (Joseph & Nist, 2006). This was chosen despite the possibility of the child may have perceived the task to be difficult due to the large number of unknown targets (Skinner, Fletcher, Wildmon, & Belfiore, 1996). A shorter instructional session, with a small number of novel target words, was prioritised in order to prevent boredom and cognitive strain caused by a longer session. Below is a sample target word list used for the passage from *Butterfly Day* (see Figure 4 below for associated passage):

love	Our	been	feel	morning	were
long	around	room	careful	clothes	head

Passages. The passages used to measure the effect of HFW training were all from the *Ready to Read* series published by the Ministry of Education and were 100-150 words in length. The *Ready to Read* series is graded based on vocabulary used in the text, length of the text, and complexity of the ideas (Ministry of Education, 2014). Based on these criteria, children 8 years and above should be at grade levels *purple*, *dark yellow*, *emerald*, and/or *white* on the colour wheel used for *Ready to Read* reading material (summarised in Appendix B). These colour grades were compiled by the Ministry of Education to allow educators to choose appropriate reading material for children at different levels of reading (Ministry of Education, 2014). Due to the large number of words in the *Ready to Read* series at this level, only the beginnings passages of each story were chosen.

Figure 4 is a sample passage from *Butterfly Day*, a book from the *Ready to Read* series written by Raymond (1998) and published by the Ministry of Education. This story is graded as light blue for guided reading and corresponds to reading level 6.5-7 (see Appendix B). The participant was asked to read pages 2- 4 (page 1 being the title page), totaling 159. The aim was to choose passages that were approximately 100-150 words in length without obstructing the flow of the story content. This was accomplished by asking the participant to stop at the end of a paragraph rather than in a middle of a paragraph or sentence.

Figure 4. A Passage from Butterfly Day (Raymond, 1998, p.2-4).

In total, the current experiment used four texts graded as purple (7.0-7.5 years' level) for *guided reading* and three at the light blue (6.5-7.0 years' level) grade for *guided reading* (one level below purple) (summarised in Appendix B). The level appropriate for *guided reading*, as opposed to the levels prescribed for *individual* or *shared* book reading, was selected as most appropriate given the procedure used involved the researcher prompting the participant during the reading exercise.

Dependent Measures. Single word reading accuracy was measured by noting the total number of words read correctly on the HFW cards presented. Transfer to passage reading was measured by adding up the number of words read correctly in the passage that were also target HFWs presented on word cards.

Passage reading accuracy was determined by dividing the total number of words correctly identified by the total number of words in the passage (approximately 100-150 words). Speed was measured by determining the total time taken to read the passage, divided by the total number of words in the passage.

Reading comprehension ability was assessed with five to six multiple-choice questions asked after passage reading. Three of these questions were *explicit* and two to three were *implicit*. Questions related to both explicit and implicit information were asked to assess different levels of student understanding. Explicit questions required answers directly found in the text whereas implicit questions required students to make inferences by integrating the text content with prior knowledge (Duke & Pearson, 2008). Each question was presented orally in a multiple-choice format with three options. Multiple choice questions were used as an alternative to open-ended questions in order to reduce the stress of assessment that may arise when children with weak language ability, or self-perception of ability, are asked to orally express their ideas (Brookhart, 2015). It also provided a clear marking schedule which was not open to interpretation by the researcher. All the questions and multi-choice options were constructed based on guidelines published by Brookhart (2015).

Procedure. The baseline and intervention sessions were conducted individually in a quiet room at the school. Each session took 20-30 minutes. The one-on-one sessions were explained to the child in the following way:

"During this session, we will first read a list of words, then read a part of a story, then answer some questions about the story. If there is any word you don't know, try to sound it out and take your time and I'll help you if you get really stuck". Extra prompts were provided during drill training such as "We've gone through the list twice, so now let's try to go as quickly as we can".

The child was initially presented with approximately twenty to 20-25 HFWs that were present in the HFW list and the chosen passage for any given session. These were not presented in any particular order. The child's ability to identify the words was probed and the correctly identified words were stacked in one group while the incorrect words were stacked in another group. From the stack of incorrect words, 10-15 words were chosen as targets for training. For the baseline phase, the child was asked to read the target words and asked to "sounds out the word" if it was unknown. If the child was unable to do so, he was told the correct pronunciation of the word, usually followed by the participant repeating the correct response. He was then given a verbal example of the word in context (i.e. "*Walk around the puddle* for the target word *around*). The intervention phase involved the same process with an added drill training procedure.

During drill training, the child was asked to repeatedly read target words in rotation. Once the child was able to identify the words correctly independently, he was prompted to identify the

word faster as the researcher increased the speed in which the target words were rotated through. The order in which the cards were presented was changed at each rotation. Each word was considered mastered/known once it could be identified correctly within one second at least two times. Rather than repeating one word or a small group of words until mastery, the complete set of target words were trained together. This was done to decrease the possibility of the participants utilising maladaptive memory strategies to memorise the target words (Kornell, 2009).

This is supported by research conducted by Kornell (2009) and Griffin and Joseph (2015) which highlight a *spacing effect* when investigating the usefulness of different flashcard training methods. Specifically, this research suggests that spacing the time between exposures through using a larger pile (i.e. 10-20 cards) of training cards was more effective at facilitating word acquisition than training four smaller piles of training cards (i.e. four sets of five cards). According to Kornell (2009), the use of variable memory heuristics and attentional factors are two explanations of these observed *spacing effects*. When smaller sets of cards are used, the participant may be inclined to use surface memory cues such as the first letter of the word, whereas the variety of words provided by a large set of cards may lead to more detailed memory cues (Ehri, 2014; Kornell, 2009). This way, the participant is also required to pay greater attention to all the letters in each word rather than just a few. For these reasons, a complete set of 10-15 HFWs words were targeted and were ordered randomly at each rotation.

Once the participant had completed drill training, they were asked to read a part of a passage from the *Ready to Read* series of books such as seen in Figure 4. The purpose of this was to deduce changes to passage reading accuracy, speed, and comprehension caused by the

training, similar to the reading fluency assessment recommended by Hudson, Lane and Pullen, (2005). The students were given affirmative praise intermittently throughout story reading and were given a partial or full verbal prompt if they were stuck on a word for more than 3 seconds. When a word was read incorrectly it was followed by a full verbal prompt of the correct response. The aim was to provide the least intrusive prompts while at the same time adapting these prompts with due consideration for the repercussions caused by prolonged pauses and lack of scaffolding to the child's comfort and self-esteem (Kaniuka, 2010).

The comprehension questions consisted of five to six implicit and explicit multiple choice questions. The researcher read out each question to the participant and then presented the three different answer options. As these were presented orally, the researcher used three fingers to represent each option to emphasise that they were three different options. The comprehension questions used and their options for the *Butterfly Day* passage in Figure 4 were:

1. What kind of plant do the caterpillars in the story eat?

a. Flowers

b. Swan plant

c. Rosemary

2. What were they doing when they noticed the butterfly had hatched?

a. In maths class

b. Playing outside

c. Singing a song

3. What did they see the butterfly doing after it came out of its chrysalis?

a. eating

b. flying out the window

c. slowly opening and shutting its wings

4. Why do you think the butterfly was clumsily going around the room?

a. Because it had just hatched

b. Because it was missing its family

c. Because it was scared of the children

5. Why did the children have caterpillars in their classroom?

a. because they were interested in how caterpillars become butterflies

b. because they wanted to cook them

c. because their class was in a garden

In summary, the procedure for Experiment 1 had seven different components:

1. The child's prior knowledge of the HFWs was probed by asking him to read aloud approximately 20-25 words contained in the passage.
2. From these, 10-15 incorrectly identified words were selected for training with flashcards.
3. These target words were re-presented to the child with a verbal example to contextualise them. Incorrect responses were prompted.
4. Drill training was carried out (only for the intervention phase).
5. The child was asked to read a passage from *Ready to Read* materials.
6. Five to six multi-choice comprehension questions were asked.
7. The child was asked to complete a feedback form (mentioned in detail below).

Social validity. Feedback questionnaires were given to the participant to ascertain how he felt about the drill training intervention. The purpose of this was, firstly to confirm that he did not feel distressed as a result of the training and secondly, to give an overall idea of student

acceptability of drill training. The participant was asked to fill out an evaluation form which was made available to the classroom teacher to view. The researcher read out five different questions: *How did you feel about the lesson, How did you feel after reading the story, How do you think your tutor feels when you read a story* and *Would you like to come for another lesson?* The child was asked to rate his response on a series of five faces expressing different emotions. Figure 5 below is an example of question 1 and the full form can be found in Appendix C.

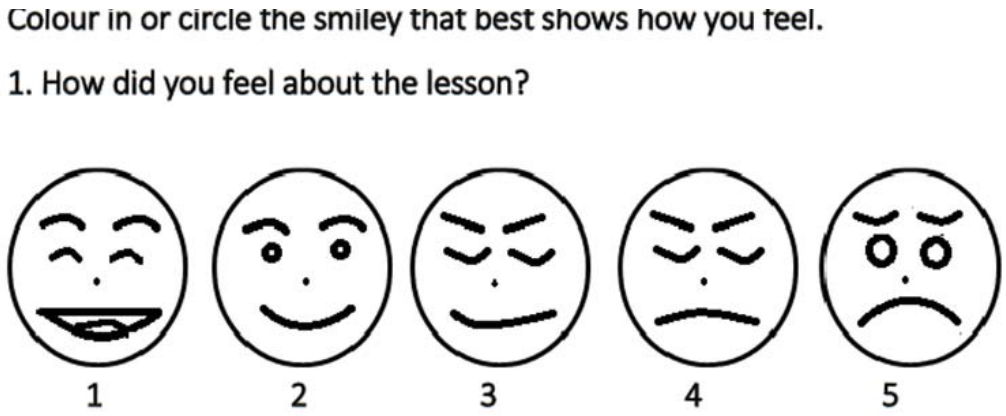


*Figure 5.*Question 1 from the feedback form used.

Reliability. In order to examine inter-observer reliability, another person was asked to listen to all the recordings of the passages read by the participant and circle the words that were read incorrectly. Inter-observer reliability was measured by dividing the total number of agreements with the total number of disagreements and multiplying by 100.

All the comprehension questions and multi-choice options were reviewed by the researcher's supervisors to check for appropriateness to the story content and to the age group involved. Appendix D has another sample of questions and options.

The student feedback questionnaire was reviewed by the school principal and the participant's teacher to be sure that the questions were appropriate and easily understood.

Data Analysis. The dependent measures of interest were: single word reading accuracy, word reading transfer, and passage reading accuracy, speed, and comprehension. A comparison was made between these measures before and after the implementation of drill training using visual analysis and Cohen's d classification of significance. Visual analysis was used as the primary mode of analysis whereas Cohen's d was the secondary mode of analysis used to confirm the visual analysis. Essentially, the Cohen's d effect size value was used to quantify the effect that was observed through visual analysis (Kratochwill et al., 2012).

Percentage accuracy of single word reading:

$$\frac{Total\ number\ of\ target\ words\ independently\ identified\ on\ HFW\ cards}{Total\ number\ of\ targetted\ HFWs} \times 100$$

Percentage transfer of single word reading:

$$\frac{Total\ number\ of\ target\ words\ independenently\ identified\ in\ the\ passage}{Total\ number\ of\ target\ HFWs} \times 100$$

Percentage passage reading accuracy:

$$\frac{Total\ number\ of\ words\ independently\ identified\ in\ the\ passage}{Total\ number\ of\ words\ in\ the\ passage} \times 100$$

Passage reading speed (words per second):

$$\frac{Total\ number\ of\ words\ read\ in\ the\ passage}{time\ taken\ (seconds)}$$

The speed measurement represented the time taken to read the whole passage. This included the words read incorrectly that were followed by prompts. The overall time taken therefore represents the overall difficulty experienced during passage reading rather than the specific time taken to read each word.

Passage reading comprehension (percentage correct questions):

$$\frac{Total\ number\ of\ correct\ responses}{Total\ number\ of\ questions} \times 100$$

***Visual analysis*.** Visual analysis is the traditional mode of analysis for single case research designs. One of the resounding limitations of visual analysis is the lack of standard rules from which visual trends are judged and the associated increased likelihood of both type I and type II error (Bulté & Onghena, 2008; Bulté & Onghena, 2009; Horner, Swaminathan, Sugai & Smolkowski, 2012).

A number of authors have attempted to address the need for decision-making standards for visual analysis. The current research will comment on the results obtained based on the review of single case research standards published by Kratochwill et al. (2012) and the visual analysis guidelines published by Franklin, Allison, and Gorman (2014). These publications highlight six features that determine whether a single case design depicts a causal relationship between independent and dependent variables.

When analysing the data obtained within a given phase (baseline or intervention phase), analysis of the *trend*, *level*, and *variability* of data points was examined (Kratochwill et al., 2012). According to Kratochwill et al. (2012), the *trend* refers to the slope of the line of best fit

across the data within each phase, the *level* refers to the mean scores within each phase, and v*ariability* refers to the variance observed that diverges from the line of best fit within each phase.

When analysing data obtained between phases, *overlap* and *effect immediacy* was examined. *Overlap* refers to the number of points that overlap between the phases whereas the *effect immediacy* refers to the difference in the level of scores obtained in the last session of the first phase and first three sessions of the second phase (Kratochwill et al., 2012). A minimum of three consecutive data points was used as a standard for judging whether an effect existed (Franklin et al., 2014).

***Cohen's d effect size*.** A limitation of visual analysis is the inability to obtain a quantified effect (Bulté & Onghena, 2009; Kratochwill et al. 2012). Guidelines for visual analysis may be used to make overall conclusions about the causal effect of the intervention but the comparison of the effect across participants and between research studies is limited without a quantified effect size (Bowman-Perrott et al., 2013; Bulté & Onghena, 2009). Due to this limitation, the current research analysis will also include Cohen's d effect size analysis which, according to Kratochwill et al. (2012) is commonly used to quantify the effects in single case study to support visual analysis.

Cohen's d classification of effect size significance was used to compare the measures of interest obtained in the baseline phase compared to the intervention phase. A review of single case research carried out by Bowman-Perrott et al. (2013) reports Cohen's d as one of the most

commonly used effect size estimate when comparing the effects of two treatments. It was calculated using the following formula:

$$Cohen's\ d = \frac{\bar{x}_{phase\ b(intervention)} - \bar{x}_{phase\ a(baseline)}}{Sd_{pooled}}$$

$$SD_{pooled} = \frac{\sqrt{(SD_{phase\ b}{}^2 - SD_{phase\ a}{}^2)}}{2}$$

A Cohen's d value of 0.2 is viewed to be a small effect size, 0.5 a medium effect size, and 0.8 or above a large effect size (Bowman-Perrott et al., 2013). It is recommended that effect size analysis is usually used in a confirmatory capacity in conjunction with visual analysis (Horner et al., 2005; Thomas, Salazar, & Landers, 1991).

Results

Visual analysis of graphic information and Cohen's d effect size values for each variable of interest will be the focus of the following section.

Word Reading Accuracy and Word Reading Transfer. The results of Experiment 1 showed a clear increase in word reading accuracy in the intervention phase. This is evident in the visual analysis of the black line in Figure 6 which shows two distinct non-overlapping trends of data obtained for the baseline phase compared to the intervention phase. A large effect size was also obtained (see Table 1). The large effect size and distinct visual change can be attributed to the fact that previously unknown target word lists (starting at 0-30% accuracy) were usually trained to mastery (increasing to 100% accuracy) during the training phase.

Conversely, the transfer of these words to in-text passage reading shows no differences between trends obtained in the baseline and intervention phase (depicted with the grey line in Figure 6), resulting in a small effect size in Table 1. This suggests that even though word recognition was achieved when read in isolation, the training did not transfer to reading these same words in the text.

Table 1.

Cohen's d effect sizes for each measurement

	Baseline		Intervention		Effect size
	Mean	SD	Mean	SD	
Word reading accuracy	23.77%	3.89%	94.11%	4.96%	15.46
Word reading transfer	50.50%	8.32%	53.85%	7.69%	0.38
Passage reading accuracy	62.72%	8.83%	72.84%	10.25%	1.05
Passage reading comprehension	63.49%	17.55%	83.93%	23.60%	0.99
Passage reading speed	0.2827	0.0561	0.3324	0.0722	0.77

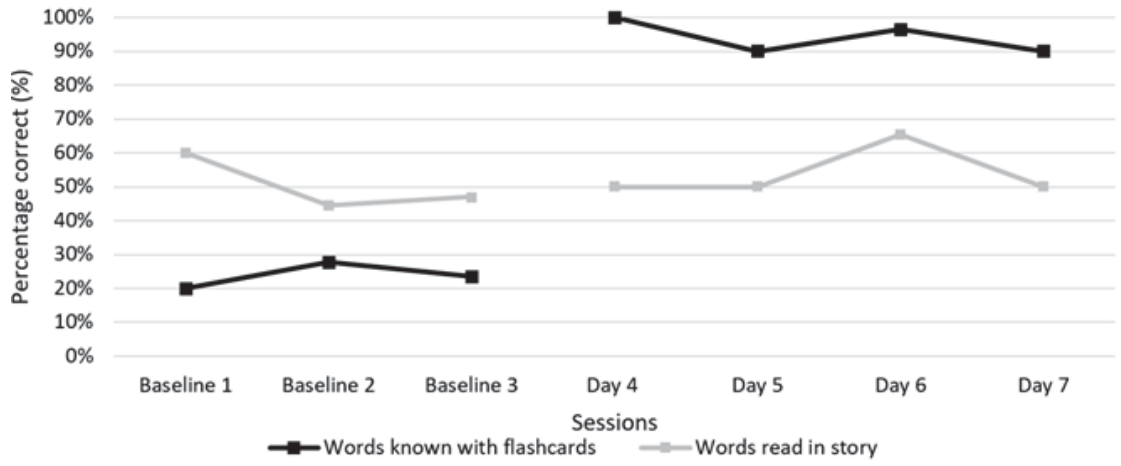


Figure 6. Word reading accuracy (black line) and transfer (grey line) per session.

Passage Reading Accuracy, Speed, and Comprehension. The measures taken for passage reading accuracy, speed, and comprehension all seem to show an increase in scores obtained in the baseline phase versus the intervention phase. Table 1 shows an average increase of passage reading accuracy by 10%, a 20% increase in reading comprehension, and a 0.05 second increase in speed after the intervention phase.

However, the lines of best fit in Figures 7-9 show increasing slopes in the data across phases, regardless of the point of phase change. This may reflect inherent increasing in these measures due to variables other than the effect of the intervention. Since the Cohen's d effect size calculation is determined by the difference of means between phases, the observed slope may have driven these values up. The validity of the effect sizes for word reading accuracy, speed, and comprehension depicted in Table 1 may be compromised due to this.

Visual analysis also shows variability within phases. The variability observed for comprehension scores was the highest with an average standard deviation of 17.5% at baseline and 23% in the intervention phase at seen in Table 1. Passage reading accuracy had the least variability with a standards deviation of 8.8% at baseline and 10.2% in the intervention phase.

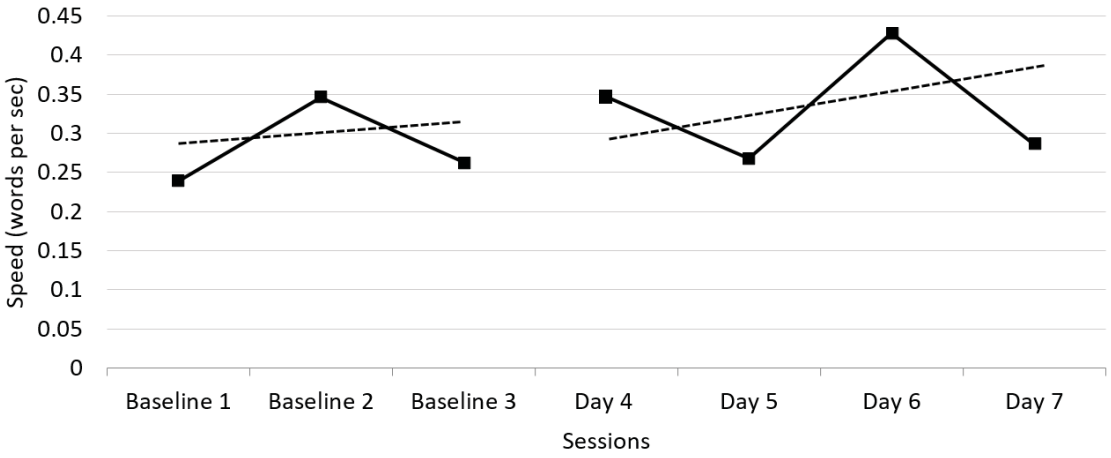


Figure 7. Passage reading speed across sessions.

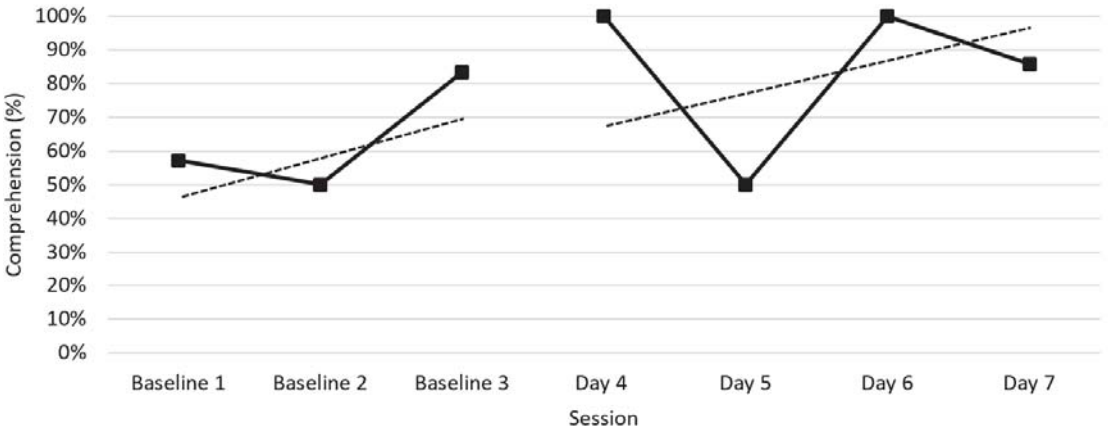


Figure 8. Passage reading comprehension across sessions.

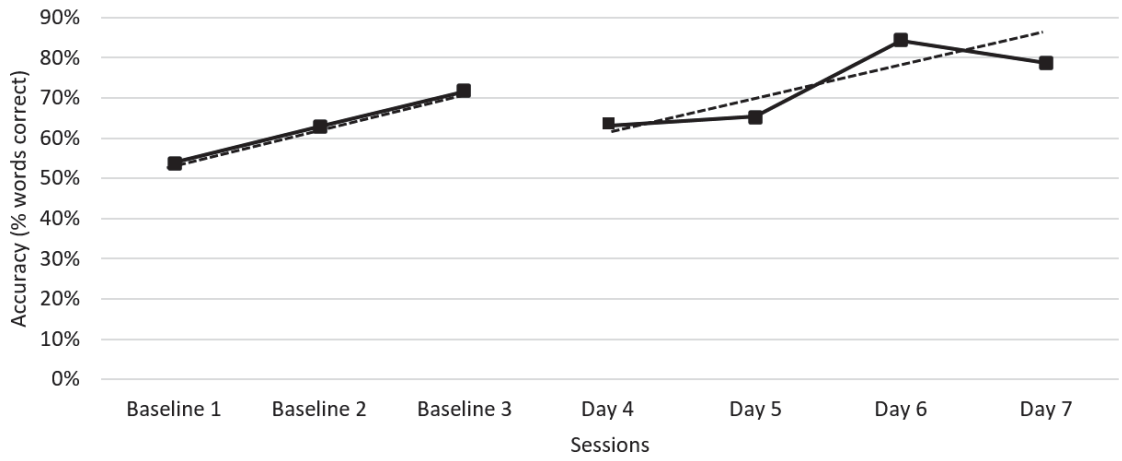


Figure 9. Passage reading accuracy across sessions.

Social validity. The participant was asked to fill out a feedback form after each intervention session. As seen in Figure 10, the participant chose the happiest or the second happiest face across all sessions for the first, second, and third questions (*How did you feel about the lesson, how did you feel after reading the story, and how do you think your tutor feels when you read a story?*).

He also picked the happy faces for the last question (*Would you like to come for another lesson?*) for 2 out of 4 sessions. During his last two intervention sessions, he picked the saddest two faces for question 4. This suggests that the participant enjoyed the majority of the intervention but was not eager to come for another lesson. This may have been because the participant was reading at below the average for his age and the lesson may have been difficult for him even though he enjoyed the lesson.

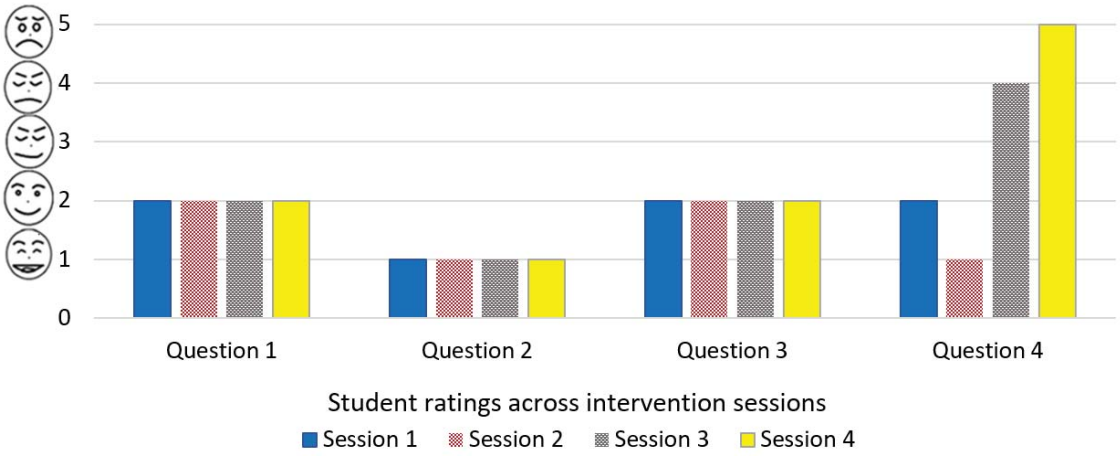


*Figure 10.*Child feedback on questions 1-4 across four intervention sessions.

Summary. The results of the first research question, which were related to word reading accuracy and transfer, showed a dramatic improvement in word reading accuracy in isolation and a very large effect size. In contrast, the training had only a small effect on the transfer of these same words in the passage. So the answer to the first research question was that the word training effects were limited to words read in isolation.

The results of question 2, which were related to passage reading variables, showed that after word training, passage reading accuracy increased by 10% on average, comprehension increased by 20% on average, and speed increased by 0.05 seconds on average. Cohen's d effect sizes for all three variables were large. Visual inspection suggests the large effect sizes observed may be due to overall increasing trends in the data regardless of the point at which the intervention was applied. Essentially, the lines of best fit for accuracy, speed, and comprehension all show an increasing trend, making it difficult to make substantial claims regarding the answer to the second research question/s.

The results from the social validity questionnaire showed that the child was very happy about the lessons and felt good after reading the passages. When asked whether he was happy to attend another lesson, the child gave a positive response for 2 out of 4 lessons.

Discussion

Experiment 1 with one participant was implemented to determine whether training of HFWs would lead to significant improvement in word reading accuracy and transfer and passage reading accuracy, speed, and comprehension. Its secondary purpose was to gain an understanding of the social validity of the word training. Its third purpose was to inform the procedure used for Experiment 2. Seven sessions with durations of 20-30 minutes were undertaken. In each session, 10 to 15 HFWs were trained, one passage 100 to 150 words long was read, and five multi-choice questions were answered.

The results obtained for word reading accuracy and transfer are in agreement with what was found by Martin-Chang et al. (2007), mentioned in the literature review. In this study, the authors compared the acquisition, transfer, and maintenance of words trained in isolation to those trained in passages. They found that words trained in isolation did not transfer to in-text reading whereas those trained in passages did transfer. Since experiment 1 involved training words in isolation alone, the results obtained by Martin-Chang et al. (2007) may explain why the training in the current study did not transfer to in-text reading.

Somewhat contrary to this, the study conducted by Tan and Nicholson (1997) found that training words in isolation versus in sentences lead to equal transfer to text reading competencies. The main distinction between the research procedures undertaken by Martin-

Chang et al. (2007) compared to Tan and Nicholson (1997) was that Tan and Nicholson (1997) had used a contextualizing phrase to accompany any unknown words. This was a two-word phrase such as *lemonade drink* with the word *lemonade* that accompanied each word regardless of whether it was presented during isolated word training, sentence word training, or the control condition. This was used to help the child understand the meaning of the word. This suggests that one possible modification that could be made to training words in isolation is a contextual phrase to make their meanings clear.

Substantial claims regarding the outcome of passage reading accuracy, speed, and comprehension cannot be made due to the increasing slopes observed regardless of the point at which the intervention phase had commenced. The increasing slopes observed here may be the result of a gradual increase in reading competency due to the exposure to new vocabulary and overall reading experience provided by this study. It may also reflect development due to usual classroom instruction.

Single word training did not seem to cause a large enough effect that could be observed above these existing slopes. Failure to observe effects above this trend suggests training did not have an impact on reading accuracy, speed, and comprehension. One reason for this may be the lack of single word reading transfer. Another reason is the possible time lapse between word training and passage reading. The group of words trained may have been stored in short term memory and these visual representations may have faded over time (Kornell, 2009). A post-reading measure of word reading accuracy would give an indication of whether memory for target words maintained during passage reading and may give further insight into the factors influencing passage reading accuracy, speed, and comprehension.

Lessons learned from experiment 1. A number of lessons learned from Experiment 1 were used to inform the procedure undertaken in Experiment 2. The comparison between the procedure used for Experiment 1 and Tan and Nicholson's (1997) study suggested that adding a contextual phrase may aid the skill transfer of single word training to reading the target words in the text. The failure to observe a distinct change in passage reading accuracy, speed, and comprehension after the intervention suggested a post-passage reading measure of word reading accuracy would be useful in ruling out the possibility of lost memory over time. Other lessons learned from Experiment 1 were that each session could be undertaken within the planned time frame of 20-30 minutes and that the materials used were suitable for the planned age group of participants.

Chapter 4 - Experiment 2

The main aim of experiment 2 was to answer the primary research questions. The primary research questions were answered through the measurement of the dependent variables: word reading accuracy, word reading transfer, passage reading comprehension, passage reading speed, and passage reading accuracy. Its secondary aim was to give insight into the social validity of single word training. Social validity was measured through a student feedback questionnaire.

Modifications were made to Experiment 2 as a result of lessons learned from Experiment 1 in the form of an additional contextual phrase and a measure of maintenance. A number of other modifications were made to increase the suitability of the intervention to the participants involved.

Method

The present study included five participants from ages ranging from 8 to 9 years at Reading recovery levels of 13 to 16. A multiple baseline experimental design was used to stagger the onset of the intervention for each participant.

The training procedure involved the training of up to 10-15 HFWs contained in Fry's (1997) *1000 Instant words* list and/or NZCER *Essential Word list* (Elley, Croft, & Cowie, 1977). Both lists were used in an effort to increase the chances of targeting novel words in the training. Each set of target words were aimed to be unfamiliar to the participant and were part of the passage used at each session. The dependent variables were the same as those used in Experiment 1.

Statistical analysis was an extra component undertaken for Experiment 2 which was made possible through the Wampold and Worsham (1986) method of randomisation (also mentioned by Arndorfer & Allen, 2001; Ferron & Sentovich, 2002; and Koehler & Levin, 1998). This method required each participant to be randomly allocated to a different intervention onset point. Each participant received three to seven intervention and baseline sessions based on their random allocation. Overall, there were a total of 10 sessions undertaken over a two to three-week period, each with duration of 20-30 minutes.

Participants. Five participants from the same school as the participant in Experiment 1 were recruited for Experiment 2. According to the school, participant 1 was 8 years and 3 months old female at a Reading Recovery level of 16, participant 2 was an 8 years and 6 months old male at level 16, participant 3 was a 9 years and 7 months old male at level 14, participant 4 was a 9 years and 5 months old male at level 14, and participant 5 was a 9 years (0 months) old male at a Reading Recovery level of 15. All of the participants had normal oral language for their age and were not part of the school's ESOL program. Reading level information for each child was provided by the school.

Dependent Measures. The dependent variables measured were the same as in Experiment 1.

Materials Used for Training Programme. The flashcards used in Experiment 2 were modified in three ways compared to the word cards used in Experiment 1. Firstly, instead of using HFWs from a generic HFW list provided by the participants' school used in Experiment 1, Experiment 2 used Fry's (1997) *1000 Instant words list* and *NZCER Essential Word list* (Elley et

al., 1977). These lists are described to include the most frequently encountered words from fiction and non-fiction library books from a wide range of different subject matters (Fry, 1997). Fry (1997) claimed to account for 65% of all words in any written material in the first 300 words of his *1000 Instant words list* HFW list. The least common words within these lists were prioritised to increase the chances of targeting unknown words. Two word lists (the *1000 Instant words list* and the *NZCER Essential Word list*) were used to increase the chances of including words that were in the passages as well as in a HFW lists.

Secondly, two to four word phrases were added to the back of each card to contextualise the target words. For instance, the HFW *own* from Fry's 1000 words list was modified to *owner* to fit the form of the word in the passage like in Experiment 1, but this time the phrase "*the dairy owner*" was printed at the back as seen in Figure 11. The target word in each phrase was underlined to draw attention to it. This modification was made to increase the chances of the trained words transferring to in-text reading (Martin-Chang et al., 2007; Tan and Nicolson, 1997). It was also made to make the meaning of the word clear and to incorporate different forms of learning to keep the intervention interesting (Kornell, 2009).

Figure 11. The front and back of a sample flashcard.

Thirdly, three extra passages from the *Ready to Read* series published by the Ministry of Education were used for Experiment 2, making a total of ten passages rather than seven used for

Experiment 1. Six texts of these were at the purple grade (aimed at the average reading level for children 7.0-7.5 years old) for *guided reading* and the remaining four were at the light blue (6.5-7.0-year level) grade for *guided reading* (Appendix B). Like in Experiment 1, the level appropriate for *guided reading* was used due to the extent of reading guidance provided by the researcher during the intervention.

Procedure. The procedure for Experiment 2 was also altered based on the lessons learned from Experiment 1. Firstly, as mentioned earlier, there was an accompanying two to four-word phrase that was read out loud to contextualise each word (e.g. *the dairy owner*). The participants were asked to only read the underlined target word while the researcher read the rest of the phrase. The contextual phrases were repeated at least three times intermittently throughout the training session while the single word training made up the majority of the drill training procedure.

Secondly, a measure of word reading maintenance was added to Experiment 2. The maintenance measure was aimed to give an indication of whether words trained in isolation in a short period maintained over time. It was added to Experiment 2 to test the possibility that the results may represent lost memory of target words over time due to the insufficiently short duration of each intervention session. For instance, the difference between word accuracy and transfer observed in Experiment 1 may be an attribution of the time lapse between the end of the single word training and when these measurements were taken (Kornell, 2009). Therefore, a measure of maintenance may increase the validity of claims made regarding the effect of the intervention.

Maintenance was obtained by re-presenting the participants with their past targets on flashcards. It was measured three weeks from the start of the study. Due to varying student absenteeism, the time at which the maintenance measure was taken ranged from 3 days to a week after each participant's last training session. It was calculated by dividing the number of target words correctly identified on flashcards by the total number of original target words trained. Each child was tested on a unique set of words since the target words were individualised to each child's ability to recognise them prior to the training. Maintenance measures were only taken for the intervention sessions.

Additionally, as the participants of Experiment 2 were more able readers, there were instances where the pre-training probes (step 1 below) led to only seven to ten incorrect words. This would have led to instances of less than 10 target words (step 2 below), which is contradictory to the research conducted by Kornell (2009) mentioned in the procedure for Experiment 1. Kornell (2009) emphasised the use of at least 10 target words to reduce the chances of the child using maladaptive memory strategies to learn the targets. Due to this, it was decided that a minimum of ten target words would be used regardless of if it included known words. If less than 10 words were incorrect, a random selection of known words would be included in the 10 target group.

The procedure undertaken for Experiment 2 is summarised below. The alterations to the procedure from that used for Experiment 1 are bolded.

1. Each child's prior knowledge of the HFWs was probed by asking them to read aloud approximately 20-25 words contained in the passage.

2. From these, a maximum of 10 incorrectly identified words were selected for training with flashcards.
    a. **If less than 10 words were incorrect, a random selection of known words were selected to make a total of 10 target words.**
3. These target words were re-presented to the child with a **short written example contextualing each word**. Incorrect responses were prompted.
4. Drill training was carried out (only for the intervention phase).
5. The child was asked to read a passage from *Ready to Read* materials.
6. Five to six multi-choice comprehension questions were asked.
7. Each child was asked to complete a feedback form after each lesson.
8. **A maintenance measure was taken 3 weeks from the start of the first baseline session.**

Inter-observer reliability. In order to obtain inter-observer reliability, one other person was asked to listen to the recordings of the passages read by the participants. Interrater reliability was measured by dividing the total number of agreements with the total number of disagreements and multiplying by 100. The total percentage of agreement across all sessions for participant 1 was 95%, for participants 2 was 93%, for participant 3 was 96%, for participant 4 was 99%, and for participant 5 was 97%.

Social Validity. A feedback questionnaire used in Experiment 1 was also used in Experiment 2.

Design. A within-subject multiple baseline design was used for Experiment 2 to investigate the influence the intervention had on measurements of word reading accuracy, word reading transfer, and passage reading accuracy, speed, and comprehension.

The Wampold-Worsham method of randomisation was used to improve the statistical validity of the multiple baseline design (Arndorfer & Allen, 2001; Ferron & Sentovich, 2002; Koehler & Levin, 1998; Wampold & Worsham, 1986). A number of publications aimed at reforming the multiple baseline design have suggested that the best way to satisfy the assumption of randomisation put forth by traditional statistics is through randomly allocating the onset of the intervention phase for each participant (Arndorfer & Allen, 2001; Ferron & Sentovich, 2002; Koehler & Levin, 1998; Wampold & Worsham, 1986). The Wampold-Worsham method of randomisation involved the randomised staggered onset of intervention sessions. It led to baseline and intervention phases of different lengths and the random allocation of participants into one of five possible starting points. The lattice below depicts the arrangement of intervention onsets used across the multiple baseline design in which *A* represented the baseline phase and *B* represented the intervention phase. All participants received the same story sequence where stories 1-10 each represented a particular story.

Table 2.

Randomly allocated intervention onsets for each participant

	Story 1	*Story 2*	*Story 3*	*Story 4*	*Story 5*	*Story 6*	*Story 7*	*Story 8*	*Story 9*	*Story 10*
Participant 1	A	A	A	B	B	B	B	B	B	B
Participant 2	A	A	A	A	B	B	B	B	B	B
Participant 3	A	A	A	A	A	B	B	B	B	B
Participant 4	A	A	A	A	A	A	B	B	B	B
Participant 5	A	A	A	A	A	A	A	B	B	B

Note. Participant 1 had her/his first intervention session at story 4, participant 2 had his/her first intervention session at story 5 and so on.

At least four participants were required for the Wampold-Worsham method to reach significance at the .05 level and it was also the minimal number recommended for any multiple baseline design (Ferron & Sentovich, 2002; Wampold & Worsham, 1986). For this reason, a minimum of five participants were requested from the school to allow for the possibility of one child dropping out of the study.

Each participant started the intervention at different points based on their predetermined random allocation depicted in Table 2. In accordance with the Wampold-Worsham (1986) method, randomisation without replacement was used to assign each participant as *participant 1*, *participant 2*, *participant 3*, *participant 4*, or *participant 5* in the above grid in Table 2. This means that the first participant had one in five chance of receiving their given allocation, the second participant had one in four chance of receiving their given allocation, the third participant had a one in three chance, and so on. Based on this, there were 120 ($5 \times 4 \times 3 \times 2 = 120$)

possible combinations and the likelihood of obtaining the current combination was 1/120. Table 11 in Appendix E depicts 10 out of these 120 possible combinations.

Data Analysis. The measures of interest; single word reading accuracy, word reading transfer, and passage reading accuracy, speed, and comprehension, were calculated using the same calculations outlined in Experiment 1. Also similar to Experiment 1, each measure of interest was analysed through visual analysis and Cohen's d effect sizes. An additional analysis carried out for Experiment 2 was statistical analysis.

Visual analysis. The visual analysis used for Experiment 2 had an extra component to that applied for Experiment 1. Multiple baseline designs allow *vertical analysis* across series due to the staggered allocation of intervention onsets. *Vertical analysis* is the process of examining the data across series (between participants) and comparing the absence or presence of an effect. For instance, participant 2's progress at story 5 (his first intervention session) can be compared participants 3-5's progress at story 5 (who were still in their baseline phase) and participant 1's progress at story 5 (who was in his second intervention session). If a change occurs in one series in which the intervention was applied but not the others, it increases the validity of inferences that suggest the intervention has had an effect (Swaminathan, Sugai, & Smolkowski, 2012).

Effect size analysis using Cohen's d. Cohen's d effect size values were computed the same way as in Experiment 1.

Statistical analysis. Assessment of statistical significance in Experiment 2 was made possible through the Wampold-Worsham method (1986) of randomisation, applied when designing the multiple baseline experiment. A randomisation procedure such as the Wampold-

Worsham method was chosen over widely applied parametric analysis because randomisation tests are unrestricted by assumptions of a normal distribution, homogeneity of variance, and series independence (Bulté & Onghena, 2008; Bulté & Onghena, 2009; Ferron & Sentovich, 2002). It was also chosen over non-parametric analysis due to its suitability to small sample within-subject data (Bulté & Onghena, 2008).

The statistical analysis involved the calculation of a test statistic for the obtained data. This was calculated by finding the difference between the means obtained in the baseline and intervention phases of the study. The sum of these values across all five participants made up the test statistic for each variable of interest (Ferron & Sentovich, 2002; Koehler & Levin, 1998; Wampold & Worsham, 1986). See Table 10 in Appendix E for individual test statistics and their sums for each variable of interest (word reading accuracy, word reading transfer, and passage reading accuracy, speed, and comprehension).

The formula used to calculate the test statistic for each variable is below. The one-tailed option (i.e. $\bar{x}_{phase\ b} - \bar{x}_{phase\ a}$), as opposed to two-tailed (i.e. $|\bar{x}_b - \bar{x}_a|$) was chosen to increase the power of the analysis (Bulté & Onghena, 2008):

$$\sum \bar{x}_b - \bar{x}_a$$

A distribution of test statistics for each of the 120 possible combinations was then determined through the use of the permn(5) command from the combinat package in R. R is a statistical language which uses a number of packages to make statistical analyses and graphical displays. Some of the multiple baseline graphs used for visual analysis were also constructed in

R using the graph(design= ”MBD”) command in the SCRT package (Bulté & Onghena, 2008). The test statistics for each possibility was further calculated and ordered to allow for the extrapolation of a distribution of test statistics using Excel. From this distribution, the p-value of the obtained data was compared to the chosen cut-off p-value of 0.05. Observations in the above or below 5% of all test statistics were therefore deemed as significantly different to what would have been obtained by chance. The results section may further clarify this procedure by referring to the outcomes of each stage of analysis depicted in Tables 12-16 and Figures 17-21 in Appendix E.

One of the limitations of randomisation tests is that they do not take into consideration individual trends. Individual differences in responsiveness to the intervention may alter the overall p-value obtained for each variable, concealing the fact that one participant may have responded well to the intervention while another did not. Due to this, visual analysis is still emphasised as a vital mode of analysis for the current multiple baseline design.

Results

Experiment 1 and 2 were synonymous in their purpose in answering the primary research questions. Statistical analysis through the Wampold-Worsham method of randomisation was a further analysis applied in Experiment 2 to answer the same questions.

Statistical Analysis. The results of the statistical analysis have been commented on in relation to the four main measures of interest: word reading accuracy, word reading transfer, and passage reading accuracy, speed, and comprehension. Table 3 shows the obtained test statistic and p-value for each of these variables. It shows that word reading accuracy and transfer both have large test statistics and the p-values obtained are significant at the 0.05 significance (p=.01 for word reading in isolation and p= .03 for word reading transfer to in-text reading as seen in Table 3). The p-values give an indication of whether the scores obtained are significantly different to what could have been achieved by chance. The small p-values obtained for word reading accuracy may be due to the fact that the majority of previously unknown target words were usually all trained to mastery during the training phase.

Tables 12 and13 and Figures 17 and 18 in Appendix E make explicit the process through which the p-values were computed. Table 11 in Appendix E gives an example of 10 out of the 120 possible combinations of intervention onsets that may have arisen by chance. From these, Tables 12 and 13 in Appendix E depict the test statistics for each possible 120 combinations. The top left is the obtained test statistic and the rest are what could have been achieved by chance if the combinations of onsets were different. Figures 17 and 18 in Appendix E give a visual depiction of where the obtained test statistic lies in relation to the whole distribution of 120

possible test statistics. The vertical line in each graph represents the obtained test statistic. Obtained test statistics on the far right or far left of the distribution indicate small p-values.

Table 3.

Overall test statistic and p-value for each measure of interest

Measure of interest	Test statistic	p-value
Passage reading accuracy	-0.02	0.54
Passage reading speed	-0.26	0.38
Passage reading comprehension	0.27	0.41
Word reading accuracy	4.04	**0.01**
Word reading transfer	1.17	**0.03**

Note. Table 10 in Appendix E gives a breakdown of how these test distributions were computed. Tables 11-16 and Figures 17-21 in Appendix E explain how the p-values were computed.

Passage reading accuracy, speed, and comprehension did not have significant p-values at the 0.05 level of significance. This suggests that the obtained test statistics were not significantly different to what would have been achieved by chance if any other random allocation had been chosen. Figures 18-21 in Appendix E further confirmed the p-value by graphically displaying the test statistics of all 120 random combinations (Tables 14-16 in Appendix E) relative to the obtained test statistics.

Visual Analysis of Trends and Cohen's d Effect Size Analysis. Where the statistical analysis above seeks to make general conclusions about the effect of the intervention across the whole sample, visual analysis of trends and Cohen's d effect size analysis will give an indication of individual responsiveness to the intervention. The graphs represented in this section are the same as those represented in Experiment 1, except they are represented as multiple baseline data. The broken line in these graphs represents the transition from baseline (phase A) to intervention (phase B) phases. Each subgraph represents the observations from each individual participant. The visual analysis of these graphs will follow the same guidelines for evaluating within-phase and between phase data outlined in Experiment 1 with an added component of *visual analysis*.

Word reading accuracy. The graphic display of word reading accuracy, represented as the hollow circle in Figure 12, shows a change in the *level* (defined earlier as the average of within phase data) for word reading accuracy scores in the intervention phase compared to the baseline phase. There also seems to be low variability within phases and low overlap between the scores obtained between phases, suggesting a training effect. This is confirmed in the large effect sizes based on Cohen's d classification observed for all five participants in Table 4.

Word reading transfer. Word reading transfer for participant 1, 2 and 4 seems to show similar trends, where there is a distinct increase in the *level* of scores obtained in the baseline compared to the intervention phase and these distributions do not overlap. The effect of the intervention is confirmed by large effect sizes of 2.0, 2.1, and 1.6 in Table 5. The reading transfer scores obtained by participants 3 and 5 also had moderate effect sizes of 0.4 and 0.5 but visual analysis of Figure 12 shows that both participants reached a higher score one session prior to the implementation of the intervention. This suggests ceiling effects and puts into question the

validity of claims regarding the effect of the intervention based on just effect size and statistical analysis outcomes.

High levels of word reading transfer obtained in the baseline phase may be attributed to practice effects as the baseline procedure involved one exposure to the target words in isolation and one exposure within a contextualising phrase to give the word meaning. The contextualising phrase alone may have also increased the chances of the word transferring to in-text reading as it made the meaning of the word clear. There is also the possibility that the target words within the passage could be guessed using contextual cues (Share, 2004), leading to higher transfer scores than word accuracy scores. These factors are considered in the discussion below as reasons for the observed ceiling effects.

Vertical analysis reveals some pattern of scores across series where the majority of participants experienced a decrease in word reading transfer at sessions 2 and 10 and an increase in sessions 3, 6, and 9.

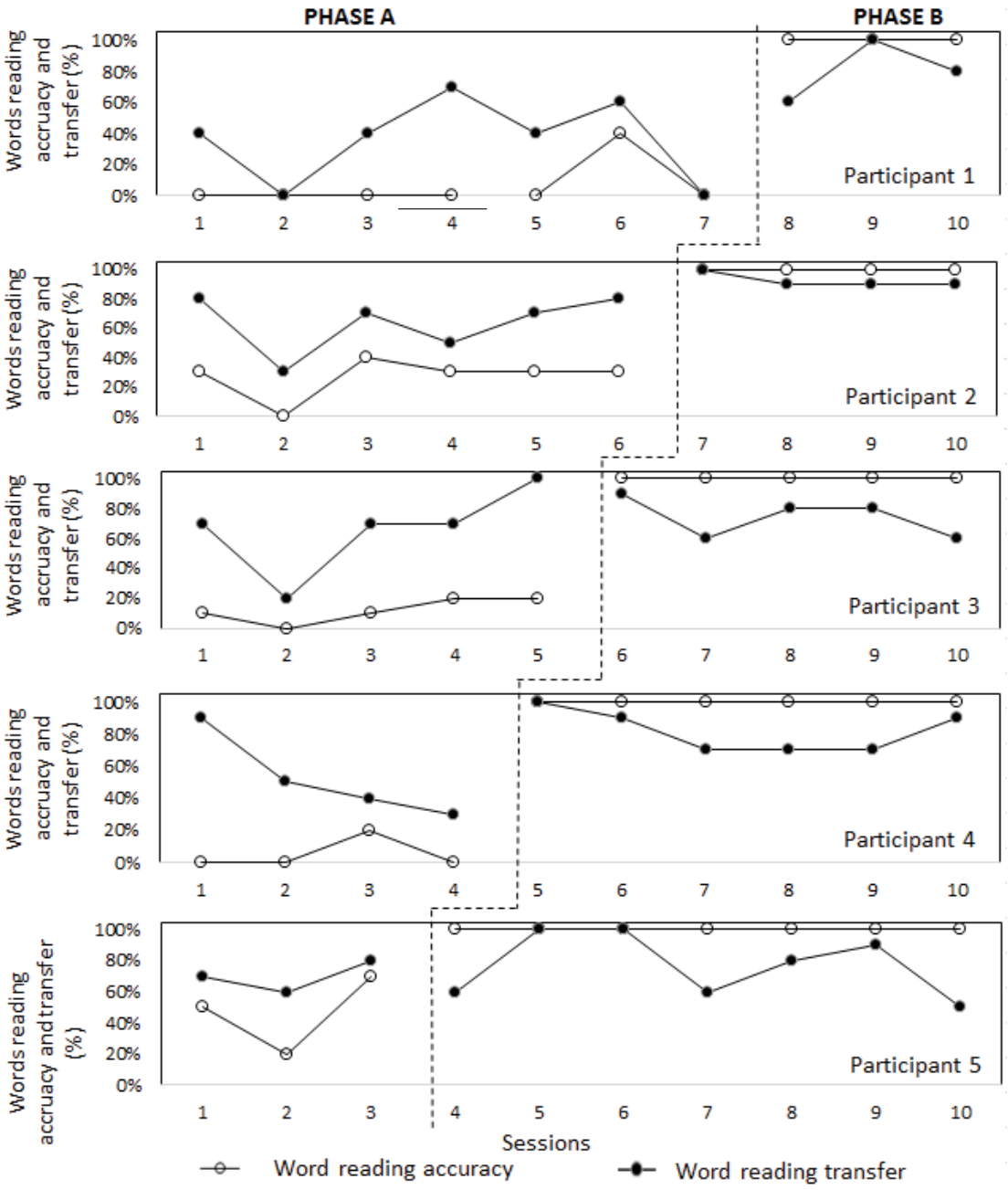


Figure 12. Word reading accuracy and transfer per session for each participant.

Table 4.

Cohen's d effect sizes for word reading accuracy.

	Baseline		Intervention		Effect size
	Mean	SD baseline	Mean	SD intervention	
Participant 1	5.7%	14.0%	100.0%	0.0%	9.5
Participant 2	26.7%	12.5%	100.0%	0.0%	3.0
Participant 3	12.7%	7.5%	100.0%	0.0%	17.8
Participant 4	5.0%	8.7%	100.0%	0.0%	14.9
Participant 5	46.7%	20.5%	100.0%	0.0%	3.5

Table 5.

Cohen's d effect sizes for word reading transfer.

	Baseline		Intervention		Effect size
	Mean	SD baseline	Mean	SD intervention	
Participant 1	35.7%	25.0%	80.0%	16.3%	2.1
Participant 2	63.3%	18.0%	92.5%	9.4%	2.0
Participant 3	66.6%	25.8%	74.0%	12.0%	0.4
Participant 4	52.5%	22.8%	81.7%	8.7%	1.6
Participant 5	70.0%	8.2%	77.1%	17.0%	0.5

Passage reading accuracy. Effect sizes for passage reading accuracy in Table 6 show negative values for participants 1, 2 and 5, a very small positive value for participant 3, and a large positive effect size for participant 4. The visual analysis of the graphical information in Figure 13 does not suggest any distinct differences in average word reading accuracy obtained in the baseline and intervention phases. Large variability within phases and overlap between phases is apparent.

Vertical analysis of trends shows a pattern across series of data. All participants experienced a steady increase in accuracy between sessions 1-3, a drastic decrease in session 4, and an increase in sessions 5-6. The majority of participants also consistently experienced another dip in performance at sessions 7 and 10. These trends were present regardless of the onset of the intervention and suggest the presence of a story effect.

Table 6.

Cohen's d effect sizes for passage reading accuracy.

	Baseline		Intervention		Effect size
	Mean	SD baseline	Mean	SD intervention	
Participant 1	75.6%	6.5%	74.7%	2.1%	-0.2
Participant 2	84.5%	6.3%	78.0%	2.2%	-1.3
Participant 3	74.0%	7.7%	68.0%	9.3%	-0.7
Participant 4	64.2%	10.0%	73.5%	8.5%	1.1
Participant 5	80.3%	6.8%	75.8%	11.1%	-0.4

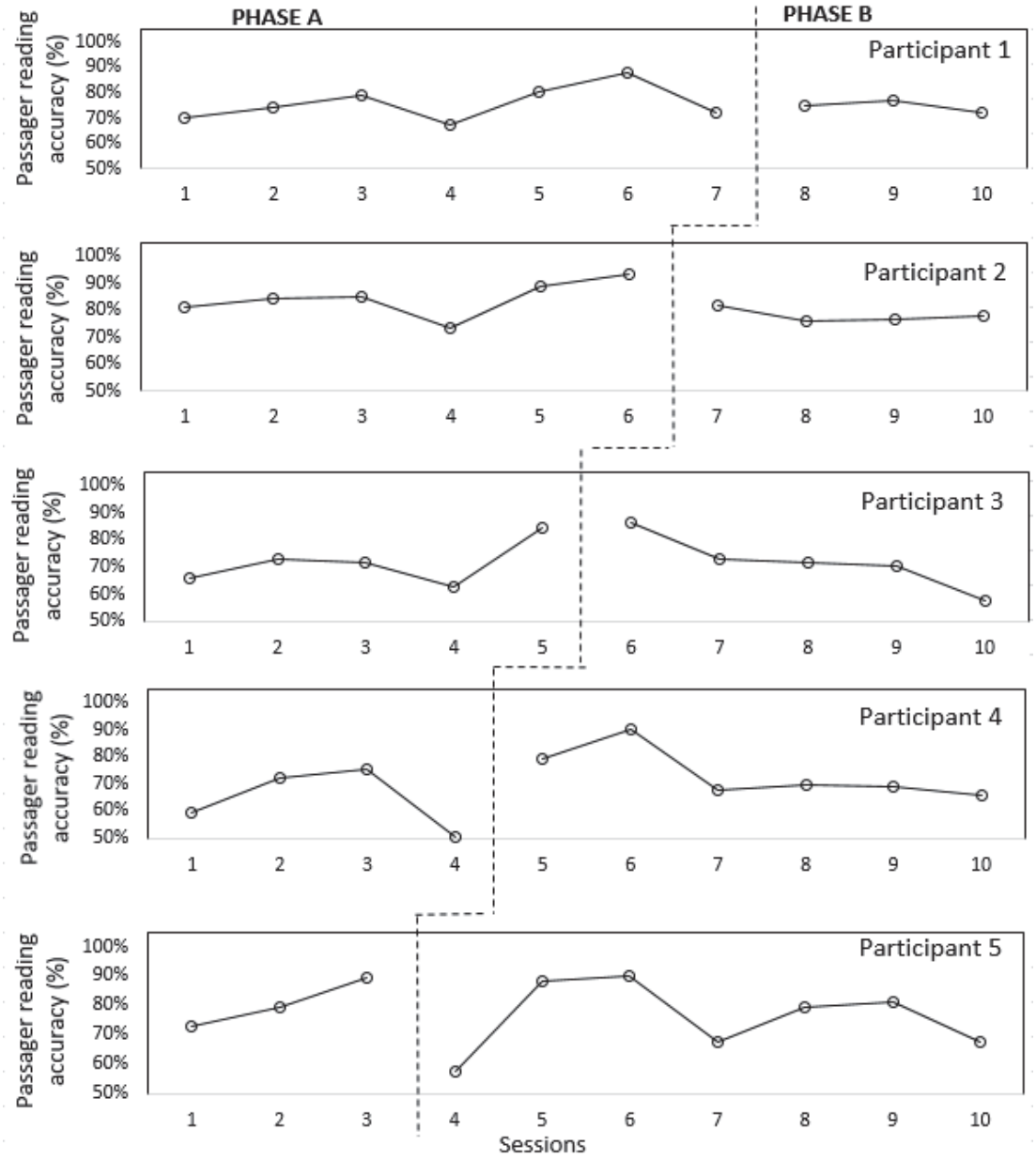


Figure 13. Passage reading accuracy per session for each participant.

Passage reading accuracy scores obtained here may be partly attributed to the percentage of word transfer achieved. Since the average percentage transfer obtained for Experiment 2 ranged from 83-92%, and 7-10 new words were trained, an average of 5.8 (83% of 7) to 9.2

(92% of 10) new words were transferred each session. Additionally, since the overall number of words in each passage was 100-150 words, 4% (5.8/150) to 9% (9.2/100) of words in each passage were target words that transferred. Therefore, theoretically, an increase of 4-9% in passage reading accuracy may be attributed to successful single word reading transfer. This idea is explored further in the discussion.

Passage reading speed. Effect sizes for passage reading speed in Table 7 shows that four out of five participants had negative large effect sizes, suggesting the average speed for baseline sessions was faster than the average speed in the intervention sessions for the majority of participants. However, as seen in Figure 14, there were two missing data points due to corrupted audio recording files which may have influenced the overall averages obtained. These were session 8 for participant 2 and session 6 for participant 5.

Table 7.

Cohen's d effect sizes for passage reading speed (words per second).

	Baseline		Intervention		Effect size
	Mean	SD baseline	Mean	SD intervention	
Participant 1	0.51	0.13	0.46	0.04	-0.51
Participant 2	0.59	0.10	0.46	0.06	-1.58
Participant 3	0.47	0.09	0.37	0.09	-1.04
Participant 4	0.40	0.08	0.43	0.11	0.07
Participant 5	0.48	0.05	0.39	0.06	-1.74

Visual analysis of Figure 14 shows large variability within phases and large overlap between phases. Vertical analysis of trends across series again shows a pattern across series of data. The majority of participants show a steady increase in words per second between sessions

1-3, a decrease in session 4, and high scores for sessions 5 and 6. They also experienced decreased performance in session 7 and 10. Again, these trends were present regardless of the onset of the intervention.

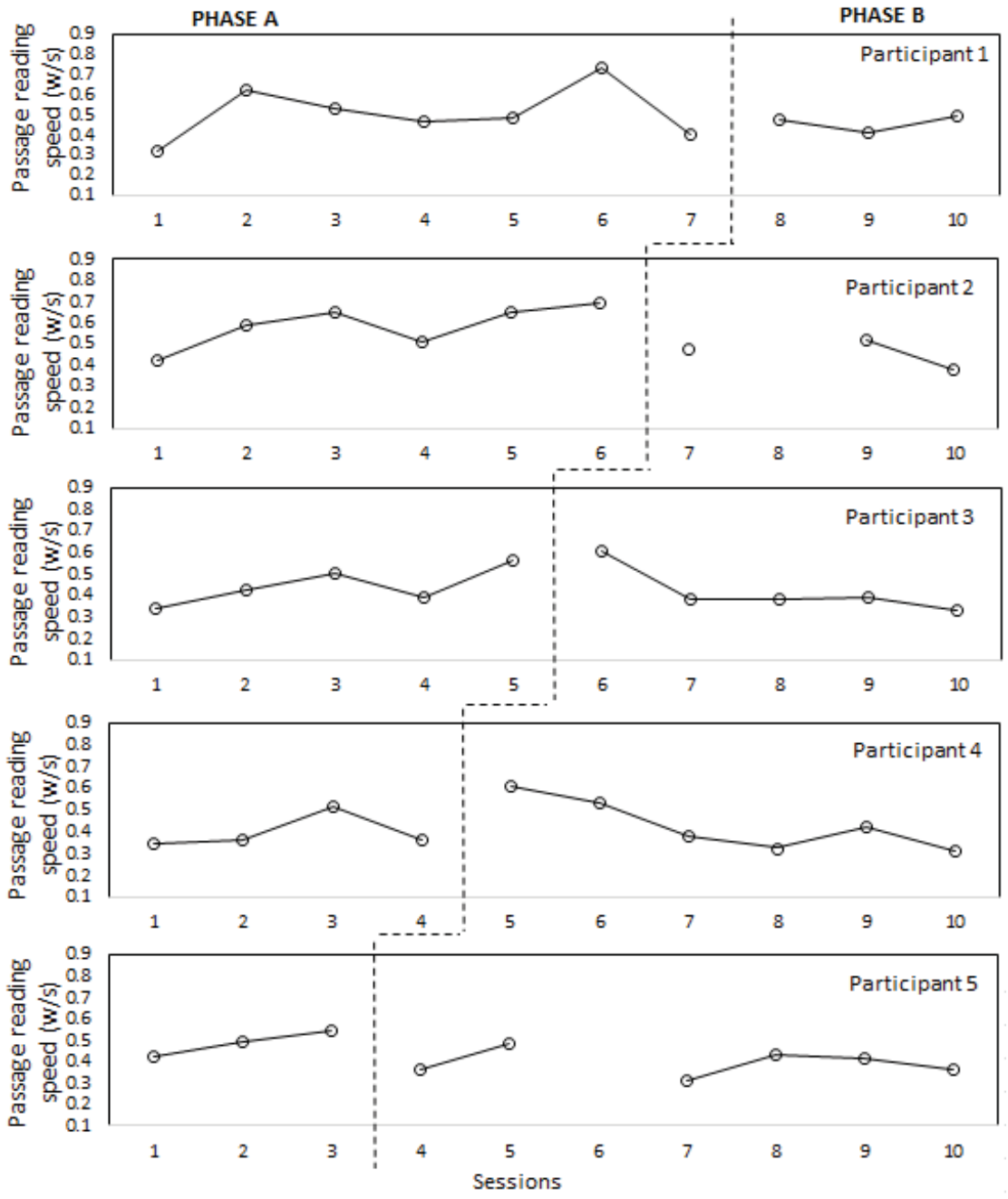


Figure 14. Passage reading speed per session for each participant.

Passage reading comprehension. Participant 4 and 5 had positive and moderate to large effect sizes for passage reading comprehension. The other three participants did not have significant effect sizes based on Cohen's d classification of significance (Bowman-Perrott et al., 2013)

Visual analysis of vertical trends in Figure 15 again seems to show large variability within phases and large overlap between phases, similar to the results observed for passage reading accuracy and speed. This makes it difficult to make claims regarding the effect of the intervention. It is also noteworthy that two out of the five participants obtained a score of 100% in their first baseline session and the other three obtained scores of 80% on their first baseline sessions, which suggests the presence of a ceiling effect.

Table 8.

Cohen's d effect sizes for passage reading comprehension.

	Baseline		Intervention		Effect size
	Mean	SD baseline	Mean	SD intervention	
Participant 1	71.4%	19.5%	73.3%	9.4%	0.14
Participant 2	86.7%	10.3%	85.0%	16.6%	-0.15
Participant 3	73.3%	30.3%	70.0%	29.4%	-0.11
Participant 4	70.0%	25.8%	83.3%	18.0%	0.65
Participant 5	80.0%	0.0%	85.7%	14.0%	0.61

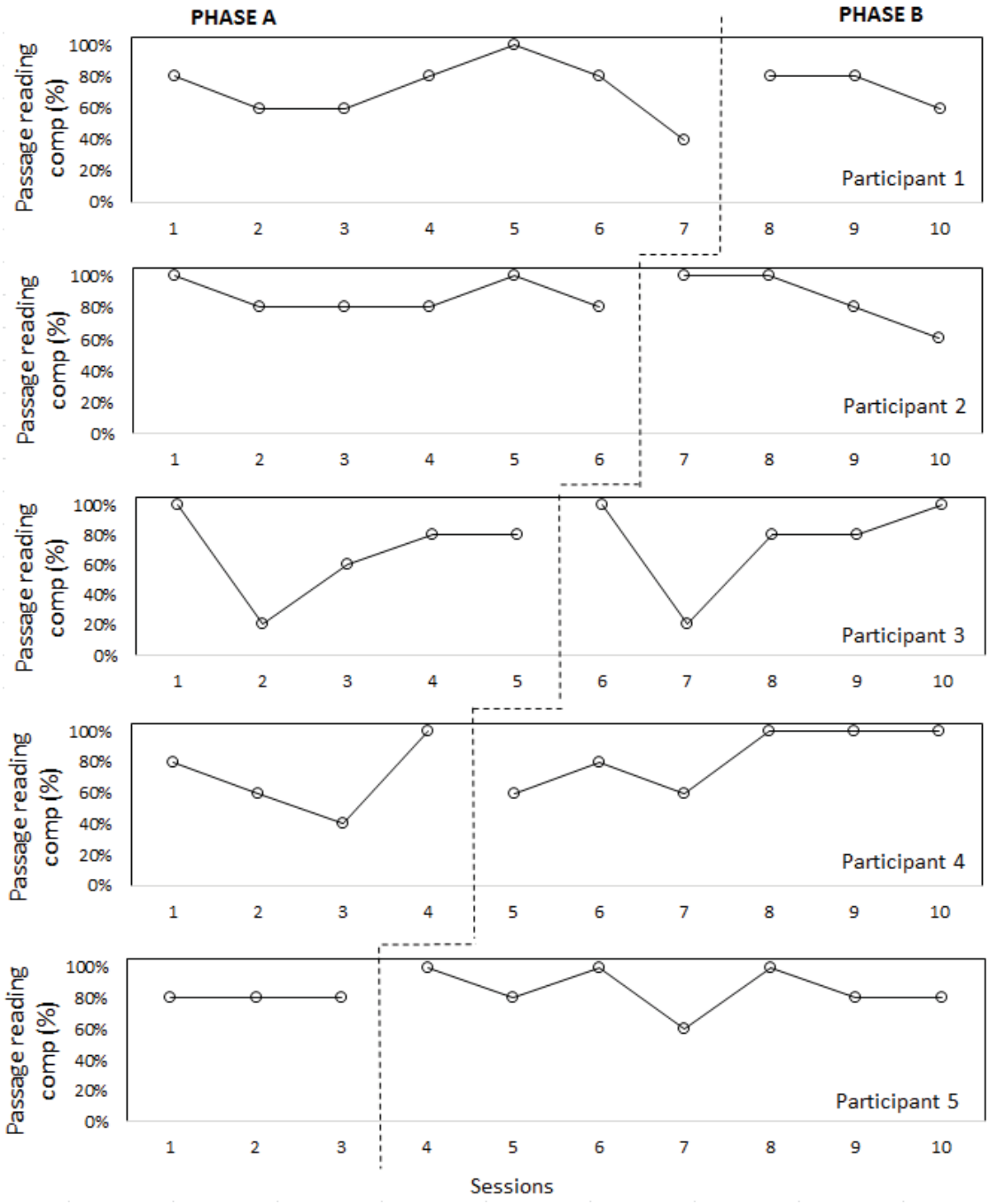


Figure 15. Passage reading comprehension per session for each participant.

Social validity. Figure 16 depicts the student feedback taken after each intervention sessions. Two out of five participants chose the happiest faces (face 1-2) across all questions after every session. Four out of five participants chose happy to moderate faces (face 1-3) for all questions across all sessions. However, participant 4 chose the second saddest face (face 4) for question 4 (i.e. *Would you like to come for another lesson*) on day 5. Participant 3 also chose the moderate face (faces 3) for question 4 on day 8. This suggests that the participants enjoyed the lesson, as indicated by their ratings on questions 1-3, but may have disliked an aspect of it, as indicated by their less positive responses on question 4. Similar to the participant in Experiment 1, the participants of this study were poor readers relative to their age group and some may have found it difficult to participate in the training and read the passages.

Figure 16. Student feedback questionnaire.

Maintenance of Single Word Reading. Measures of maintenance were taken for Experiment 2 to account for the possibility that the decreased memory of target words may have been the main reason for the lack of training transfer. Maintenance measures were taken three weeks from the first day of the experiment. Due to variable student attendance, the time at which the measure was taken for each participant varied between 3 days to a week after their last training session. Table 9 shows the maintenance of word reading accuracy obtained for each child at each intervention session. The mean maintenance for each participant ranged from 61-85%. The scores suggest that the majority of trained words had maintained for at least 3 days. This suggests the outcomes for passage reading accuracy, speed, and comprehension may not be entirely attributed to the lost retention of target words due to the time lapse between training and the passage reading measurement taken directly afterward.

Table 9.

Percentage maintenance obtained for participant 1-5 across each training session.

	Participant 1	Participant 2	Participant 3	Participant 4	Participant 5
Day 4					60%
Day 5				70%	40%
Day 6			80%	90%	50%
Day 7		90%	60%	90%	70%
Day 8	70%	80%	70%	90%	70%
Day 9	60%	50%	80%	70%	70%
Day 10	90%	100%	100%	100%	70%
Mean Maintenance	**73%**	**80%**	**78%**	**85%**	**61%**

Note. The spaces left blank were baseline sessions in which they did not receive training.

Summary

Experiment 2 was a multiple baseline design carried out to determine whether training HFWs would lead to significant improvement in word reading accuracy, word reading transfer and passage reading accuracy, speed, and comprehension. The secondary purpose of Experiment 2 was to gain an understanding of the social validity of single word training using flashcards. The procedure used in Experiment 2 was a modification of Experiment 1. The overall procedure involved ten sessions with durations of 20 to 30 minutes. In each session, 10 to 15 HFWs were trained with a small phrase contextualising each word, one passage approximately 100 to 150 words long was read, five multi-choice questions were answered, and five feedback questions were rated.

Research Question 1. The results showed that all five children made strong improvements in single word reading accuracy and were usually able to read the target words with 100% accuracy after training. Assessment of word reading transfer showed that 3 out of 5 children made average gains of 29-45% in word reading transfer in the intervention phase. These children were able to read 80-91% of trained words in the passage in the intervention phase. Ceiling effects were apparent for the remaining two children who seemed to have made only moderate gains of 7-9% on average between the baseline and intervention phase.

Research Question 2. The results showed that the word training had little impact on passage reading accuracy except for one of the children who made an average improvement of 9%. The remaining four participants did worse in the intervention phase compared to the

baseline phase on average. Visual analysis also did not show any distinct differences in the data obtained in the baseline and intervention phases.

Similar results were obtained for passage reading speed in that four of the five children decreased in reading speed in the intervention passages compared to the baseline passages. Only one child improved slightly in the intervention phase.

Results for passage comprehension also showed little improvement in the intervention phase. On average, there were small improvements for three of the children and slight declines for two of the children. Again, visual analysis of trends shows large variability within phases and overlap between phases, suggesting that any average improvement observed may be due to factors other than the intervention.

Social validity. The results showed that all of the children were positive about the lessons and usually wanted to come back for another lesson.

Maintenance. Single word reading maintenance measures showed that the participants maintained 61-85% of target words 3 days to a week after their last training session.

Discussion

The following discussion will be focused on the findings of Experiment 2. The primary aim of the experiment was to answer two research questions. First, will flashcard drill training enable low-achieving readers to improve their accuracy in identification of high-frequency words (HFWs) both in isolation and when reading those same words in the text? Second, will flashcard drill training lead to significant improvements in passage reading accuracy, fluency, and comprehension? The dependent variables associated with the first question were termed *word reading accuracy* and *word reading transfer*. The dependent variables associated with the second research question were termed *passage reading accuracy*, *passage reading speed,* and *passage reading comprehension*. A single case multiple baseline design was used to investigate the effectiveness of HFW training. Visual analysis, effect size analysis, and statistical analysis were utilised to explain the findings.

Statistical analysis revealed that both word reading accuracy and word reading transfer variables were significantly higher in the intervention phases compared to the baseline phase (at the .05 level of significance). The visual analysis of individual data confirmed the marked increase in word reading accuracy observed for all participants and the increase in word reading transfer observed for three out of five participants. Maintenance measures suggested that the majority of trained words maintained at least 3 days after training. The results obtained for passage reading accuracy, speed, and comprehension were not significantly different between the baseline and intervention phases based on both statistical analysis and visual analysis.

The interpretations of these results will be discussed in light of past research. The results of the social validity questionnaire will also be discussed. Lastly, the limitations of the current study together with recommendations for future research will be mentioned. The subsequent concluding chapter will summarise the overall contribution of Experiment 1 and 2 to the application of HFW training in the classroom and the theoretical understanding of word recognition training.

Research Question 1: Word reading accuracy and transfer. Single word HFW training significantly increased word reading accuracy for all participants. Evidence for this was provided through visual analysis together with Cohen's d effect size and statistical analysis. This suggests that it was possible for low-achieving readers to learn a set of 10-15 words through drill training within 15-20 minutes per session. Kornel's (2009) assertion that word lists made up of 10 or more words cannot be held in short term memory, suggests that the target words used in the current study were learned rather than arbitrarily held in short term memory. The maintenance measure confirms this by showing that the majority of words were maintained when presented again 3 days (from the last training session) to 2 weeks (from the first training session) after training.

Share (2004)'s self-teaching hypothesis argues that the acquisition of words while reading them in passages is the main route through which orthographic representations of novel words are formed. This is theorised to be because it helps to link unfamiliar words with their equivalents in oral language (Ehri, 2005; Share 2004). However, the current results for word reading accuracy suggest that word acquisition through single word drill training may be a legitimate alternative to acquiring new vocabulary through passage reading. This is important

because many poor readers find passage reading strenuous and boring, decreasing their motivation to engage in reading activities (Walczyk & Griffith-Ross, 2007) and further worsening their reading deficit due to avoidance (Nathan & Stanovich, 1991; Richards, 2000).

The second component of the first research question was whether HFW training of 10-15 target words would transfer to the recognition of those same words while reading a passage. Statistical analysis of the observations obtained for all five participants showed that the word reading transfer scores obtained overall were significantly higher in the intervention phase compared to the baseline phase. Visual analysis of individual data indicated that three out five participants had visually distinct non-overlapping sets of observations obtained in the baseline versus intervention phase whereas the other two had some overlap between phases. Overlap of observations between phases makes it difficult to claim that the intervention caused a significant change in word transfer for those participants (Kratochwill et al., 2012). This suggests that single word training clearly transferred to in-text word reading for at least three out of five participants whereas the other two may not have experienced significant transfer.

For the three participants who did experience distinct increases in transfer to in-text word reading, a major contributor to word reading transfer may have been the incorporation of the short contextualising phrase. The contextualising phrase may have facilitated the long-term memory of the target word as the number of cognitive connections made to a given reading stimulus is theorised to increase long-term recall (Cao et al., 2013). Skill transfer is another possible reason the phrase training may have led to increased word reading transfer (Fisher, Greer, Fuhrman, & Querim, 2015; Share, 2004). This is when the congruence between the training environment and the context in which the skill must be applied results in greater transfer

of new skills (Fisher et al., 2015). This may especially be the case since the participants were asked to read aloud the target word within the phrase while the researcher read the other words. This was repeated intermittently throughout training. For example, for the phrase *dairy owner*, the researcher would read the word *dairy* and the student would read the word *owner*. This way, the student may have learned how to identify the single target word amongst other words placed before and after it in a phrase. The repetition of this procedure intermittently during single word training may have further consolidated the association between the skills required when reading the word in isolation to reading it in a text. Similarly, task congruence is another possible reason the phrase training may have led to increased word reading transfer (Martin-Chang et al. 2007).

The reasons for the lack of observed word reading transfer for the remaining two participants (participant 3 and 5), despite also receiving the contextualising phrase, may be attributed to ceiling effects. Ceiling effects may be attributed when a change in the independent variable fails to effect change in a dependent variable which may have already reached high scores prior to the change in the independent variable (Ary, Jacobs, Sorensen, & Walker, 2013). The possibility that participant 3 and 5 may have experienced ceiling effects for word reading transfer is supported by the fact that both these participants started off with the highest average single word transfer percentages in the baseline phase compared to the baseline averages obtained by the other three participants.

Practice effects, the influence of the contextual phrase, and the use of contextual cues within the passage, may explain the ceiling effects observed as well as explain why transfer scores were consistently higher than word accuracy scores in the baseline phase. Firstly, practice effects may have been a factor because the procedure in the baseline phase involved single

exposure to each target word in isolation and then in a contextual phrase. Secondly, the contextual phrase and its facilitation of a deeper understanding of what the words mean and how they are used in a sentence may have also increased the chances of the target words being read in the passage. Lastly, research has found that children use the contextual cues within passages to identify unknown words, which may further explain why higher word transfer scores were observed compared to word reading accuracy (Share, 2004). This is especially the case for children with reading difficulty who may lack the ability to use other strategies to identify unknown words (Ehri, 2014). These three factors may explain the ceiling effects observed and also suggest that, in some cases, single exposure of the word in isolation together with the contextual phrase may have sufficed in eliciting in-text word reading transfer. This also suggests that participants 3 and 5, who experienced high scores within the baseline phase, may have been more receptive to these single exposures in the baseline phase compared to the other participants who may have required the drill training to in order achieve the same results.

Research Question 2: Passage reading accuracy, speed, and comprehension. The passage reading accuracy scores obtained in Experiment 2 may be partially attributed to the high percentages of single word transfer obtained. Given that training was found to allow the majority of participants to identify the target words in the text, a proportional increase in overall passage reading accuracy may be attributed to this effect alone. Contrary to the expectations that passage reading accuracy would increase at least partially from successful single word reading transfer alone, the results of the current study showed that all but one participant obtained an average decrease in passage reading accuracy. Similar to these findings, Tan and Nicholson (1997) also trained approximately 8% of words in their passages and found no change in passage reading

accuracy or speed. The authors attributed this to the small ratio of trained words to the number of words in the rest of the passage. The same attributions could be made for the results obtained from the current study.

The results of passage reading accuracy, speed, and comprehension showed no statistical difference between the baseline and intervention phases. Visual representations of these variables may reveal individual responsiveness to the intervention and it too suggests the intervention did not result in changes to passage reading accuracy, speed, or comprehension. One reason for this may have been the excessive difficulty of the passages chosen. The Reading Recovery levels of the participants were between levels 14 and 16 whereas the *Ready to Read* passages used were at the light blue and purple levels, aimed at Reading Recovery levels 17-20 (see Appendix B). These were chosen to allow room for improvement after the intervention. However, due to this, the participants may have found the reading material too difficult, inhibiting the effectiveness of the intervention. The use of difficult reading material may have decreased the transfer of training through several hypothesised mechanisms. Difficult reading material may have distracted attention away from the target words in the passage (Copeland & Gedeon, 2014), increased the overall cognitive load required to read (Ehri, 2014; Tønnessen & Uppstad, 2015), increased student anxiety and decreased performance (Eysenck, Derakshan, Santos, & Calvo, 2007), and/or decreased the student's ability to use contextual cues to identify unknown words (Ehri, 2014; Share, 2004).

The results may give an indication of whether the reading material was in fact too difficult for the participants. The obtained average passage reading accuracy for all five participants in the baseline phase ranged 64-84% across both phases. According to Clay (2001)

and Hudson et al. (2005), texts is usually considered as able to be read independently if they are read with accuracy percentages of 95% or more. Rates of 90-94% accuracy are appropriate for guided reading whereas accuracy rates 89% and lower are considered too difficult (Clay, 2001). This suggests that the reading levels may have been too difficult for the participants of the study.

Low reading ability may not have only influenced passage reading capabilities but may also reflect inherent difficulty in word learning. Apel (2009), Ehri (2014), and Wolf and Bowers (1999) suggest that children with lower reading ability employ cognitive learning strategies that lead to surface learning of words, associated with non-detailed orthographic representations of words in their mental lexicon. On the other hand, competent readers are hypothesised to form more detailed orthographic representations in memory (Apel, 2009; Ehri, 2014; Ehri and Wilce, 1985). Ehri (2014) attributes such differences in orthographic representations to the use of different memory mnemonics to identify words. Poor readers may use the arbitrary shape of the word to store it in memory whereas a good reader with well-developed phonological awareness may be able to use the letters in the words to identify and retrieve the orthographic representation of it from memory. Since the target words used in the current research were orthographically dissimilar to each other (see a sample list in Chapter 3), single word reading accuracy after training may have been equal for all the participants regardless of reading ability. However, when exposed to the same words in the passage, containing other words with similar orthographies, the accuracy with which the same words were identified may have dropped due to the use of surface mnemonics to learn them during training.

Despite the participants showing low accuracy rates, the average speed obtained by each participant seems to be normal for their reading levels. The average speed of passage reading for

all participants ranged 0.37-0.59 words per second. According to the oral reading norms collated by Hasbrouck and Tindal (2006) of children from grades 1-8 in the US, these reading rates are normal for low achieving 7-8 year old readers. Since the participants of the current research were reading at a 6-7 year old level, these results suggest the students of the current study were reading the presented passages at the normal rate for their reading level. Therefore, although passage reading accuracy indicates that the text may have been too difficult, the speeds obtained may reflect normal reading rate. The mismatch between accuracy and speed may be common as indicated by the longitudinal study conducted by Landerl and Wimmer (2008) over eight years which found that many of the 115 children studied had relatively stable reading speed compared to reading accuracy as they progressed in school.

Individual differences between participants may give an indication of why the majority of participants did not experience increases in passage reading competencies after training. For the majority of participants, average passage reading speed and accuracy decreased after the intervention despite high word transfer percentages. However, participant 4 was the only participant who experienced an increase in passage accuracy and speed (Table 6 and 7) while also having a similarly large transfer percentage to the other four participants (Table 5). The results show that participant 4 also had the largest average change in word reading accuracy (Table 4) between post and pre-training, suggesting that he acquired the most words through training, which then transferred to in-text reading. This suggests that there may be a tipping point at which single words training, given transfer, had an impact on passage reading accuracy and speed large enough to be observed.

The difference between the comprehension scores obtained in the baseline compared to the intervention phase was also not statistically significant. This may be contrasted with Tan and Nicholson's (1997) study that found a significant increase in comprehension scores after single word training. The methodology used in the current study may have influenced the results due to the use of HFW, rather than content words. The target words in Tan and Nicholson (1997) were content words such as *lemonade* and *raspberry* which may have added more to the comprehension of the text relative to HFWs such as *our, around,* and *been* used in this current research. Although the HFW lists used in the current study included common nouns such as *mother* and *summer,* any given target list only had up to four nouns. This suggests that the target lists used in the current study may not have added as much to the understanding of the text as the list of content words used by Tan and Nicholson (1997). Additionally, the small group of nouns used in the current study were still those most commonly encountered in text, in contrast to content words which may have been rarer and therefore may have stood out more in the text. This may explain why past research using content words has found an increase in reading comprehension whereas the current research using HFWs did not.

Social Validity. The outcomes of the student feedback questionnaire suggested the intervention was enjoyed by the majority of participants. This suggests that single word training is a socially valid form of reading intervention. Past research has also found that students tended to report drill training as a positive experience (Nist & Joseph, 2008). This may be especially useful for students with reading difficulty who may find it strenuous to build sight word vocabulary through reading experience alone (Nathan & Stanovich, 1991; Richards, 2000).

Maintenance. Where classroom training of HFWs is usually gradual and regular, increasing the likelihood of retention, the current research trained a novel set of words each session, decreasing the likelihood of retention. The current measurement of training maintenance was aimed to account for the possible loss of word retention as a contingency in the research method. Without it, the possibility that the time lapse between training and measurements of reading ability may have been the overriding explanation for the failure to induce reading accuracy, speed, and comprehension. The results indicated that the participants maintained the majority of target words assessed at least 3 days after the last training session. This suggests that factors other than training retention may explain the failure to see a change in passage reading competencies after training.

Limitations. Several limitations of the current study are apparent. Some of these have become apparent through the results and some reflect inherent assumptions in the measurements made. These limitations led to the recommendations for future research that may seek to investigate the same variables and/or seek to use the same measurement materials**.**

Story effect and baseline stability. The staggered multiple baseline design employed for Experiment 2 allows vertical analysis across series of data sets, making it possible to detect possible patterns in the data present outside of experimental manipulation (Kratochwill et al., 2012). A noticeable pattern in passage reading accuracy and speed is apparent in the scores achieved by all five participants, regardless of the onset of the intervention. All the participants experienced an increase in accuracy and speed between sessions 1-3, decreases in session 4, increases at sessions 5-6, and another decrease in performances at sessions 7 and 10. This

suggests that there may be a story effect in the data where the sequence of stories may have influenced performance enough to conceal any other effects present.

Variability caused by the story effect inhibited the obtainment of a stable baseline. It is preferred in single case research to obtain a stable baseline prior to the implementation of an intervention phase (Kratochwill et al., 2012). This is because variability in the baseline phase, especially in the same expected direction of the intervention, would lead to inconclusive inferences about the effect of the intervention (Kratochwill et al., 2012). The baseline phase could not be delayed in Experiment 2 due to the predetermined allocation of intervention onsets in compliance with the Wampold-Worsham method of randomisation.

The data analysis method used may have caused the passage reading variables (speed, accuracy, and comprehension) to be more vulnerable to variability caused by the story effect than the word reading variables (word reading accuracy and transfer). Word reading accuracy and transfer may have been a more reliable measure of training effects because the trained words were also the same words they read in the passage. On the other hand, the effect of training on passage reading accuracy, speed, and comprehension was determined by a comparison between reading performance on passages presented in the baseline versus intervention phases. Therefore, the passage reading variables may have been more influenced by this story effect because it involved the comparison between passages rather than a comparison between the chosen 10-15 target words.

Ceiling effect. A ceiling effect is when the majority of scores at baseline are at or near the maximum and changing the independent variable does not seem to influence change in the

dependent variable/s (Ary et al., 2013). It can occur as part of a treatment effect and/or as part of a measurement effect during data gathering (Ary et al., 2013). The most apparent ceiling effect obtained in Experiment 2 was observed for word reading transfer and comprehension. In the word reading transfer scores obtained for Experiment 2, three out of the five participants had at least one instance in which they scored 80-100% in the baseline phase. Four out of five participants in Experiment 2 also scored 100% at least once in the baseline phase for comprehension. This suggests that participants were already scoring high in the baseline phases, not leaving much room for improvement after the intervention. However, the number of instances in which the ceiling was reached for word reading transfer was higher in the intervention phase relative to the baseline phase. The same was not apparent in the comprehension scores observed. This suggests that, although a ceiling effect seems to be present in the results for both transfer and comprehension, the effects of the intervention for word transfer may still be observed through the number of times the participant reached the ceiling.

Suitability of measures. The suitability of strategies used to measure the variables of interest may have caused instability in the results and the ceiling effects mentioned earlier. The measurement of reading comprehension through multiple choice options was the only assessment employed in the current study that was different to that used by the majority of past research (Martin-Chang et al., 2005; Tan & Nicholson, 1997). Past research has used open-ended questions which may be a valid assessment of higher-order understanding (Palmer & Devitt, 2007) but may pose problems for objective assessment of responses, as mentioned in the limitations of Tan and Nicholson's (1997) research.

Multiple choice questions were used in the current study to reduce the stress of assessment that may be faced by students with weak language ability or self-perception of ability when they are asked to orally express their ideas (Brookhart, 2015). Research conducted by Segool et al. (2013) compared assessment anxiety experienced by 355 primary school children when they were assessed in their classrooms versus when they were tested outside their classrooms as part of state implemented testing. They found that the students reported significantly higher cognitive and physiological symptoms of assessment anxiety when they were tested outside their classrooms. Given that the current intervention was carried out in a room outside the classroom with participants with a history of poor reading performance, the decision was made to use multiple choice questions as the least anxiety provoking form of reading comprehension assessment. Multiple choice questions were also chosen to allow a clear marking schedule that could reduce the chances of assessment error and be easily replicated by other implementers (Tan & Nicholson, 1997).

Suitability of materials used. A major assumption made in the current research was the reading levels of the participants nominated for the study. The validity of teacher testing of reading levels has been questioned by the number of studies (Compton, Fuchs, Fuchs, & Bryant, 2006; Limbos & Geva, 2001; Jacob, 2005). Some attribute this to the inadequate oral testing of early reading ability (Compton et al., 2006; Fuchs et al., 2001; Limbos & Geva, 2001) and others attribute this to high teacher accountability for student progress (Jacob, 2005). Research also questions the validity of teacher and school nominations for interventions (Meisinger, Bradley, Schwanenflugel, Kuhn, & Morris, 2009; Siegle, Moore, Mann, & Wilson, 2010). This may have contributed to the low passage reading accuracy observed as the passages chosen for the current

study were chosen to be a maximum of two colour grades higher than the reported ability of the students nominated. Incorrect teacher grading may have exaggerated the difference between the reading level required to read the text and student reading abilities.

Future recommendations. The obtainment of a stable baseline is outlined as important for the validity of results gained through single case design (Kratochwill et al., 2012). Future research that aims to use a multiple baseline design using a method of randomisation such as the Wampold-Worsham method may benefit from delaying the intervention onsets for all participants until they all experience a stable baseline. This way, a stable baseline will be achieved without compromising the random allocation of intervention onsets. Possible increases in the time and resources needed to carry out the experiment may need to be accounted for if this strategy is used.

A disadvantage of using a single case design over an experimental design using a control group is the inability to control for variability caused by the different stories used. Such variability may be decreased by using a reading resource which includes a smaller range of reading levels. Future research single case research using the *Ready to Read* series may benefit from using books from only one level. This might limit the number of resources available and require the incorporation of other resources such as the PM storybook collection published by Nelson Education which is also graded to align with reading levels.

Assessment of reading ability prior to the commencement of the study may also allow a more appropriate pairing of reading resources with each child's ability. This is useful due to the

limitations of depending on teacher nominations (Meisinger et al., 2009) and teacher assessment (Limbos & Geva, 2001) to determine the child's ability.

Future research aimed to reduce student anxiety during the assessment of reading comprehension in a way that is objective and easily replicated may choose to use a large number of multiple choice questions. Where open-ended questions are believed to result in the better assessment of higher order understanding than multiple choice questions, Palmer and Devitt (2007) point out that it is the number of multiple choice questions that influence whether they assess more than just surface learning. In their study, Palmer and Devitt (2007) assessed the extent at which multiple choice questions versus essay questions measured the subject knowledge of pupils in a university undergraduate class. They found that the use of a large variety of multiple choice questions led to the assessment of deeper content understanding as well as surface recall of information whereas essay writing alone did not always assess deeper understanding. Therefore, in order to utilise the benefits of multiple choice comprehension questions such as less student anxiety (Brookhart, 2015) and more assessment impartiality, future research may choose to use a larger number of multiple choice questions in order to achieve the same results as could be achieved with open-ended questions.

Chapter 5 - Conclusion

New Zealand has one of the largest gaps in the reading attainment of good and poor readers amongst all OECD countries (Ogle et al., 2003). Tunmer et al. (2004) attribute this to large differences in literacy skills at school entrance that become more distal as children progress through school. It is apparent that the ability of the education system to allow these children to catch up to their peers is paramount for equal education opportunities for children from different economic and cultural backgrounds (Tunmer et al., 2004). The current research contributes to this by investigating one common difficulty faced by poor readers which is reading fluency (Nathan & Stanovich, 1991; Richards, 2000).

The main theoretical construct associated with fluency is automatisation, which represents the ability to identify words rapidly without conscious effort (Ehri, 2014). A number of rapid word identification strategies have been researched and applied to increase fluency but have not trained HFWs (Martin-Chang et al., 2007; Tan & Nicholson, 1997) and those that have trained HFWs have not measured direct transfer of trained words to in-text reading (McArthur et al. 2013; Watts & Gardner, 2013). This is a significant gap in the past research as HFW training is already commonly applied in the first years of school. The current research was carried out in light of the current need for research evidence related to the use of HFW training as a commonly utilised strategy to increase reading vocabulary and fluency.

Specifically, the aim of the study was to investigate the effect of HFW training on single word reading accuracy, its transfer to in-text word reading, and whether this would impact passage reading accuracy, speed, and comprehension. Two experiments were carried out in order

to allow the first experiment to inform changes to the word training strategy used in the second experiment. The results of the first research question for Experiment 2 showed significant increases in word reading accuracy and transfer to reading the same words in the assessment passages after the implementation of drill training. Conversely, the results of the second research question showed no change in passage reading accuracy, speed, or comprehension after single word training. Additional findings of the current research were that the words trained largely maintained over time and that the participants enjoyed the training procedure.

An important contribution of the current study is the use of the Wampold-Worsham method of randomisation. The application of this method during the design of Experiment 2 allowed the calculation of a significance value which served to compliment and quantify the results observed through visual analysis (Bowman-Perrott et al., 2013). Koehler and Levin (1998) have pointed out that the application of randomisation methods may be the cornerstone of statistical credibility for research with small sample sizes. However, despite its merits, it is rare to see single case research methodology utilise randomisation methods. The current research may, therefore, serve to inform future research that may aim to use a randomisation procedure within a multiple baseline research design. Efforts were made to make the randomisation method and analysis as explicit as possible in recognition of this.

The main contribution of the current study is its applicability to classroom contexts. The results of the present study suggest that HFW training is likely to result in novel word acquisition and the more accurate identification of trained words in the text, especially if the word training is complimented with contextualising phrases. This is important since HFW training is already

common in New Zealand schools (Ministry of Education, 2009). The single case design used in the current research also makes it easily replicated in the classroom (Kratochwill et al., 2012).

References

Ainsworth, M. K., Evmenova, A. S., Behrmann, M., & Jerome, M. (2016). Teaching phonics to groups of middle school students with autism, intellectual disabilities, and complex communication needs. *Research in Developmental Disabilities*, *56*, 165-176.

Apel, K. (2009). The acquisition of mental orthographic representations for reading and spelling development. *Communication Disorders Quarterly*, *31*(1), 42-52.

Apel, K., Thomas-Taite, S., Wilson-Fowler, E.B., Brimo, D. (2012). Acquisition of initial mental graphemic representations by children at risk for literacy development. *Applied psycholinguistics*, *33*(2), 365–391.

Archer, N., & Bryant, P. (2001). Investigating the role of context in learning to read: A direct test of Goodman's model. *British Journal of Psychology, 92*(4), 579-591.

Arndorfer, R. E., & Allen, K. D. (2001). Extending the efficacy of a thermal biofeedback treatment package to the management of tension-type headaches in children. *Headache: The Journal of Head and Face Pain*, *41*(2), 183-192.

Ary, D., Jacobs, L. C., Sorensen, C. K., & Walker, D. (2013). *Introduction to research in education*. Belmont, CA: Cengage Learning.

Binamé, F., & Poncelet, M. (2016). The development of the abilities to acquire novel detailed orthographic representations and maintain them in long-term memory. *Journal of Experimental Child Psychology*, *143*, 14-33.

Bosse, M. L., Chaves, N., Largy, P., & Valdois, S. (2015). Orthographic learning during reading: The role of whole-word visual processing. *Journal of Research in Reading*, *38*(2), 141-158.

Bowman-Perrott, L., Davis, H., Vannest, K., Williams, L., Greenwood, C., & Parker, R. (2013). Academic benefits of peer tutoring: A meta-analytic review of single-case research. *School Psychology Review*, *42*(1), 39-55.

Brookhart, S. M. (2015). Making the most of multiple choice. *Educational Leadership, 73*(1), 36-39.

Bulté, I., & Onghena, P. (2008). An R package for single-case randomization tests. *Behavior Research Methods*, *40*(2), 467-478.

Bulté, I., & Onghena, P. (2009). Randomization tests for multiple-baseline designs: An extension of the SCRT-R package. *Behavior Research Methods*, *41*(2), 477-485.

Cao, F., Rickles, B., Vu, M., Zhu, Z., Chan, D. H. L., Harris, L. N., ... & Perfetti, C. A. (2013). Early stage visual-orthographic processes predict long-term retention of word form and meaning: A visual encoding training study. *Journal of Neurolinguistics*, *26*(4), 440-461.

Chamberlain, M. (2013). *PIRLS 2010/2011 in New Zealand: An overview of findings from the third cycle of the progress in international reading literacy study (PIRLS).* Wellington, New Zealand: Ministry of Education.

Clay, M. M. (2001). *Running records*. Portsmouth, NH: Heinemann.

Clemens, N. H., Shapiro, E. S., Wu, J.-Y., Taylor, A. B., & Caskie, G. L. (2014). Monitoring early first-grade reading progress: A comparison of two measures. *Journal of Learning Disabilities, 47*(3), 254-270.

Compton, D. L., Fuchs, D., Fuchs, L. S., & Bryant, J. D. (2006). Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of decision rules and procedures. *Journal of Educational Psychology*, *98*(2), 394-409.

Copeland, L., & Gedeon, T. (2014). What are you reading most: Attention in elearning. *Procedia Computer Science*, *39*, 67-74.

Dennis, M. S., Sorrells, A. M., & Falcomata, T. S. (2016). Effects of two interventions on solving basic fact problems by second graders with mathematics learning disabilities. *Learning Disability Quarterly*, *39*(2), 95-112.

Duke, N. K., & Pearson, P. D. (2008). Effective practices for developing reading comprehension. *The Journal of Education*, *189*(1), 107-122.

Ehri, L. C. (1987). Learning to read and spell words. *Journal of Reading Behavior*, *19*(1), 5-31.

Ehri, L. C. (1992). Reconceptualising the development of sight word reading and its relationship to recoding. In Gough, P. B., Ehri, L. C. Teiman, R. (1992). *Reading Acquisition* (pp. 107-144). Hillsdale, NJ: Lawrence Erlbaum.

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, *9*(2), 167-188.

Ehri, L. C. (2014). Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning. *Scientific Studies of Reading*, *18*(1), 5-21.

Ehri, L. C., & Saltmarsh, J. (1995). Beginning readers outperform older disabled readers in learning to read words by sight. *Reading and Writing*, *7*(3), 295-326.

Ehri, L. C., & Wilce, L. S. (1985). Movement into reading: Is the first stage of printed word learning visual or phonetic? *Reading research Quarterly*, *20*(2), 163-179.

Elley, W.B., Croft, C, & Cowie, C. (1977). *A New Zealand basic word list. Revision of the Dolch basic sight vocabulary list.* Wellington, New Zealand: New Zealand Council for Educational Research.

Ellis, N. C., Natsume, M., Stavropoulou, K., Hoxhallari, L., Van Daal, V. H., P., Polyzoe, N., Tsipa., M., Petalas, M. (2004). The effects of orthographic depth on learning to read alphabetic, syllabic, and logographic scripts. *Reading Research Quarterly*, *39*(4), 438-468.

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, *7*(2), 336-353.

Ferron, J., & Sentovich, C. (2002). Statistical power of randomization tests used with multiple-baseline designs. *The Journal of Experimental Education*, *70*(2), 165-178.

Fisher, W. W., Greer, B. D., Fuhrman, A. M., & Querim, A. C. (2015). Using multiple schedules during functional communication training to promote rapid transfer of treatment effects. *Journal of Applied Behavior Analysis*, *48*(4), 713-733.

Franklin, R. D, Allison, D.B., & Gorman, B. S. (2014). *Design and analysis of single-case research.* New York, NY: Psychology Press.

Fry, E. (1980). The new instant word list. *The Reading Teacher*, *34*(3), 284-289.

Fry, E. (1997). *1000 instant words*. Lincolnwood, IL: Contemporary Books.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, *5*(3), 239-256.

Gast, D. L., & Ledford, J. R. (2014). *Single case research methodology: Applications in special education and behavioral sciences*. New York, NY: Routledge.

Griffin, C., & Joseph, L. M. (2015). Supplemental flashcard drill methods for efficiently helping at-risk kindergartners make letter-sound correspondences: Does presentation arrangement of words matter?. *Reading Psychology*, *36*(5), 421-444.

Guttentag, R. E., & Haith, M. M. (1978). Automatic processing as a function of age and reading ability. *Child Development*, *49*(3) 707-716.

Hasbrouck, J., & Tindal, G. A. (2006). Oral reading fluency norms: A valuable assessment tool for reading teachers. *The Reading Teacher*, *59*(7), 636-644.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional children*, *71*(2), 165-179.

Horner, R. H., Swaminathan, H., Sugai, G., & Smolkowski, K. (2012). Considerations for the systematic analysis and use of single-case research. *Education and Treatment of Children*, *35*(2), 269-290.

Hudson, R. F., Lane, H. B., & Pullen, P. C. (2005). Reading fluency assessment and instruction: What, why, and how?. *The Reading Teacher*, *58*(8), 702-714.

Jacob, B. A. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools. *Journal of Public Economics*, *89*(5), 761-796.

Joseph, L. M., & Nist, L. M. (2006). Comparing the effects of unknown-known ratios on word reading learning versus learning rates. *Journal of Behavioral Education*, *15*(2), 69-79.

Kaniuka, T. S. (2010). Reading achievement, attitude toward reading, and reading self-esteem of historically low achieving students. *Journal of Instructional Psychology*, *37*(2), 184-189.

Koehler, M. J., & Levin, J. R. (1998). Regulated randomization: A potentially sharper analytical tool for the multiple-baseline design. *Psychological Methods*, *3*(2), 206-217.

Kornell, N. (2009). Optimising learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology*, *23*(9), 1297-1317.

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2012). Single-case intervention research design standards. *Remedial and Special Education*, *34*(1), 26-38.

Kuhn, M. R., & Stahl, S. A. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology*, *95*(1), 3-45.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*(2), 293-323.

Lai, S. A., Benjamin, R. G., Schwanenflugel, P. J., & Kuhn, M. R. (2014). The longitudinal relationship between reading fluency and reading comprehension skills in second-grade children. *Reading & Writing Quarterly: Overcoming Learning Difficulties, 30*(2), 116-138.

Landerl, K., & Wimmer, H. (2008). Development of word reading fluency and spelling in a consistent orthography: An 8-year follow-up. *Journal of Educational Psychology*, *100*(1), 150-161.

Leinonen, S., Müller, K., Leppänen, P. H., Aro, M., Ahonen, T., & Lyytinen, H. (2001). Heterogeneity in adult dyslexic readers: Relating processing skills to the speed and accuracy of oral text reading. *Reading and Writing, 14*(3), 265-296.

Limbos, M. M., & Geva, E. (2001). Accuracy of teacher assessments of second-language students at risk for reading disability. *Journal of Learning Disabilities, 34*(2), 136-151.

Maggin, D. M., & Odom, S. L. (2014). Evaluating single-case research data for systematic review: A commentary for the special issue. *Journal of School Psychology, 52*(2), 237-241.

Martin-Chang, S. L., Levy, B. A., & O'Neil, S. (2007). Word acquisition, retention, and transfer: Findings from contextual and isolated word training. *Journal of Experimental Child Psychology, 96*(1), 37-56.

Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning a meta-analysis. *Review of Educational Research, 80*(3), 300-335.

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, 48*(4), 391-407.

Meisinger, E. B., Bradley, B. A., Schwanenflugel, P. J., Kuhn, M. R., & Morris, R. D. (2009). Myth and reality of the word caller: The relation between teacher nominations and prevalence among elementary school children. *School Psychology Quarterly*, *24*(3), 147-159.

Ministry of Education. (2009). *After one year at school: Reading and writing standards.* Retrieved from: http://nzcurriculum.tki.org.nz/National-Standards/Reading-and-writing-standards/The-standards/After-one-year

Ministry of Education. (2014). *Guided reading texts: Ready to Read levelled texts*. Retrieved from: http://literacyonline.tki.org.nz.

Ministry of Education. (2015). *Decile ratings 2015*. Retrieved from: http://www.education.govt.nz/school/running-a-school/resourcing/operational-funding/school-decile-ratings/#Find

Ministry of Education. (2016). *School student population.* Retrieved from: http://www.educationcounts.govt.nz/find-school/school/profile?school=1458&district=7606&region=2.

Muter, V., & Snowling, M. (1998). Concurrent and longitudinal predictors of reading: The role of metalinguistic and short-term memory skills. *Reading Research Quarterly*, *33*(3), 320-337.

Nathan, R. G., & Stanovich, K. E. (1991). The causes and consequences of differences in reading fluency. *Theory Into Practice*, *30*(3), 176-184.

Nation, K., & Cocksey, J. (2009). Beginning readers activate semantics from sub-word orthography. *Cognition*, *110*(2), 273-278.

Nation, K., & Snowling, M. J. (1998). Individual differences in contextual facilitation: Evidence from dyslexia and poor reading comprehension. *Child Development*, *69*(4), 996-1011.

Neumann, M. M., Hood, M., Ford, R. M., & Neumann, D. L. (2011). The role of environmental print in emergent literacy. *Journal of Early Childhood Literacy*, *0*(0), 1-28.

Nist, L., & Joseph, L. M. (2008). Effectiveness and efficiency of flashcard drill instructional methods on urban first-graders' word recognition, acquisition, maintenance, and generalization. *School Psychology Review*, *37*(3), 294-308.

Ogle, L. T., Sen, A., Pahlke, E., Jocelyn, L., Kastberg, D., Roey, S., & Williams, T. (2003). *International comparisons in fourth-grade reading literacy: Findings from the progress in international reading literacy study (PIRLS) of 2001*. Washington, DC: U.S. Government Printing Office.

Ouellette, G. (2010). Orthographic learning in learning to spell: The roles of semantics and type of practice. *Journal of Experimental Child Psychology*, *107*(1), 50-58.

Ouellette, G., & Fraser, J. R. (2009). What exactly is a yait anyway: The role of semantics in orthographic learning. *Journal of Experimental Child Psychology*, *104*(2), 239-251.

Perfetti, C. A., & Lesgold, A. M. (1979). Coding and comprehension in skilled reading and implications for reading instruction. In Resnick, L. B. & Weaver, P. A. (Eds). *Theory and Practice of Early Reading* (Vol 1, pp. 57-84). Mahwah, NJ: Lawrence Erlbaum.

Pikulski, J. J., & Chard, D. J. (2005). Fluency: Bridge between decoding and reading comprehension. *The Reading Teacher*, *58*(6), 510-519.

Raymond, J. (1998). *Butterfly day*. Wellington, NZ: Ministry of Education.

Richards, M. (2000). Be a good detective: Solve the case of oral reading fluency. *The Reading Teacher*, *53*(7), 534-539.

Samuels, S. J., & Flor, R. F. (1997). The importance of automaticity for developing expertise in reading. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, *13*(2), 107-121.

Schwanenflugel, P. J., Hamilton, A. M., Kuhn, M. R., Wisenbaker, J. M., & Stahl, S. A. (2004). Becoming a fluent reader: Reading skill and prosodic features in the oral reading of young readers. *Journal of Educational Psychology, 96*(1), 119-129

Seymour, P. H., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*(2), 143-174.

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition, 55*(2), 151-218.

Share, D. L. (2004). Orthographic learning at a glance: On the time course and developmental onset of self-teaching. *Journal of Experimental Child Psychology, 87*(4), 267-298.

Siegle, D., Moore, M., Mann, R. L., & Wilson, H. E. (2010). Factors that influence in-service and preservice teachers' nominations of students for gifted and talented programs. *Journal for the Education of the Gifted*, *33*(3), 337-360.

Skinner, C. H., Fletcher, P. A., Wildmon, M., & Belfiore, P. J. (1996). Improving assignment preference through interspersing additional problems: Brief versus easy problems. *Journal of Behavioral Education, 6*(4), 427-436.

Smith, J. D. (2012). Single-case experimental designs: A systematic review of published research and current standards. *Psychological Methods*, *17*(4), 510-550.

Sprenger-Charolles, L., & Siegel, L. S. (2013). Reading acquisition and dyslexia in languages varying in orthographic depth: From behavior to brain. *Perspectives on Language and Literacy*, *39*(1), 23-31.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*(4), 360-407.

Tan, A., & Nicholson, T. (1997). Flashcards revisited: Training poor readers to read words faster improves their comprehension of text. *Journal of Educational Psychology, 89*(2), 276-288.

Therrien, W. J. (2004). Fluency and comprehension gains as a result of repeated reading a meta-analysis. *Remedial and Special Education, 25*(4), 252-261.

Thomas, J. R., Salazar, W., & Landers, D. M. (1991). What is missing in p<. 05? Effect size. *Research quarterly for exercise and sport, 62*(3), 344-348.

Tønnessen, F. E., & Uppstad, P. H. (2015). *Can we read letters? Reflections on fundamental issues in reading and dyslexia research*. Rotterdam, Netherland: Sense Publishing.

Tunmer, W. E., & Nicholson, T. (2011). The development and teaching of word recognition skill. In M. L. Kamil, P. D. Pearson, E.B. Moje & P. Afflebach (Eds.), *Handbook of reading research* (pp. 405-431). Abingdon, OX: Routledge.

van den Boer, M., & de Jong, P. F. (2015). Parallel and serial reading processes in children's word and nonword reading. *Journal of Educational Psychology, 107*(1), 141-151.

Walczyk, J. J., & Griffith-Ross, D. A. (2007). How important is reading skill fluency for comprehension?. *The Reading Teacher, 60*(6), 560-569.

Wampold, B. E., & Worsham, N. L. (1986). Randomization tests for multiple-baseline designs. *Behavioral Assessment. 8*(2), 135-143.

Watts, Z., & Gardner, P. (2013). Is systematic synthetic phonics enough? Examining the benefit of intensive teaching of high frequency words (HFW) in a year one class. *Education 3-13, 41*(1), 100-109.

Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology*, *91*(3), 415-438.

Appendix A: Information and consent forms

MASSEY UNIVERSITY
INSTITUTE OF EDUCATION
TE KURA O TE MATĀURANGA

Title of study: Does training children to read frequently occurring words faster improve their reading speed, accuracy, and comprehension?

INFORMATION SHEET FOR PARENT(S)/CAREGIVER(S)

My name is Butul Iqbal and I would like to invite your child to participate in a research study on reading that I will be undertaking for my master's thesis in educational psychology.

Reason for the study:
Many of the words that children read in their books at school are repeated again and again. These commonly used words in books (e.g., was, where, could) are called high frequency words. The most frequently occurring 220 words in children's books account for more than six out of ten of all the words that children read in books. The aim of my study is to investigate whether teaching these high frequency words increases reading speed, accuracy and comprehension. Findings from this research will contribute to gaining a better understanding of how to improve reading comprehension.

What the study will involve:
If your child agrees to participate, they will attend 15 lessons for about 30 minutes each, in which your child will go through a short list of words, read a short passage and answer some comprehension questions. In some lessons your child will be reading through the list of words once before reading the passage and answering the comprehension questions. In other lessons, your child will be trained to read the words as quickly as possible before moving onto reading the passage and answering the comprehension questions. Each reading lesson will take place in a quiet area at school that is close to your child's classroom and classmates.

We will have one 30 minute lesson per day for up to 15 days. Your child will be asked for their assent before I begin the lesson each day and they will be able to accept or decline participation. Your child will also be allowed to go back to the classroom at any time if they wish to during the lesson.

There is no obligation for you to accept this invitation. It will not affect your child in any way at school if you do not want your child to participate. If you do participate you can still withdraw your child from the study at any time up until it finishes at the end of July 2016.

If you agree to your child being in the study, please sign and return the attached Consent Form. Please indicate on the consent form if you would like a summary of the results of the study and I will provide this to you when the study is complete.

If you allow your child to participate, both you and your child have the right to:

- decline to answer any particular question
- ask any questions at any time during the study
- provide information on the understanding that your name and your child's name will not be used unless you give permission to the researcher

Your child's test scores will not be shared with anyone except the researcher and the researcher's supervisors. Data and Consent Forms will be stored separately in a locked cabinet for 6 years. At the end of this time the researcher's supervisors will dispose of the data and Consent Form.

Ethical Approval
This project has been evaluated by peer review and judged to be low risk. Consequently, it has not been reviewed by one of the University's Human Ethics Committees. The researcher named above is responsible for the ethical conduct of this research.

If you have any concerns about the conduct of this research that you wish to raise with someone other than the researcher(s), please contact Dr Brian Finch, Director (Research Ethics), telephone 06 356 9099 extn 86015, email humanethics@massey.ac.nz.

Project Contacts
If you have any questions or require further information, please feel free to contact myself or my supervisors:

Butul Iqbal
Postgraduate student
Massey University at Albany
butoolis@gmail.com
02102278599
Supervisors:
Prof Tom Nicholson
Professor of Literacy Education
Institute of Education
Massey University at Albany
T.Nicholson@massey.ac.nz
09 414-0800 ext 43519
Mob: 021 1085 923

Dr Jane Prochnow
Senior Lecturer
Institute of Education
Massey University at Albany
J.E.Prochnow@massey.ac.nz
(06) 356 9099 ext. 84463

Title of study: Does training children to read frequently occurring words faster improve their overall reading speed, accuracy, and comprehension?

CONSENT FORM FOR PARENT(S)/CAREGIVER(S)

- I have read the Information Sheet and have had the details of the study explained to me. My questions have been answered to my satisfaction, and I understand that I may ask further questions at any time.
- I agree to my child participating in this study under the conditions set out in the Information Sheet, and understand that my child will also be asked for his/her assent.

Your signature: _____ Date: _____

Your full name (printed): _____

Your relationship to child: _____

Your contact phone (optional): _____

Your contact street address (optional): _____

Your child's name: _____

Your child's date of birth: _____ Ethnicity (optional): _____

MASSEY UNIVERSITY
INSTITUTE OF EDUCATION
TE KURA O TE MATĀURANGA

Title of study: Does training children to read frequently occurring words faster improve their reading speed, accuracy, and comprehension?

INFORMATION SHEET FOR PRINCIPAL AND BOARD OF TRUSTEES

Dear Principal,

My name is Butul Iqbal and I would like to invite your school to participate in a research study on reading that I will be undertaking for my master's thesis in educational psychology.

Reason for the study:
Many of the words that children read in their books at school are repeated again and again. These commonly used words in books are called high frequency words. The most common 220 words, for example, account for more than six out of ten of all the words that children read in books. The aim of my study is to investigate whether teaching these high frequency words increases reading speed, accuracy and comprehension. Findings from this research will contribute to gaining a better understanding of how to improve reading comprehension.

Nature of the study:
It would involve **five** children who are **eight to nine years old** reading at Reading Recovery levels of **15** to **18** (6-7 year reading levels).

Participation for each child will involve 15 lessons. Each lesson will take about 30 minutes in which I will ask the child to go through a list of high frequency words, read a short passage and answer some comprehension questions. In some lessons the child will be reading through the list of words once before going onto reading the passage and answering the comprehension questions. In other lessons, the child will be trained to read the words as quickly as possible before moving on to reading the passage and answering the comprehension questions. These two types of lessons will make up the baseline and experimental observations.

I will ask for the child's consent before I begin each lesson, which they may accept or decline. He/she will also be allowed to go back to the classroom at any time if they wish during the lesson.

I will need access to a **quiet area** to carry out the lesson.

I have a parent/caregiver consent and information form that can be sent home with the child to explain the study to the parents and to seek their consent.

For ethical reasons, the child's assessments are confidential to the researcher and the researcher's supervisors. Data and Consent Forms will be stored separately in a locked cabinet for 6 years. At the end of this time the researcher's supervisors will dispose of the data and Consent Form.

Ethical Approval

This project has been evaluated by peer review and judged to be low risk. Consequently, it has not been reviewed by one of the University's Human Ethics Committees. The researcher named above is responsible for the ethical conduct of this research.

If you have any concerns about the conduct of this research that you wish to raise with someone other than the researcher, please contact Dr Brian Finch, Director (Research Ethics), telephone 06 356 9099 extn 86015, email humanethics@massey.ac.nz.

Project Contacts

If you have any questions or require further information, please feel free to contact myself or my supervisors:

Butul Iqbal
Postgraduate student
Massey University at Albany
butoolis@gmail.com
02102278599

Supervisors:

Prof Tom Nicholson
Professor of Literacy Education
Institute of Education
Massey University at Albany
T.Nicholson@massey.ac.nz
09 414-0800 ext 43519
Mob 021 1085 923

Dr Jane Prochnow
Senior Lecturer
Institute of Education
Massey University at Albany
J.E.Prochnow@massey.ac.nz
(06) 356 9099 ext. 84463

Title of study: Does training children to read frequently occurring words faster improve their overall reading speed, accuracy, and comprehension?

CONSENT FORM FOR PRINCIPAL

- I have read the Information Sheet and have had the details of the study explained to me. My questions have been answered to my satisfaction, and I understand that I may ask further questions at any time.

- I agree to the children at my school participating in this study under the conditions set out in the Information Sheet, and understand that they and their parents will also be asked for their consent.

Your signature: _____ Date: _____

Your full name (printed): _____

Full names of the child/children nominated to participate:

_____

_____

_____

_____

_____

MASSEY UNIVERSITY
INSTITUTE OF EDUCATION
TE KURA O TE MATĀURANGA

Title of study: Does training children to read frequently occurring words faster improve their reading speed, accuracy, and comprehension?

ASSENT FOR CHILDREN

My name is Butul Iqbal and I am studying at Massey University. I am trying to find a good way to help children read. With your help I want to find out if flashcards would be useful to help you learn new words quickly. A flashcard is a small card with a word on it.

The lesson with flashcards will take 30 minutes each day. In each lesson we will work on learning as many words as you can, then read a story and answer some questions about the story. We will have one lesson per day for 15 days. This is not a test so I will not tell anyone about how you did in the lessons, but you are allowed to tell your teacher whether you liked the lesson or not.

You do not have to do this activity if you do not want to. If you do agree to take part in it but change your mind, you can stop at any time.
If you have any questions you can ask them anytime.
You don't have to answer a question I ask you if you don't want to.
If you would like to be in this study please write your name on this page to show that you would like to be in the study.
My name is _____

Appendix B – Summary of Reading levels

Table 10.

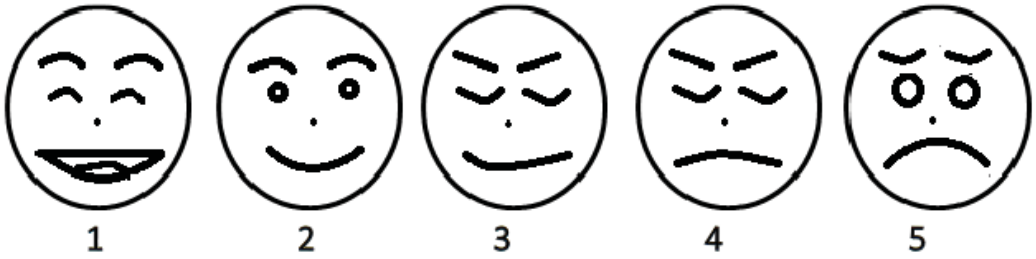
Summary of the Ready to Read levelled texts (as depicted by the colour wheel), Reading Recovery levels, and student ages based on the National Standards for Literacy Achievement (Ministry of Education, 2009).

Ready to Read level	Reading Recovery level	Age (years)
White	23-26	8-9
Emerald	23-26	8-9
Gold	21-22	7-8
Purple	19-20	7-8
Turquoise	17-18	6-7
Orange	15-16	6-7
Green	12-13	5-6
Blue	9-11	5-6
Yellow	6-8	5-6
Red	3-5	5-6
Magenta	0-2	5-6

Appendix C - Student Evaluation form

Colour in or circle the smiley that best shows how you feel.

1. How did you feel about the lesson?



2. How did you feel after reading the story?

3. How do you think your tutor feels when you read a story?

4. Would you like to come for another lesson?

Appendix D - Sample comprehension questions

1. What did the Weka find (E)?

a. A piece of bark

b. A stick

c. a black tail feather with strip

2. What did Weka do with the feather (E)?

a. Kept it for herself

b. **Ate it**

c. Tried to find its owner

3. What kind of animal is Weka looking for to return the feather to? (I)

a. A horse

b. A bird

c. A Dolphin

4. Who did Weka meet first (E)?

a. His friend pigeon

b. A nosy owl

c. Kahu the harrier Hawk

5. Why did Weka want to return the feather? (I)

a. Because birds can't survive without their feathers

b. Because it was beautiful and Weka thought it should be returned.

c. Because it was Weka's job

Questions	Child 1	Child 2	Child 3	Child 4	Child 5
1					
2					
3					
4					
5					

Appendix E - Statistical Analysis

Table 11.

Table of obtained test statistics for each participant across each measure of interest.

	Participant 1	Participant 2	Participant 3	Participant 4	Participant 5	Overall test statistic
Passage accuracy	-0.01	-0.06	0.00	0.09	-0.05	-0.02
Passage speed	-0.05	-0.13	-0.02	0.03	-0.09	-0.26
Passage comprehension	0.02	-0.02	0.08	0.13	0.06	0.27
Word accuracy	0.94	0.73	0.88	0.95	0.53	4.04
Word transfer	0.44	0.29	0.08	0.292	0.071	1.17

Note. The obtained test statistics represents the mean of the intervention phase subtracted by the mean of the baseline phase for each participant across each measure of interest. The overall test stat is the sum of the mean differences for each measure of interest (Koehler & Levin, 1998; Wampold & Worsham, 1986).

Table 12.

Table of 10 out of 120 (5! = 120) possible combinations of starting points.

	1	2	3	4	5	6	7	8	9	10
Participant 1	4	8	8	8	8	8	8	8	8	8
Participant 2	5	7	7	7	7	7	6	6	6	6
Participant 3	6	6	5	5	4	4	7	7	5	5
Participant 4	7	4	6	4	5	6	5	4	7	4
Participant 5	8	5	4	6	6	5	4	5	4	7

Note. Combination 1 depicts the obtained random allocation where participant 1 started at story 4, participant 2 at story 5, participant 3 at story 6 and so on. The rest depicts the multiple other possibilities of allocations that could have arisen.

Table 10.

All 120 possible test statistics computed for word reading accuracy.

4.040	3.747	3.736	3.553	3.651	3.541	3.760	3.467	3.476
3.367	3.464	3.281	3.506	3.397	3.486	3.303	3.118	3.191
3.506	3.323	3.203	3.130	3.413	3.303	3.675	3.382	3.372
3.189	3.287	3.177	3.395	3.102	3.126	3.069	3.167	2.931
3.156	3.099	3.122	2.939	2.768	2.893	3.209	2.973	2.905
2.780	3.049	2.939	3.525	3.232	3.256	3.199	3.297	3.061
3.525	3.232	3.242	3.132	3.230	3.047	2.992	2.882	3.006
2.949	2.763	2.711	2.992	2.809	2.723	2.775	3.059	2.823
3.256	3.199	3.222	3.039	2.868	2.993	2.976	2.919	2.962
2.852	2.681	2.733	3.242	3.132	3.222	3.039	2.853	2.927
2.723	2.775	2.689	2.615	2.630	2.755	3.278	3.042	2.974
2.849	3.118	3.008	2.998	2.762	2.714	2.662	2.931	2.748
2.744	2.692	2.724	2.599	2.584	2.658	3.211	3.028	2.908
2.834	3.118	3.008						

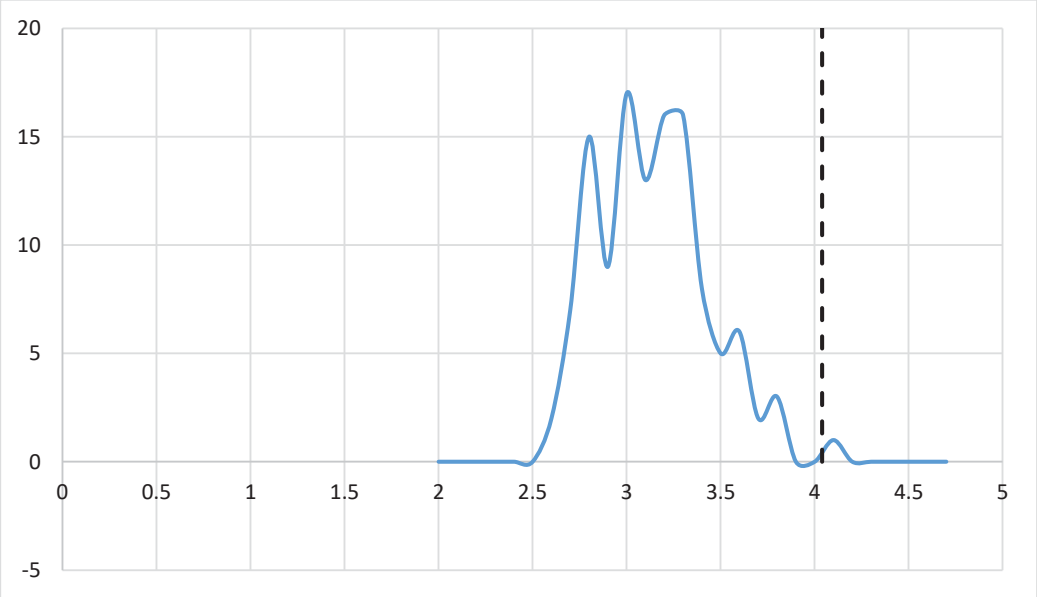


Figure 17. Distribution of 120 test statistics against the obtained test statistic of 4.04 (depicted by the vertical line) for word reading accuracy, resulting in a p-value of 0.01. Note: A smooth trend line was applied to this distribution.

Table 11.

All 120 possible test statistics computed for word reading transfer

1.178	1.082	1.174	1.106	1.284	1.258	1.106	1.011	1.106
1.011	1.189	1.189	0.969	0.874	0.966	0.897	1.076	1.049
0.945	0.945	0.942	0.969	0.969	0.942	0.841	0.745	0.837
0.769	0.947	0.921	0.894	0.799	0.858	0.811	0.989	0.942
0.722	0.674	0.754	0.685	0.828	0.849	0.745	0.698	0.742
0.721	0.757	0.730	0.922	0.827	0.887	0.839	1.018	0.970
0.797	0.702	0.797	0.702	0.881	0.881	0.714	0.619	0.678
0.631	0.809	0.762	0.690	0.690	0.655	0.702	0.702	0.655
0.730	0.682	0.762	0.694	0.837	0.858	0.658	0.611	0.694
0.599	0.742	0.789	0.641	0.545	0.637	0.569	0.747	0.599
0.721	0.498	0.545	0.530	0.557	0.521	0.542	0.959	0.911
0.955	0.934	0.970	0.943	0.887	0.839	0.887	0.839	0.875
0.875	0.750	0.703	0.747	0.726	0.762	0.735	0.822	0.822
0.818	0.845	0.845						

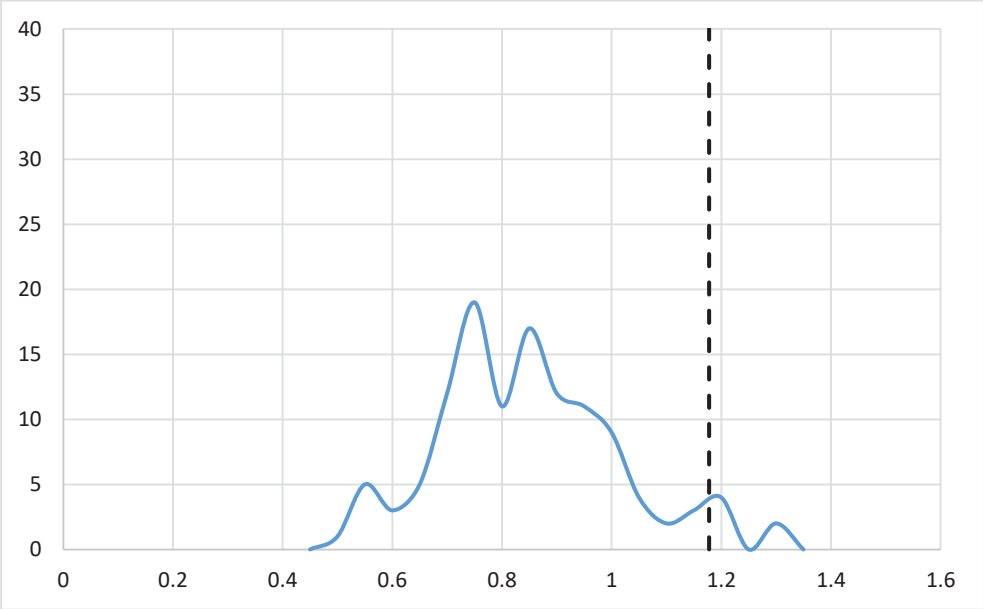


Figure 18. Distribution of 120 test statistics against the obtained test statistic of 1.17 (depicted by the vertical line) for word reading transfer, leading to a p-value of 0.03.

Table 12.

All 120 possible test statistics computed for passage reading accuracy

-0.023	-0.014	-0.010	-0.004	0.039	0.042	-0.038	-0.028	-0.044
-0.009	0.034	0.009	-0.069	-0.035	-0.050	-0.044	-0.007	0.021
0.004	-0.021	0.018	-0.011	-0.005	-0.001	-0.043	-0.034	-0.030
-0.024	0.019	0.022	-0.065	-0.056	-0.049	0.016	0.059	0.003
-0.075	-0.009	-0.077	-0.071	-0.013	0.047	0.030	-0.026	0.043
-0.016	-0.032	-0.029	-0.064	-0.055	-0.048	0.018	0.061	0.004
-0.057	-0.047	-0.063	-0.028	0.015	-0.010	-0.110	-0.076	-0.088
-0.023	0.014	-0.017	-0.036	-0.061	-0.020	0.011	0.017	-0.039
-0.083	-0.017	-0.085	-0.079	-0.021	0.039	-0.097	-0.032	-0.119
-0.085	-0.026	0.005	-0.097	-0.063	-0.078	-0.072	-0.035	-0.007
-0.055	-0.024	-0.058	-0.086	-0.064	-0.005	0.029	-0.028	0.042
-0.018	-0.034	-0.030	0.014	-0.042	0.008	-0.023	-0.039	-0.064
-0.018	-0.049	0.002	-0.058	-0.080	-0.051	-0.017	-0.042	-0.004
-0.033	-0.027	-0.023						

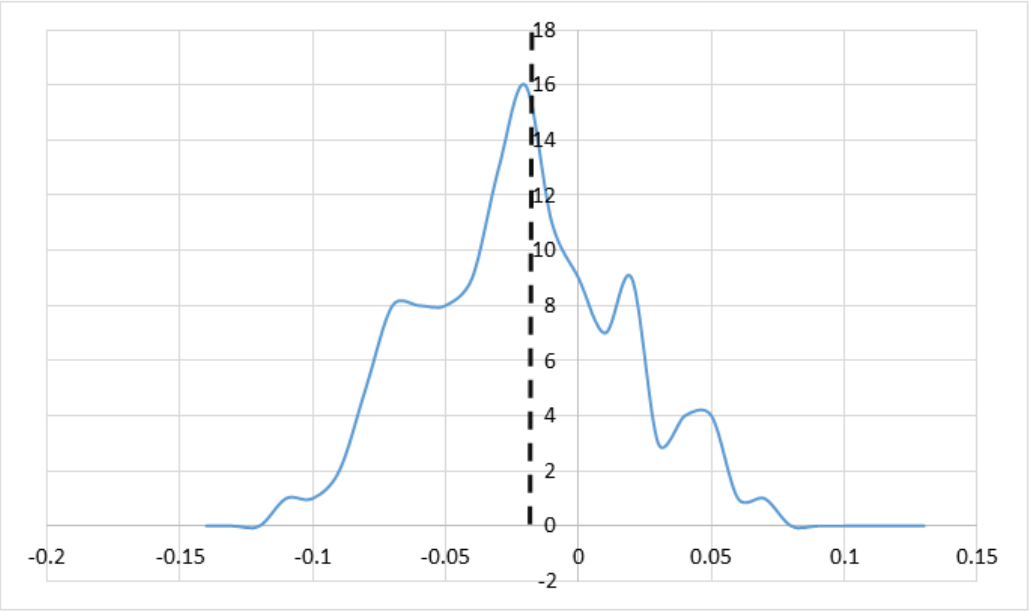


Figure19. Distribution of 120 test statistics against the obtained test statistic of -0.023 (depicted by the vertical line) for passage reading accuracy, results in a p-value of 0.54.

Table 13.

All 120 possible test statistics computed for passage reading speed.

-0.26	-0.25	-0.29	-0.22	-0.22	-0.27	-0.26	-0.24	-0.26
-0.14	-0.14	-0.24	-0.27	-0.15	-0.29	-0.22	-0.22	-0.17
-0.14	-0.24	-0.17	-0.22	-0.21	-0.27	-0.29	-0.28	-0.32
-0.25	-0.24	-0.30	-0.28	-0.27	-0.29	-0.13	-0.12	-0.27
-0.29	-0.14	-0.31	-0.25	-0.25	-0.16	-0.13	-0.27	-0.16
-0.25	-0.24	-0.29	-0.26	-0.25	-0.27	-0.11	-0.10	-0.25
-0.26	-0.25	-0.27	-0.15	-0.14	-0.25	-0.26	-0.15	-0.27
-0.11	-0.11	-0.15	-0.14	-0.24	-0.15	-0.11	-0.10	-0.24
-0.32	-0.17	-0.34	-0.28	-0.28	-0.18	-0.32	-0.16	-0.31
-0.20	-0.20	-0.16	-0.32	-0.21	-0.34	-0.28	-0.28	-0.23
-0.20	-0.16	-0.22	-0.27	-0.27	-0.18	-0.15	-0.30	-0.18
-0.27	-0.26	-0.31	-0.15	-0.29	-0.15	-0.19	-0.18	-0.29
-0.16	-0.20	-0.18	-0.27	-0.27	-0.22	-0.19	-0.30	-0.22
-0.27	-0.27	-0.32						

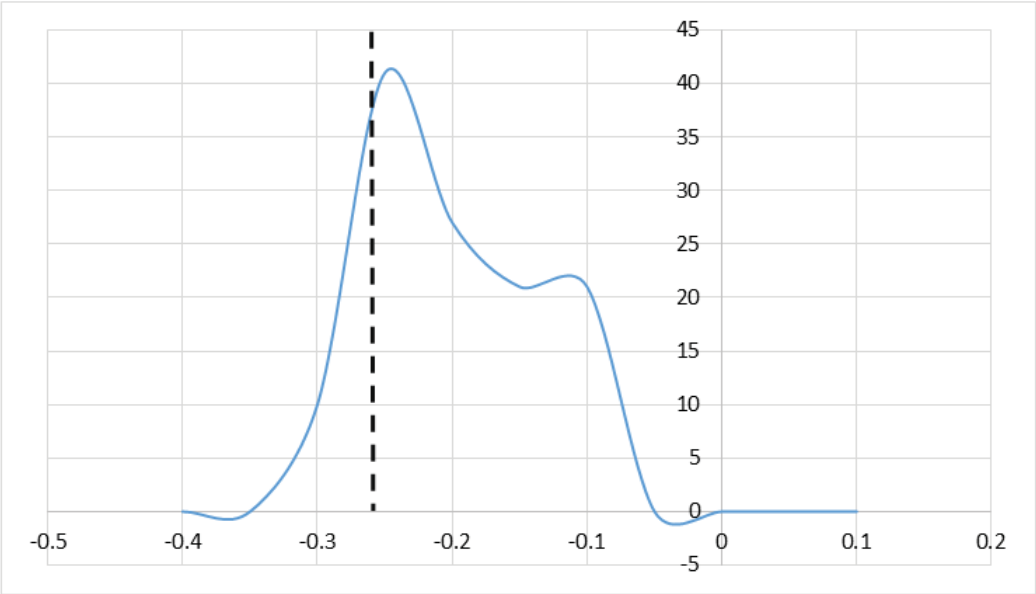


Figure 190. Distribution of 120 test statistics against the obtained test statistic of -0.26 (depicted by the vertical line) for passage reading speed, leading to a p-value of 0.38.

Table 14.

All 120 possible test statistics computed for passage reading comprehension

0.27	0.32	0.38	0.38	0.31	0.36	0.14	0.19	0.35
0.29	0.22	0.33	0.37	0.31	0.26	0.26	0.41	0.34
0.16	0.27	0.26	0.33	0.11	0.16	0.07	0.12	0.17
0.17	0.10	0.15	0.24	0.29	0.33	0.26	0.19	0.31
0.35	0.28	0.37	0.37	0.39	0.31	0.13	0.25	0.23
0.30	0.22	0.27	0.30	0.35	0.39	0.32	0.25	0.37
-0.01	0.04	0.21	0.14	0.07	0.19	0.40	0.34	0.27
0.20	0.35	0.36	0.19	0.30	0.27	0.27	0.05	0.17
0.47	0.39	0.48	0.48	0.50	0.43	0.33	0.26	0.46
0.39	0.41	0.40	0.28	0.22	0.17	0.17	0.32	0.25
0.35	0.34	0.37	0.43	0.30	0.23	0.31	0.44	0.41
0.49	0.40	0.45	0.17	0.30	0.39	0.40	0.31	0.43
0.41	0.42	0.30	0.37	0.50	0.44	0.14	0.25	0.24
0.31	0.09	0.14						

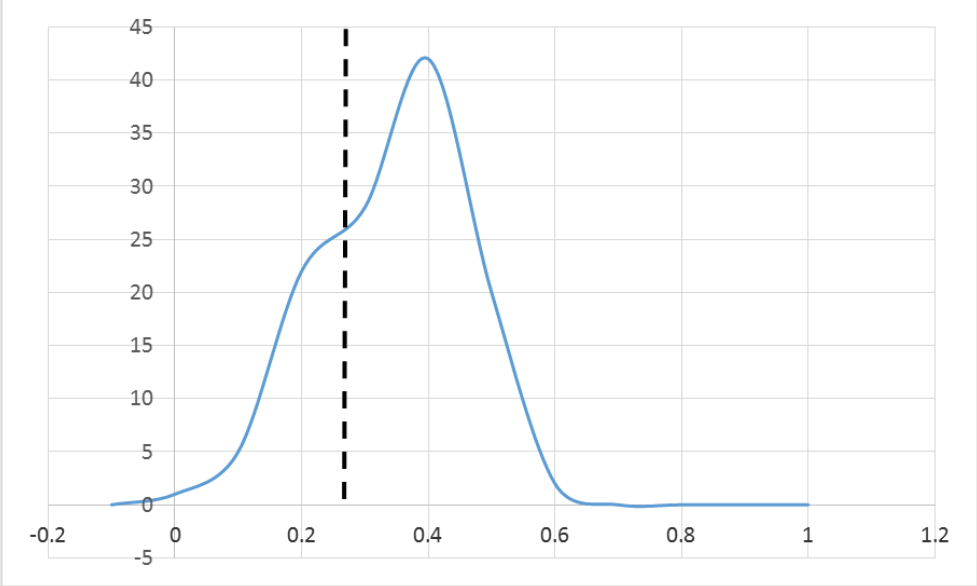


Figure 21. Distribution of 120 test statistics against the obtained test statistic of 0.27 (depicted by the vertical line) for passage reading comprehension, resulting in a p-value of 0.41.