

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Predicting Reading Recovery
Selection and Outcomes -
Is It Possible?

A thesis presented in partial fulfilment of the
requirements for the degree of
Master of Education (Literacy)
at Massey University, Manawatū, New Zealand

Athena London

2017

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Abstract

Purpose: Predictive early literacy assessments are useful to identify students who are at risk of reading difficulty. This study investigated the use of six early literacy assessments, administered when students first entered school (Time 1), and in the middle of their first year at school (Time 2), in order to predict which students would be selected for Reading Recovery and to identify the Reading Recovery (RR) outcomes for students who participated in the intervention.

Method: Unpublished data from a longitudinal study (Early Literacy Project, Chapman, Arrow, Tunmer, & Braid, 2016) was analysed to find predictive links between assessment results and later reading outcomes, for a cohort of 300 5-year-old children in New Zealand primary schools.

Results: It was not possible to predict which students would be selected for Reading Recovery due to the variations in RR selection processes. It was found that children who participated in RR were more likely to be referred on for further support the lower their phonological awareness scores were. It was also found that if a child scored 20 points or less, in a combination of Time 1 assessments (letter names, letter sounds and three measures of phonological awareness), they were likely to have a body of literacy abilities that meant they would be working at least a year below the National Standard by the end of their second year at school.

Implications: The findings indicate that standardising the selection of students for RR may mean students with the lowest literacy attainment all get support. In addition, early literacy assessments, including measures of phonological awareness, should be administered early in a child's schooling and those identified as being at risk of reading difficulty should receive literacy support without delay. Addressing students' low levels of phonological awareness in the first year of schooling may lead to better outcomes for students who participate in RR.

Keywords: *phonological awareness, Reading Recovery, early literacy assessment, letter names, letter sounds, New Zealand, timing of assessment, vocabulary*

Acknowledgements

This study is the result of a team of people who supported me along the journey. Firstly, I would like to thank Dr Alison Arrow and Prof Tom Nicholson who have supervised this study. Their guidance has supported me to dig deeper and I have learnt so much through the process. I would also like to thank the leaders of the Early Literacy Project, James Chapman, Alison Arrow, William Tunmer and Christine Braid, for sharing their unpublished data with me. For guidance about analysis, I would like to thank Dr Jonathan Godfrey from the Statistical Consultancy Service. I would also like to thank Dr Penny Bilton for support deciphering SPSS and for her encouragement.

I would like to acknowledge TeachNZ, who awarded me study leave from my role as a Resource Teacher of Literacy so I could investigate my burning questions. These study awards are so valuable to the teachers who receive them – thanks TeachNZ!

Finally, I need to thank all my extended whanau for their tireless support and encouragement. I am grateful to Tim, who convinced me I was capable of post graduate study. To my sons, Eric, Jonathan, and David: thank you for patiently listening to my latest findings, and for taking charge of the extra chores so I could be free to type. And thank you Richard, for setting up quiet spaces in the midst of renovations so I could study, for technical support at all hours of the day and night, and for always believing in me.

Thanks whanau!

Contents

| | |
|---|-------------|
| List of Tables | viii |
| List of Figures | ix |
| Introduction | 1 |
| Rationale | 2 |
| Thesis overview | 4 |
| Key terms and definitions | 4 |
| Literature Review | 6 |
| Scope of review | 6 |
| Organisation of review | 7 |
| Models of reading acquisition | 8 |
| Skills needed for successful reading | 14 |
| Early literacy assessment | 16 |
| The ideal time to administer early literacy assessments | 19 |
| Reading Recovery | 20 |
| The 13% - what is known about students who are referred on? | 25 |
| Treatment resisters | 28 |
| Summary | 32 |
| The present study | 36 |
| Methodology | 38 |
| Participants | 38 |
| Design | 41 |
| Data collection | 42 |
| Measures | 43 |
| Results | 48 |
| Analysis | 48 |
| Predicting selection for Reading Recovery | 52 |
| Predicting Reading Recovery outcomes | 61 |
| Summary | 67 |
| Discussion | 69 |
| Predicting Reading Recovery selection | 69 |

| | |
|--|-----------|
| Predicting Reading Recovery outcomes | 72 |
| Reading outcomes for the cohort..... | 73 |
| Limitations of this study..... | 74 |
| Conclusions and Implications..... | 77 |
| Implications for future research | 79 |
| Implications for New Zealand educators | 81 |
| Summary | 83 |
| APPENDIX | 85 |
| Means and standard deviations for all groups and all variables | 85 |
| References | 86 |

List of Tables

| | |
|---|------------|
| Table 1 - Profile of ELP Participants in 2015..... | 40 |
| Table 2 - Minimum Number of Cases Required to Run Logistic Regression | 50 |
| Table 3 - Pearson Correlation Matrix of Time 1 Assessment Tasks..... | 53 |
| Table 4 - Pearson Correlation Matrix of Time 2 Assessment Tasks..... | 55 |
| Table 5 - Logistic Regression Predicting Likelihood of Reading Recovery Selection Based on Knowledge of Letter Names, Letter Sounds, Phonological Awareness and Vocabulary at Time 1 | 157 |
| Table 6 - Logistic Regression Predicting Likelihood of Reading Recovery Selection Based on Knowledge of Letter Names, Letter Sounds and Phonological Awareness at Time 2 | 59 |
| Table 7 - Summary of Time 1 (T1) Combined Scores for Students Who Have Exited Reading Recovery..... | 64 |
| Table 8 - Students Predicted as Having Reading Difficulty by the End of Year 2 Using the Time 1 Combined Score | 66 |

List of Figures

| | |
|--|-----------|
| Figure 1 - Correlation Between Pairs of Time 1 Assessment Tasks | 54 |
| Figure 2 - Correlation Between Pairs of Time 2 Assessment Tasks | 56 |
| Figure 3 - Reading Levels at Time 3 (T3) When Students Began Entering Reading Recovery (RR) | 60 |
| Figure 4 - Comparison of Scores for the Successful and Unsuccessful RR students for Time 1 (T1) and Time 2 (T2) Variables That Showed the Greatest Difference Between the Groups..... | 62 |
| Figure 5 - Time 1 (T1) Combined Assessment Scores Compared with Time 5 (T5) Reading Levels | 65 |

Introduction

The ability to read underpins learning in all areas of the curriculum (Ministry of Education, 2010). For those who struggle to read, the negative Matthew effects (the rich get richer while the poor get poorer, Stanovich, 1986) lead to an ever-widening achievement gap. Because of the negative consequences of reading difficulties, it is crucial to intervene early to identify students who will need additional literacy support. Having identified these students, it is then important to ensure that the support provided is matched to the needs of the student, in order for the child to make the greatest gains possible. If students with literacy difficulties can be identified early, then support can begin early, thus reducing the subsequent achievement gap.

Reading Recovery is a one-to-one literacy intervention for children who are not making the expected progress in reading (Clay, 1993). The aim of this school-based supplementary instruction is to bring each child's reading level up to the average level of their peers so the child's needs can then be met as part of the classroom programme. In New Zealand (NZ), this programme is the only early literacy intervention subsidised by the Ministry of Education. At most recent count, two thirds of New Zealand primary schools offered Reading Recovery (Ministry of Education, 2014).

This thesis aims to fill a gap in current research regarding Reading Recovery student selection so that schools can more accurately identify those students who are most at risk of reading difficulty. If schools can identify the most vulnerable students, they can ensure

RR is provided to the students who need it most. Furthermore, if the student's needs can be identified at an early stage, support can begin earlier rather than the current NZ model of waiting until students turn six before support in the form of Reading Recovery is offered. The findings of this study could have implications for early literacy assessment and intervention in New Zealand primary schools.

Rationale

As a Resource Teacher of Literacy, the author frequently worked with students referred on from Reading Recovery for additional specialist literacy support. These students were not able to make the required progress in the Reading Recovery (RR) intervention in order to be successfully discontinued. It was observed these referred students often had low self-esteem and they actively avoided reading. This observations corresponds with the findings of Chapman, Tunmer, and Prochnow (2001) that children who exit RR and need to be referred for further help, experienced a decline in self-concept and perceptions of their own reading ability. In order to help these students, the author found it was often necessary to regress to very basic reading levels in order for the children to regain confidence in their own reading ability. The time taken to go back to such an early level means that these vulnerable students are put at greater risk, due to the widening achievement gap between them and their peers. Therefore, it is important to offer these students the most appropriate form of support, at the most appropriate time, to ensure they have the greatest opportunity to succeed as readers.

The current process for selecting students for Reading Recovery uses data from the Six Year net testing (see An Observational Survey of Early Literacy Achievement, Clay, 1993). This assessment battery is generally used in New Zealand primary schools with all children turning six. The students with the lowest overall assessments scores in each school are selected for the Reading Recovery intervention. Because this assessment only takes place after a full year of reading instruction, children at-risk of reading difficulty may not be identified early enough. This is a weakness of the current system and it is important to ask whether there is a better way to support the most vulnerable readers.

Using predictive literacy assessments at an earlier stage in primary schools would enable teachers to identify those most at risk of reading difficulty from the very beginning of their schooling, rather than waiting for them to be struggling by age six. This has several implications for schools; if teachers can identify which 5-year-old students are likely to be selected for Reading Recovery, schools may be able to offer Reading Recovery earlier. Another important implication of this study is that if effective early literacy assessment could identify which areas of literacy need to be strengthened for reading success, schools could ensure they were offering interventions that were targeted to the needs of the learners. In addition, appropriate intervention support, given to children who are aged five, may lead to better outcomes if the children participate in RR when they are six.

Thesis overview

This thesis is comprised of five chapters. The second chapter provides an overview of current research in the areas of early literature acquisition and assessment, and includes three models of literacy acquisition. Following the literature review, the methodology of this study is explained. Next, the results are analysed for correlations between each child's early literacy profile and their subsequent selection for Reading Recovery (and the outcomes for the children selected for RR). Finally, there is a discussion of the findings from this study, and the implications for teachers and researchers are identified. This work addresses a gap in current education research regarding New Zealand's Reading Recovery selection process, and aims to prove the value of administering early predictive assessments in literacy to ensure schools select the right students, for the right support, at the right time.

Key terms and definitions

Alphabetic Principle – an understanding that there are systematic and predictable relationships between letters and sounds. The knowledge that spoken words are made up of phonemes and written words are made up of letters. Each letter, or combination of letters, relates to a sound or sounds.

Letter names - the names of the symbols used in an alphabetic orthography to represent speech sounds.

Letter sounds - the sounds associated with individual letters and letter combinations in an alphabetic orthography.

Oral Language - the system through which people use spoken words to express knowledge, ideas, and feelings.

Phonological Awareness (PA) - the conscious understanding that spoken language is made up of various units of sound. These include phonemes, onset-rime units, syllables, words and sentences. Phonological awareness develops over time (Blachman, 2000) and comprises the ability to attend to, reflect on and manipulate these sound units in speech (Catts & Kamhi, 2005). Phonemic awareness is a subset of phonological awareness.

Phonological Processing (PP) – an auditory processing skill that relates to words, but occurs in the absence of print. It involves detecting and discriminating differences in phonemes or speech sounds in words, and includes phonological memory and rapid automatic naming skill. Phonological Awareness is a subset of phonological processing.

Literature Review

This review examined recent literature on reading acquisition to look for predictive measures that can identify which students are likely to be selected for Reading Recovery (RR), and also predict their RR outcomes. There is a vast quantity of national and international literature on reading development and assessment (e.g. Blachman, 2000; Chapman et al., 2001; Clay, 1991; Ehri, 2005; Greaney & Arrow, 2012; Torgesen, 1998). Therefore, in this review it was necessary to limit texts to those specifically investigating (a) current RR selection process and RR student outcomes, (b) the skills needed for reading success, (c) assessments to identify students at-risk of literacy difficulties, (d) the ideal timing of literacy screening, and (e) students who do not respond to interventions.

Scope of review

In total, 60 texts were reviewed, including a range of primary studies, meta-analyses, journal articles, and edited books, offering prominent viewpoints from the last 30 years. Literacy research is advancing quickly, therefore texts written before 1984 were not considered because the research findings are likely to have been superseded. This review focussed predominantly on quantitative research regarding literacy assessment tasks and results. Three models of reading development that underpin this study are explained: Hoover and Gough's (1990) simple view of reading, Spear-Swerling's (2013) road map for understanding reading difficulties, and Ehri's (2005) phases of word learning. The

literature presents perspectives from within NZ and abroad. Participants in the primary studies ranged from ages 3-17, though the majority of students were 5-7 year olds. In literature from the USA, this age group is generally referred to as *kindergarten* and *first grade*, which in NZ classrooms would be approximately equivalent to students in *New Entrants* to *Year Two*. Search terms included: Reading Recovery, early literacy assessment, predictive literacy assessments, letter knowledge, phonemic awareness, phonological awareness, oral language, treatment resisters and non-responders.

Organisation of review

Firstly, the review will examine conceptual models of reading development. The review then examines which skills are required for success in reading. Following this, there is an examination of assessments that could be used to predict which children will be the 'lowest achieving six year olds' for selection into RR and also which students are not likely to succeed in the traditional RR format. The review will then investigate the ideal time to administer early reading assessments, then the current Reading Recovery selection process will be investigated. Following this, the review discusses what is known about the 13% of students who are referred on from RR (Ministry of Education, 2014) and how the models of reading acquisition help educators understand the needs of the students with reading difficulties. Finally, the review will examine newly emerging research on treatment resisters. By identifying predictive measures that can be used for early literacy assessment in schools, educators can be empowered to identify those most at risk of

reading difficulty and intervene early to prevent the widening achievement gap from occurring.

Models of reading acquisition

Three models for conceptualising reading underpin the research in this thesis – they explain what reading is and how it develops. First, Hoover and Gough’s (1990) model explains what must be present in order for reading to occur successfully. After this, Spear-Swerling and Sternberg’s (1994) road map for reading difficulties provides a useful analogy, illustrating that once a reader comes off the path to reading proficiency, negative consequences begin to impact the student’s reading strategies and the student’s self-esteem. Finally, Ehri’s (2005) phases of word reading development detail what skills and knowledge readers develop as they move through the phases, and importantly for educators, Ehri explains what should be taught to students at each phase. Together, these models explain important understandings for educators of vulnerable readers.

The simple view of reading was introduced by Hoover and Gough (1990). The authors assert that comprehension of text, for information and enjoyment, is the ultimate goal of reading, and even though reading requires many skills and strategies to be orchestrated skilfully, there is a simple way to conceptualise reading. According to the simple view, reading requires two crucial parts: linguistic comprehension and decoding. The authors state, “if reading (R), decoding (D), and linguistic comprehension (L) are each thought of as variables ranging from 0 (nullity) to 1 (perfection), then the simple view of reading can be

expressed as $R = D \times L$ " (p.132). Therefore, to make reading progress, neither part can have a value of zero. This view is important because it emphasises the crucial need for decoding development in readers.

In their road map for understanding reading difficulties, Spear-Swerling and Sternberg (1994) offer a series of six stages on the road to reading proficiency. The authors compare the process of 'reading development' to following a road on which skills must be gained before progressing. Along the way, readers who fail to learn a reading skill fall by the wayside. These readers begin on a different path which is much harder and from which it is very difficult to return to the main path. For example, in the early stages of reading, if a reader fails to learn to decode effectively, the reader will develop an increasing number of negative associated behaviours and attributes; the reader may develop frustration with reading, poor motivation, lower self-concept, lower self-esteem, and poor reading strategies like guessing and solely referring to pictures and context for help. Spear-Swerling and Sternberg explain such students will read less and less, experience fewer opportunities to read and learn reading skills, and encounter thousands fewer printed words than proficient readers. The longer a poor reader remains off the path to proficient reading the harder and less likely it is for them to return. This highlights the importance of early intervention for vulnerable readers.

In discussing the six stages of their model, Spear-Swerling and Sternberg (1994) explain the *Visual Cue* stage is usually observed in pre-schoolers when they readily identify

logographic information in their environment. Visual Cue readers do not understand the alphabetic principal yet. They do not yet understand that writing English involves a code in which printed letters map onto speech sounds in spoken words. By the time students enter school, most will be moving through the *Phonetic Cue Word Recognition* stage. As phonetic-cue readers gradually develop letter-sound links they may only attend to part of a word, such as the beginning or ending sound. At this stage, readers will require phonemic awareness to be in place or they will come 'off the road'. In the *Controlled Word recognition* stage, readers will have acquired some sight words and will make full use of phonetic cues. Some letter patterns are now known on sight but multisyllabic words may still pose a problem. Readers commonly reach this stage in the second year of schooling. By the time readers reach the *Automatic Word Recognition* stage, they will have amassed a large sight word vocabulary, and larger letter-patterns are also recognised on sight. Readers in this stage will only rely on context occasionally to work out unfamiliar words. Readers often reach this stage around the time of their third year of school. By the fifth stage, *Strategic Reading*, readers begin to use comprehension strategies such as summarising. These readers have increased morphemic knowledge which enables them to learn new vocabulary. The final stage is *Proficient Reading* – this typically begins in adolescence. Spear-Swerling and Sternberg explain proficient reading requires the use of higher order comprehension strategies.

If the reader comes off the road at any place along the way, there are negative consequences that may lead to serious reading difficulties unless the reader is brought

back onto the road to reading proficiency. This view of reading acquisition matches the author's experience as a Resource Teacher of Literacy (RT:Lit); students that are referred to RT:Lit due to severe reading delay, appear to have come off the path at one of the early stages (often the phonetic-cue word recognition stage) and therefore cannot progress in their reading development until they are brought back onto the 'road'. Spear-Swerling (2013) reminds educators that "determining the underlying profile and pattern associated with poor reading is an important first step to designing appropriate intervention." (p. 428). This highlights the importance of providing different interventions for students at different stages of reading proficiency.

The third model underpinning the research in this thesis is Ehri's (2005) phases of word reading development . On the way to becoming a proficient reader Ehri proposes there are four phases that all readers move through, which is similar to the six stages of Spear-Swerling and Sternberg's (1994) model. Ehri chooses the term 'phases' rather than 'stages' because the later term implies discrete steps that must be accomplished before moving on to the next step. Instead, this model of reading acquisition acknowledges readers may use skills from more than one phase as their skills develop and this is important for educators to grasp – students can be working at different phases at the same time.

The first phase in Ehri's (2005) model is the pre-alphabetic phase. This phase is based on the meaning of words. Readers at this phase have no letter-sound knowledge. The

'reader' makes visual and contextual connections to 'read' words. For example, a child may see the 'EXIT' sign by a door at school and correctly 'read' the word, yet the same child would be unable to identify the word 'exit' without the familiar colour, font and layout. This phase is most commonly seen in pre-schoolers and readers with severe reading difficulties.

In the second phase readers, begin using partial phonetic cues. As some connections are made between letters and sounds, these readers may attempt reading words using known beginning or end sounds. These readers will lack vowel knowledge so the word 'CAT' may be read as 'COT', 'KIT' or 'CUT' unless there are other contextual cues. These behaviours are likely to be noticed in New Entrant students and older readers with severe reading difficulties.

Readers in the third phase will begin 'cipher reading'. All letter-sound correspondences are known and major grapheme-phoneme units are secure. These students begin to benefit from introducing decoding skills approach rather than the whole-word approach used up to this point. Students in Year One/Two and above are usually working at the full alphabetic phase.

By the time readers arrive at the consolidated alphabetic phase, they will know grapho-phonemic units (morphemes, onsets, and rimes). They will have developed a large bank of automatic high frequency words, including some monosyllabic words that have become

sight words. Within this phase, readers will also develop automatic recognition of frequently seen syllables which are now a consolidated unit, such as 'ing', and this will assist in reading polysyllabic words. At this phase, morphemic connections are made for retaining sight words. For example, '-ed' '-er' and '-est' are some of the first morphemic suffixes to be learnt. At this phase, readers are likely to be Year 3 and above.

The ultimate goal of Ehri's phases is for automatic reading, as is seen in mature readers who read most words automatically, with little effort required to apply strategies to work out unknown words. Similar to Spear-Swerling (2013), Ehri offers specific instructions to help educators correctly target the needs of students at the various phases. For example, when a reader is working at the Pre-alphabetic phase and the reader comes to an unknown word, prompt the reader to use picture and contextual cues together to work out the word, however, as the reader moves to the next phase, prompt for 'sounding out'. It would be inappropriate to prompt pre-alphabetic readers to use letter sounds to identify words because they have not yet made links from letters to sounds so are not capable of following this instruction.

Together the three models of reading acquisition build the foundation of this study. Hoover and Gough (1990) remind educators that strong decoding and strong comprehension are both required for proficient reading, while the models by Ehri (2005) and Spear-Swerling and Sternberg (1994) explain the skills required at each step of reading acquisition. In addition, the road map analogy reminds teachers of the importance

of intervening quickly once a student encounters difficulty, while Ehri offers guidance on providing the appropriate support to readers according to their current phase of reading.

Skills needed for successful reading

In discussing poor readers, Juel (1988) shared the sobering statistic that the “probability that a child would remain a poor reader at the end of fourth grade, if the child was a poor reader at the end of first grade, was .86” (p.440). This highlights the urgency to set every child on the path to reading success. In order to help struggling readers to progress, it is necessary for educators to understand what successful readers do well. Research on literacy acquisition has developed rapidly, and as the knowledge base builds, several key skills are repeatedly identified as being crucial for reading success. In their longitudinal study examining literacy acquisition, Juel, Griffith, and Gough (1986) followed the model of the simple view of reading and identified that good decoders have three things in common: exposure to lots of print (which builds their lexical knowledge), strong phonemic awareness, and secure knowledge of the cipher. The need for strong phonemic awareness was also identified by Castle, Riach, and Nicholson (1994) who found children who started school with low levels of phonemic awareness, were more likely to need RR support if they had not been given phonemic awareness training. Ehri (2005) also promotes the importance of grapheme-phoneme correspondences because of the way the connections bond letters to their pronunciation and meaning. Combining both viewpoints, Ehri and McCormick (1998) found phonemic awareness was a precursor to developing skill in processing grapheme-phoneme relations. More recently, Spear-

Swerling (2013) found the lowest 20-25% of readers will require explicit teaching in phonemic awareness and alphabet sound-symbol relationships in order to become successful readers. So, taken together, the evidence shows successful readers need strong phonemic awareness and strong knowledge of the letter-sound relationships.

Successful phonological processing is another area identified as a requirement for reading success. In their longitudinal study, Chapman et al. (2001) found phonological processing skills (i.e. phonological recoding and phonological awareness) were closely linked to subsequent reading success. Students were assessed on a range of skills including sound matching, phoneme deletion, phoneme segmentation and phonological recoding. Poor readers, including those who participated in RR, were found to have weak phonological processing skills in pre- and post-tests. Phonological processing weakness is also identified as a cause of difficulties acquiring early word reading skills by Liberman, Shankweiler, and Liberman (1989).

Oral language proficiency is often linked with reading success. However, there are conflicting opinions about the impact of oral language on students' responsiveness to reading instruction. Lesaux and Kieffer (2010) identified a vocabulary weakness is common amongst low SES learners and English language learners. Adams (1990) suggested vocabulary may play a role in fluent reading because it is easier to decode familiar words rather than unfamiliar. Catts, Compton, Tomblin, and Bridges (2012) and Scarborough (2005) found language delay from pre-school years tends to link to later

reading difficulty. This matches the findings of the study by Chapman et al. (2001) where an oral language measure was used (Peabody Picture Vocabulary Test (PPVT)). Chapman and colleagues found receptive vocabulary (when tested at the beginning and end of Year One) was linked to later reading outcomes. Several other investigations have found similar results (e.g. Fazio, 1997; Foorman et al., 1997; Torgesen & Davis, 1996).

Interestingly, these studies all worked with diverse populations, yet when Vellutino and colleagues worked with a mostly Caucasian, middle class group of students, they found the PPVT-R results were not related to treatment responsiveness (Vellutino et al., 1996). Further inconclusive data were found when Al Otaiba and Fuchs (2006) examined 23 studies for indicators of unresponsiveness; 15 of the studies reported on vocabulary and verbal ability, and of these, only five reported that students low in these areas were likely to be non-responders. Further research in this area would be useful to confirm whether verbal ability is linked to reading outcomes.

Early literacy assessment

In NZ, current early literacy assessment practices vary from school. The most common assessment battery is the OSELA, which is used to assess many children at age six, but there are few literacy assessments commonly administered in the first year of schooling. While a variety of assessment are included in the OSELA, the lack of specific phonemic awareness components means teachers are not alerted the phonemic awareness needs of students who may be at risk of developing reading difficulties (Greaney & Arrow, 2012). While the *Letter Identification* task in the OSELA accepts identification by letter name or

letter sound, research by Arrow (2012) indicates both types of knowledge (i.e. letter and sound) are highly correlated in the early years but they differentially predict reading outcomes, therefore, these two domains should be assessed separately. Therefore, NZ schools need alternative assessments measures, that can be used with students in their first year of school, and measure the range of skills crucial for reading success.

There are many types of early literacy assessments, but for this study, it was necessary to narrow down the selection of assessments to those measures that were predictive of later reading outcomes. It was identified that letter names (LN) assessments can be accurate predictors of learning to read (Blachman, 1984; Ehri & McCormick, 1998; Foulin, 2005). In addition, phonemic awareness is a prerequisite skills for reading success and can be easily assessed in the early years (Adams, Foorman, Lundberg, & Beeler, 1998; Stahl & Murray, 1994). Another area of assessment that is predictive of reading outcomes is phonological awareness (Blachman, 2000). So it can be seen that there are several areas of knowledge that require assessment in the early years, in order to predict later reading outcomes.

Many authors advocate the use of pairs of assessments tasks to achieve greater accuracy in predicting reading outcomes (Juel et al., 1986; Muter, Hulme, Snowling, & Taylor, 1997; Scarborough, 1998). When Scarborough (1998) investigated letter naming as a predictor, it was found LN by itself is less accurate in identifying at-risk students than results from a combination of tasks. Juel et al. (1986) undertook a longitudinal study into the acquisition of literacy and identified phonemic awareness and cipher knowledge to be predictive of

later reading outcomes. In another longitudinal study, Muter et al. (1997) reported results from their longitudinal study in the UK that investigated early literacy prediction measures for use in the first two years of learning to read. Assessments were administered prior to school entry and phoneme segmentation and letter naming were found to be predictive of reading success at the end of the first year. Therefore, it would appear that more than one assessment should be used to predict reading outcomes.

Rather than using one or two measures of literacy assessment, a multivariate screening process is recommended by a number of researchers. For example, Torgesen (1998) urges educators to use various measures of letter identification (LI) and phonemic awareness (PA) assessments together. This matches the findings of the National Reading Panel (2000) who examined best practice in early literacy education, and found LI and PA were the two indicators that are the best school entry predictors of reading success. Torgesen suggests using a range of phonological awareness tasks that tap into different forms of phonological awareness (e.g. blending and segmenting). Torgesen encourages the use of sound comparison tasks (i.e. matching) with 5-year-olds, and blending and segmenting tasks with 6-year-olds because phoneme segmentation and blending are skills that develop over time so are more appropriate with the older children. A multivariate approach is also endorsed by Clay (2005), who pointed out, *“When important decisions are to be made, we should increase the range of observations we make in order to decrease the risk that we will make errors in our interpretations”* (p. 12). In a recent New Zealand study with multivariate assessment tools, Greaney and Arrow (2012) investigated

the reading-related knowledge of five-year-olds using the eight tasks in the OSELA, the Burt word reading test (Gilmore, Croft, & Reid, 1981), a set of six phonological-based assessments (Adams et al., 1998), a phoneme segmentation task (a modified version of the test developed by Tunmer, Herriman, & Nesdale, 1988), and a pseudo-word reading task (30 words from section three of the decoding skills test by Richardson & DiBenedetto, 1985). Greaney and Arrow selected those particular tasks because they are easily administered and give a full picture of a child's literacy knowledge and skills. So it can be seen that a multivariate approach to literacy screening is desirable.

The ideal time to administer early literacy assessments

The timing of literacy assessments and subsequent literacy interventions is important to get right; the achievement gap begins to widen quickly so the longer schools wait to intervene, the wider the gap becomes. Therefore, literacy assessments should begin as early as possible so needs can be identified quickly. In order to avoid later reading difficulties, schools should put in place early assessment procedures that begin soon after children enter school rather than a wait-to-fail approach (Greaney & Arrow, 2012). The selection model used for Reading Recovery requires that schools wait until students are struggling before they are identified at age six for RR support. However, it is not clear when assessment should be carried out to minimise false positives and false negatives in predicting child outcomes. In a comprehensive review of reading literature, Scarborough (1998) found studies about predicting reading difficulties had substantial levels of false positives and false negatives. If a child is assessed as soon as he or she begins school, the

results may imply great risk of reading difficulty, when in reality the results simply show the child's lack of literacy experience at that point, and the child may make rapid progress when exposed to quality teaching (Torgesen, 1998). Therefore, the longer the child is in school, the greater the accuracy of the assessment. However, Torgesen adds it is still worth assessing early, because the negative consequences of waiting to fail are hard to undo.

In the NZ longitudinal study by Chapman et al. (2001), the timing of early literacy assessments were explored. The assessments were carried out on seven occasions spread throughout Years 1 – 3. In each of the three Year 1 testing occasions, the results of the phonological processing and word recognition measures were significantly lower for the readers who were struggling one and two years later. The assessment of letter knowledge from the beginning of Year One was also predictive, while reading book level and reading comprehension were not predictive until the end of Year 1. This suggests that early literacy assessment could begin from the beginning of Year 1 with phonological processing and word recognition.

Reading Recovery

Reading Recovery (RR) is a preventative early intervention programme designed for young children who are not making expected progress after 12 months of classroom reading instruction (Clay, 1993). This school-based reading intervention provides one-to-one teaching, which is partly funded in New Zealand (NZ) schools by the Ministry of

Education. This programme has been acknowledged internationally as an effective one-to-one intervention for at-risk readers that provides positive effects in reading fluency and general reading achievement (D'Agostino & Murphy, 2004; What Works Clearinghouse, 2007).

An important issue in school-level implementation of RR is how students are selected for the programme. Nationwide, approximately 14% in each year's cohort receive Reading Recovery support (Ministry of Education, 2014). Currently, these students are selected for RR largely based on the results of the Observational Survey of Early Literacy Achievement (OSELA, Clay, 2005). The OSELA consists of a battery of tests and observational procedures assessing letter identification, word recognition, concepts-about-print, writing vocabulary, dictation, and running records of oral reading behaviour. However, these assessments lack a measure of phonological awareness, so crucial data about the student's metalinguistic knowledge is not being collected (Center, Wheldall, Freeman, Outhred, & McNaught, 1995; Iversen & Tunmer, 1993).

Once students have been assessed by the OSELA, the outcomes are used to guide student selection for RR. As part of the roles and responsibilities of schools who offer Reading Recovery, schools are instructed to, 'select the lowest achievers...regardless of potential, absences, [or] ethnicity' (Ministry of Education, 2012). When describing the RR selection process, Serry, Rose, and Liamputtong (2014) state, "Reading Recovery adopts a non-exclusion philosophy such that all children who were tentatively selected following

screening on Clay's Observation Survey are eligible for the intervention," (P. 66).

However, anecdotal evidence in several studies indicates that this is a variable process and some schools are selecting students according to other criteria. In another study, Belgrave (2009) discussed student selection with a group of Reading Recovery teachers. Four of the five teachers felt that children with very low OSELA scores were unlikely to succeed in Reading Recovery and as a result the teachers manipulated which students they took onto the programme. These teachers indicated that they felt it was a waste of resources taking on the very lowest students because they are likely to be referred on from Reading Recovery and that it is better to take on students who will get the greatest benefit from the programme. In addition, the teachers felt that the children with very low OSELA scores were not ready to receive RR, and that they needed more experience with phonemic awareness and oral language activities before RR would be useful for them.

More recently, non-adherence to the student selection process was also raised as a concern in the evaluation of the four-year i3 scale-up of Reading Recovery in America (May, Sirinides, Gray, & Goldsworthy, 2016). In the evaluation survey data, it was found up to 22% of schools did not take the lowest achieving students into Reading Recovery. Reasons for excluding the lowest achieving students included special education status and poor attendance. This variability in student selection will obviously have an impact on the literacy achievement of those students who struggle the most.

Reading Recovery teaching is based on a strong top-down approach to fluent reading (Clay, 1991). Greater emphasis is placed on teaching word recognition skills in context and minimal word-level information is used to confirm predictions. However, Ehri, Nunes, Stahl, and Willows (2001) suggest an approach which includes phonics instruction is likely to lead to greater improvements in reading comprehension, decoding and sight-word reading than a whole-word or whole-language approach. While many authors debate the efficacy of RR (e.g. Iversen & Tunmer, 1993; Tunmer & Chapman, 2003) this review does not have the scope to cover that topic here.

As students are discontinued from RR, they are assigned one of five outcomes:

- *“Successfully discontinued* —the student is able to work effectively with their cohort without additional support.
- *Carried over*— the student is unable to complete their series of lessons in the current year and has had their lesson series continued into the following year.
- *Referred on*—the student has not reached expected level and further specialist or long-term literacy support is required.
- *Left the school* — the student left the school before completing their series of lessons (and may or may not have continued at their new school).
- *Unable to continue*—the student left the intervention before completing their support for various reasons.” (Ministry of Education, 2014, p.15)

Decisions about discontinuing a child from RR are based on several factors: a) the child's reading level, b) the child's level of independence when reading, and c) time spent in the programme. After approximately 12 -20 weeks of Reading Recovery support, a child is 'successfully discontinued' if they have reached the average reading level of their peers (i.e. a level where the child can learn effectively in the classroom without further support). The 2013 annual data collected from RR in NZ shows the majority (79%) of students who exited RR were successfully discontinued from the intervention. These students were all reading texts at or above level 13 (Green 2) and most of these students (82%) were reading texts between level 17 (Turquoise 1) and level 20 (Purple 2) by the time they exited from Reading Recovery (Ministry of Education, 2014) This follows the trends of RR 'successfully discontinued' outcomes in recent years. However, 13% of RR students did not make the expected progress and were referred on for specialist reading support (Ministry of Education, 2014).

In New Zealand, once students are 'referred on', some receive ongoing support from a specialist Resource Teacher of Literacy (RT:Lit). However, the RT:Lit service (also funded by the Ministry of Education) has very limited resources, so, for a large proportion of referred students, there is no further access to specialised literacy support. This means the students who exit Reading Recovery with the greatest need often have no further access to expert literacy support in New Zealand. Therefore, it is crucial to ensure those students

most at risk of reading difficulties are offered the correct intervention at the right time so NZ early literacy funding and support is targeted accurately.

The 13% - what is known about students who are referred on?

Nationally, 76% of the 60,272 six-year-olds had access to Reading Recovery (Ministry of Education, 2014). Within these RR schools, 8137 six-year-olds entered Reading Recovery, and 8160 exited the programme by the end of the year. While 79% of RR students who exited the programme in 2013 were successfully discontinued, 13% were referred on for further specialist support. That means approximately one in eight students who exited RR and did not benefit sufficiently from the programme and needed ongoing reading support. It has to be asked if more can be done for this vulnerable group of learners.

Looking for trends in the outcomes for Reading Recovery students, McDowall, Boyd, Hodgen, and Vliet (2005) analysed the 2003 New Zealand Reading Recovery Annual Monitoring data (Anand & Bennie, 2005). They also interviewed many teachers involved in the implementation of RR within New Zealand. It is interesting to note; when interviewed, the RR tutors and trainers did not consider it possible to predict students' rates of progress by entry scores. However, when the Reading Recovery data from student assessments prior to entry (Observation Survey of Early Literacy Achievement, Clay, 1993) was analysed the authors found there were some noticeable trends in the statistics. The most likely students to be referred were those who entered RR with Instructional Reading Level scores below 5, Burt Word Reading Test (Gilmore et al., 1981)

scores below 8, and Clay Writing Vocabulary scores below 12. In addition, Maori, Pasifika and Asian students were more likely to be referred than NZ European.

Interesting trends for referred students can also be seen in the more recent analysis of the 2013 RR annual monitoring data (Ministry of Education, 2014). The proportion of Maori (14.1%) and Pasifika (13.3%) students referred on for specialist literacy support remains higher than other ethnic groups. In addition, boys (14.7%) are more likely to be referred than girls (9.8%). Of the referred students, it is known that 66.8% were referred to RT:Lit for specialist literacy support, though unfortunately many of these students would not have received RT:Lit support due to the high demand for the limited number of spaces in the RT:Lit service.

A common finding in research is that children who do not respond in Reading Recovery have a weakness in phonological awareness (Center et al., 1995; Snow, Burns, & Griffin, 1998; Tunmer & Chapman, 2003). For example, in an evaluation of Reading Recovery, Center et al. (1995) analysed pre-test and post-test data in metalinguistic areas such as phonemic awareness and phonological recoding and found those with poor meta-linguistic skills were more likely to be referred.

The time students spent in the RR programme varied widely, however, on average, successfully discontinued students attended 79 lessons over 19 weeks, while referred students attended 90 lessons over 23 weeks (Ministry of Education, 2014). 'Time in

Reading Recovery' statistics for referred boys and girls were very similar. Ninety-nine percent of referred students entered RR reading at text level 8 or below. Burt Word scores of most referred students were 10 or less (below the lowest age band of 5.10-6.04) and these students exited RR with scores averaging 17 (still below the lowest equivalent age band of 5.10-6.04). This shows a persistent difficulty with word reading for the referred students. Furthermore, the statistics show these students also had difficulty with the Writing Vocabulary Task; scoring an average of stanine 2 on entry and only progressed to stanine 3 on exit (see Clay, 2013, for further explanation of scoring and stanines for this task). A third of referred students had already completed two years of schooling by the time they exited RR; this means these students were already seven years old and they were now a long way behind their peers.

When early literacy assessments are administered in the first year of school, the results can be compared with subsequent RR outcomes to identify correlations between the initial assessment scores and those who are successfully discontinued and those who require further support. When Chapman et al. (2001) carried out a longitudinal study of beginning literacy, they assessed a range of measures with their five-year-old students to identify links from this early knowledge to the students' participation in Reading Recovery, and subsequent outcomes. The tasks covered phonological sensitivity, word reading, spelling, self-concepts, and reading. Chapman and colleagues found the RR students' levels of phonological processing, word recognition, letter ID, and receptive vocabulary were all linked to later reading outcomes. Furthermore, when the Chapman et al. assessed

the students as they exited Reading Recovery, it was found the students' difficulties in the phonological processing and letter sound knowledge remained. Therefore, the more advanced the phonological processing skills on entry to the programme, the greater the likelihood of successful outcomes. To measure phonological processing skills, Chapman and colleagues used six tasks; phoneme deletion, phoneme segmentation, onset-rime sensitivity, pseudo-word decoding, analogical transfer and invented spelling. In addition, the Letter ID task from the Diagnostic Survey (Clay, 1985) was used to assess letter knowledge, receptive vocabulary was measured by the PPVT, and word recognition was measured using the Ready to Read Word Test (Clay, 1985).

Treatment resisters

In every literacy intervention study, there are students who do not respond to generally effective reading interventions. These students have been termed 'non-responders' or 'treatment resisters' (e.g. Blachman, 1994; Torgesen, 2000). As many as 30% of students with reading difficulties, and 50% of children with special needs, may not benefit from early literacy interventions – these are the so-called 'treatment resisters' (Al Otaiba, 2003). The issue of treatment resisters is a relatively new topic, yet there is a growing body of research emerging.

The first concern is to define the non-responders; Torgesen (2000) defines this group as children who have access to preventative programmes, but who still struggle with word reading skills, and score below the 30th percentile on Word Identification and Word Attack

tasks. An alternative method of identifying those most at risk of reading difficulty has been to select an arbitrary cut-off point across the school, e.g. the lowest 25% of readers.

According to Torgesen, this normative approach has an inherent flaw; there will always be students achieving in the 'lowest quartile' no matter what their level of attainment. This means some children in the lowest 30% may actually be achieving adequate levels of literacy skills and may not require intervention. It is important to note that the current method of selection for Reading Recovery students, is to select the 'lowest 20% of readers' from each school.

Vellutino et al. (1996) argues that many students with apparent reading disabilities are actually students who needed more effective early literacy instruction. This supports the findings of Al Otaiba and Fuchs (2006) who found when high quality classroom literacy instruction was offered over a sustained period of time, only 7.05% of students were nonresponsive (compared with 25.35% of the control students). Similar to this, Torgesen (2000) reviewed studies designed to improve the early reading skills of students with reading difficulties and found, when students were provided with effective early literacy interventions, there would always be a small but persistent group of 2% - 6% of first and second grade students who were treatment resisters. In a recent study by Carson, Gillon, and Boustead (2013), it was found a ten-week PA intervention, that was delivered by the classroom teacher alongside the regular reading programme, was highly effective.

Following the intervention, Carson et al. found only 6% the children who received the PA intervention had difficulties with decoding, while, in the classrooms with literacy

programmes that did not target PA, 26% of students had difficulties with decoding. Thus, it can be seen that an intensive PA intervention, administered in the classroom when students are 5-years-old, can have a strong influence on their literacy outcomes. It is interesting to note, the findings of Carson et al. about the percentages of unresponsive students are similar to the findings in the research by Al Otaiba and Fuchs (2006).

Across the research on treatment resisters, there lacks consistency in the data reporting; some studies work with the general student population, while others use a group of students who are already working at the lowest levels. Furthermore, some studies present statistics based on the percentage of the total student population who are treatment resisters, others report on a group of students who have participated in an intervention, and others do not specify which group is being reported on. This makes it very difficult to use the data meaningfully in generalisations to the wider population.

When examining what is known about the literacy profile of treatment resisters, Al Otaiba (2003) found, across a range of studies, a majority of treatment resisters had phonological awareness deficits (see also Al Otaiba & Fuchs, 2006). This is important when considered in terms of the finding of McKenna and VioLato (2003) that programmes that target phonological deficits for remediation tend to show better reading outcomes for the participants (see also Torgesen, Wagner, & Rashotte, 1997). Gough and Juel (1991) assert, “what seems essential is to ensure that children learn to decode in first grade. If decoding skill does not arrive then, it may be very hard to change the direction that

reading achievement takes” (p.50). Therefore, in the context of this present study examining Reading Recovery outcomes, it has to be considered whether the Reading Recovery programme is targeting phonological deficits thoroughly enough to ensure the greatest gains for treatment resistant students.

The number of treatment resisters reduces as the length and complexity of literacy interventions increases (Blachman, 1994). This has important implications for the duration of interventions offered in NZ schools; RR is currently offered for 12-20 weeks, and in some cases, is extended a few extra weeks if the child is near to the expected level for discontinuing. The duration of this intervention is long compared to many others, and RR is a daily intervention so the intensity is high. However, as Blachman does not specify the duration of the interventions being compared in her research, it is difficult to say whether this is long enough to be effective for treatment resisters.

Having acknowledged treatment resisters exist, it is important to look for ways to identify these students early, in order for intervention support to be targeted appropriately. Al Otaiba and Fuchs (2006) reviewed 23 studies involving young students with literacy difficulties and found there were several characteristics associated with unresponsiveness: (a) phonological awareness; (b) verbal memory; (c) rapid naming, (d) vocabulary, verbal ability and IQ; (e) attention or behaviour problems; and (f) home background, including socioeconomic status. In a larger review of literacy intervention responsiveness, Nelson, Benner, and Gonzalez (2003) found noteworthy predictors of reading outcomes, for

students in the early years of primary school, were rapid naming, behaviour, phonological awareness, and alphabetic understanding. In examining the data further, Al Otaiba and Fuchs (2006) found a combination of characteristics was usually present in unresponsive students; naming speed and vocabulary were both important predictive factors in anticipating which students would be nonresponsive. Furthermore, when these scores were combined with scores for sentence imitation, problem behaviour, and amount of intervention time, they correctly predicted 82.1% of nonresponsive students. However, while this combination of characteristics was linked to non-responders, this level of prediction is still not accurate enough because there would be several false positives and negatives among those identified.

Thus it can be seen that there are many unresolved issues regarding treatment responsiveness such as which early characteristics lead to accurate identification of these students, which combination of skills is most predictive, and what is the ideal intensity and duration for interventions? There is still more research to be done before we know how to teach each student in a way that will bring successful reading outcomes for all treatment resisters.

Summary

This review covered the skills needed for reading success, predictive assessments to identify at-risk students, the ideal time to administer early literacy screening, the current RR selection process, RR student outcomes, and treatment resisters. Firstly, the review

examined conceptual models of reading development that underpinned this study. The simple view of reading explained that comprehension and decoding must both be strong in order for reading to be successful. The road map showed that certain skills must be gained before progress can be made in reading. Furthermore, readers who fail to learn a reading skill begin on a different path from which it is very difficult to return. While off the path, students develop coping strategies which hinder their progress, and develop frustration with reading, poor motivation, and lower self-esteem. The third model, Ehri's (2005) phases of word learning, explained which skills are required at each phase, and which sequence to teach them in order for skilled reading to occur. These three models explain the key factors for children at risk of reading difficulties: that it is important to intervene early, with the right strategies, and to ensure decoding and comprehension are equally strong.

The second area investigated was the current Reading Recovery selection process. It was identified that the students selected for Reading Recovery are those with the lowest scores in the OSLEA in each school (i.e. not students with the lowest scores across the country). This system may mean students with the greatest need do not always get selected, and the national statistics may not accurately represent the needs of the population. Torgesen (2000) suggests a selection process with a minimum attainment criteria across the population should be adopted to ensure equity. A further issue was highlighted with RR; the current selection process does not occur until students have reached the age of six, delaying intervention support for these vulnerable learners. As

Torgesen (1998) says, “The best solution to the problem of reading failure is to allocate resources for early identification and prevention” (p.1).

Following this, the review discussed what is known about the 13% of students who are referred on from RR (Ministry of Education, 2014). The most likely students to be referred on from RR were those who entered RR with an instructional text level 8 or below, BURT word reading scores 10 or less, and Clay Writing Vocabulary scores below 12. In addition, the proportion of Maori and Pasifika referred on for specialist literacy support remains higher than other ethnic groups. In addition, boys are more likely to be referred than girls. A common finding was that children who do not respond in Reading Recovery have a weakness in phonological awareness (Center et al., 1995; Snow et al., 1998; Tunmer & Chapman, 2003) and the more advanced the phonological processing skills on entry to the programme, the greater the likelihood of successful outcomes (Chapman et al. 2001).

The review then examined which skills are required for success in reading. It was identified that good readers have strong phonemic awareness and secure knowledge of the cipher. In addition, phonological processing skill was closely linked to subsequent reading success. Receptive oral language was also found to be predictive in some studies. However, future research will be required to identify whether oral language is a significant indicator of reading outcomes.

Following this, there was an examination of predictive assessment tasks and the ideal timing for the screening to take place. A multivariate screening model was recommended to achieve the greatest predictive reliability. The most common predictive assessments included (a) letter-naming, (b) letter sounds, (c) phonemic segmentation (from age 6) and sound comparison tasks (with 5-year-olds), (d) phoneme deletion, (e) phoneme blending, (f) rapid naming, and (g) receptive vocabulary. Regarding the timing of these assessments, all authors agreed screening for literacy difficulties should begin early. The first or second term of a child's schooling were the predominant recommendations.

Finally, the review examined newly emerging research on treatment resisters which indicated many students (approximately 30% of the general population) may not benefit from early literacy interventions. However, many of these students may have progressed if they had been provided with more effective literacy instruction. Al Otaiba and Fuchs (2006) found, when high quality classroom literacy instruction was offered over a sustained period of time, only 7.05% of students were nonresponsive, while Torgesen (2000) identified that, despite the most high quality interventions being offered, there was always a small but persistent group of 2% - 6% of first and second grade students who were treatment resisters. Furthermore, the majority of these treatment resisters had phonological awareness deficits. To reduce the number of treatment resisters, it is known that as the length, intensity and complexity of literacy intervention increases, the number of non-responders reduces (Blachman, 1994) .

In summary, the literature on the topic of early literacy acquisition was useful in clarifying the skills needed for successful reading outcomes, identifying which successful early literacy assessments are most predictive, and the ideal timing of those assessments. However, further research is required using NZ data, to identify the most predictive combination of screening tasks so we can ensure funding for early literacy interventions is used in a way that brings about the best outcomes for students. By identifying predictive measures that can be used for early literacy assessment in schools, educators may be empowered to identify those most at risk of reading difficulty and intervene early to prevent the widening achievement gap from occurring.

The present study

Using assessment data from a larger, longitudinal research project, this thesis investigates whether there are any correlations between children's literacy skills at school entry (aged 5), and their subsequent selection for Reading Recovery. Of the students selected for Reading Recovery, this study will investigate any correlations between the students' early literacy assessments and their subsequent Reading Recovery outcomes. Can the data tell us whether students will be selected for RR, and whether these students will be successful in the RR intervention?

This study will investigate two questions:

1. Which early literacy assessments predict selection for Reading Recovery?
2. Which early literacy assessments predict the outcomes in Reading Recovery?

Based on the research findings in the Literature review, there are two hypotheses. The first hypothesis is that children who start school with low phonological awareness (PA), low alphabet knowledge and low oral language, are most at risk of reading difficulty, and therefore are most likely to be selected for Reading Recovery. The second hypothesis is that, of the students who are selected for Reading Recovery, the higher their letter-naming and PA scores are, the more likely they are to be successfully discontinued, while those students with the lowest PA and lowest alphabet knowledge are most likely to be referred on for further support.

Methodology

The purpose of this thesis was to examine the relationship between students' literacy skills at school entry and their subsequent Reading Recovery selection and outcomes. The first aim of the thesis was to identify any school entry assessment tasks that were predictive of future selection for Reading Recovery. The second aim was to identify any assessments tasks that were predictive of the two predominant RR outcomes (successfully discontinued or referred on). This quantitative research project was undertaken as secondary data analysis using unpublished data from a larger longitudinal study that was in progress (Early Literacy Project, Chapman et al., 2016). This chapter discusses the methods and materials used in this thesis. First, the participants are described, then an explanation of the research design is given. This chapter ends with an outline of the data collection and assessment tasks.

Participants

The students were all participants in the Early Literacy Project (Chapman et al., 2016). From the ELP data-set, the data for 300 students from 38 primary schools, were used in this thesis. All participating schools were based in the lower half of the North Island of New Zealand. The cohort of students whose data were used in this study were all aged five at the time the first data were collected in February 2015 (T1), and had a mean age of 6 years, 9 months by the time the final data were gathered in November 2016 (T4). All

students had little to no formal reading instruction at the point of initial testing and few had any reading ability (mean score from Clay's (2005) Word Reading subtest = 0.62 words).

Of the 38 schools that participated in the longitudinal study, 31 offered the Reading Recovery intervention within the school. Of the participating children, 224 attended schools that offered Reading Recovery. Of these students, 34 participated in Reading Recovery. For the purposes of this thesis, the groups of participants were identified as follows: (1) the *Main Group* (MG, N=300), (2) the subset of *Reading Recovery Schools* (RRS, n=224), and (3) the subset of students who *Participated in Reading Recovery* (PRR, n=34). The third group was further divided according to RR outcomes, and results were analysed for the following two groups: students who were *Successfully Discontinued* from RR (SD, n=11), and those who were *Referred on from RR* (RO, n=6).

The MG students (see Table 1) were predominantly NZ European, Maori or from the Pacific Islands (Pasifika), and there were approximately 10% more boys than girls. The MG schools represented the full range of decile ratings 1 – 10. These deciles are the SES ratings determined by New Zealand census data, based on variables such as household income and occupations of parents. The deciles are used by the Ministry of Education to allocate funding; 1 is the lowest rating and receives the most funding support. Of the schools that offered Reading Recovery in this study, none were decile 10. From the students that participated in RR, (a) there were significantly more boys than girls, (b)

Maori and Pasifika students featured prominently (making up 53.3% of the total RR students), and (c) the majority of students (97.1%) were from mid-to-lower decile schools.

Table 1 - Profile of ELP Participants in 2015

| | | All Students | Students in schools that offer RR | Students that participated in RR |
|-------------|-------------|-------------------------|--|---|
| | | N=300 | n=224 | n=34 |
| Gender | Male | 55.3% | 52.7% | 64.7% |
| | Female | 44.7% | 47.3% | 35.3% |
| Ethnicity | NZ European | 61.0% | 63.0% | 36.7% |
| | Maori | 26.4% | 26.0% | 40.0% |
| | Pasifika | 6.5% | 5.3% | 13.3% |
| | Asian | 3.2% | 2.4% | 3.3% |
| | Other | 2.9% | 3.3% | 6.7% |
| Decile Band | Low 1-3 | 30.7% | 24.1% | 38.2% |
| | Mid 4-6 | 49.0% | 55.4% | 58.9% |
| | High 7-10 | 20.3% | 20.5% | 2.9% |

Design

This thesis was undertaken as secondary data analysis using an unpublished data set from the Early Literacy Project (ELP). Secondary data analysis of the ELP data was a cost-effective way to access longitudinal data from a large group of emergent literacy learners. The ELP was a randomised intervention-control design, but the data reported in this thesis does not distinguish between the intervention and control groups as there were no differences in those groups after one year of ELP. Another advantage of using secondary data analysis was that separate Ethics Committee approval was not required for this study as ethics approval was granted for the primary study. The names of all participants were deleted so data was only identifiable by project student ID, which ensured the anonymity of all participants. An additional benefit of using the ELP data is that the information was gathered from New Zealand schools, which makes this data uniquely relevant for New Zealand classroom teachers.

The ELP was a longitudinal study following learners' literacy progress for three years from entry to school in February 2015. The data used in this thesis was taken from ELP data collected in 2015 and 2016, which covered collection times 1 – 5. The data were gathered by trained data-collectors who went into schools to assess the children.

- Time 1 (T1) – February 2015 (initial assessment data gathered)
- Time 2 (T2) – June/July 2015 (second set of assessment data gathered)
- Time 3 (T3) – Nov 2015 (reading levels collected)

- Time 4 (T4) - June 2016 (Reading Recovery selection data collected from schools)
- Time 5 (T5) – November 2016 (Outcomes for RR students and reading levels collected)

Data collection

As this study was based on longitudinal data from the Early Literacy Project, the first year of data had been collected before this study began; therefore, the selection of assessment tasks was outside the control of this thesis. From the ELP battery of assessments, the tasks used in this study were selected because they measured aspects of early literacy which had been shown to be predictive of later reading outcomes. In the study by Chapman et al. (2001) difficulties in phonological awareness and letter-sound knowledge at age five were both linked with Reading Recovery outcomes, and Castle et al. (1994) found that children with low levels of phonemic awareness are more likely to need RR support. Therefore, assessments of phonological awareness and letter sounds were included in this study.

Furthermore, Ehri and McCormick (1998) promoted letter-name recognition as one of the best predictors in learning to read. As Arrow (2012) explained, it is important to assess both letter names and letter sounds, because they are differentially predictive of outcomes, therefore, a letter naming assessment was included in the data set. Finally, a measure of oral language was selected for this study to investigate whether verbal ability was linked to reading outcomes. While some researchers debate the influence of oral

language on reading outcomes, several researchers found language delays from early years tended to link to later reading difficulty (Catts et al., 2012; Chapman et al., 2001; Scarborough, 2005) so oral language warranted investigation in this study.

All participants were assessed on a variety of measures over five data collection points. At *Time 1* the students had just entered New Entrants/Year One and by *Time 5* the students had almost completed Year Two. The data set included Reading Recovery selection and outcome data as collected by each school. In addition, reading levels were collected at Times 3 and 5 by classroom teachers as part of their ongoing monitoring. Below is a description of all assessment measures used in this study. For all assessment tasks, reliability is reported if available (scores close to 1 show high reliability):

Measures

Letter Identification

This task was used to find out which alphabetic symbols were recognised by the student. Students were asked to name upper and lower case letters, when presented with them in random order. One point was scored for each correct answer. The upper-case letter-name scores were out of 26, and the lower-case scores were out of 28 (due to an additional formation of the letters 'a' and 'g'); a total score of 54 letter-names. This task was modified from an original task by Clay (2005) where letter name knowledge and letter sound knowledge were assessed as part of the same knowledge in the Observation Survey. As children are usually

expected to know all letter names by the end of Year 1 (Ministry of Education, 2010), ceiling effects mean that letter name knowledge was only assessed in the first year of schooling (Time 1 and Time 2).

Knowledge of the alphabetic principle

Letter-sound (LS) knowledge is used to measure knowledge of the alphabetic principle. Letter-sound knowledge draws on the understanding that letters represent sounds. The Literacy Learning Progressions (Ministry of Education, 2010) identified the need for letter-sound knowledge to be in place by the end of Year 1. In this task, modified from the *Letter Identification* task by Clay (2005), students were instructed to provide the sound for each of the 26 upper-case and 28 lower-case letters (due to an additional formation of the letters 'a' and 'g'); a total of 54 letter-names). Some letters represent several sounds (e.g. 'A' can represent the sounds /ā/, /ǎ/, /ǔ/), therefore these variations were accepted as well. This task also has a ceiling effect so was only administered at Time 1 and Time 2.

Three phonological awareness tasks were selected from the Comprehensive Test of Phonological Processing, 2nd Edition (CTOPP 2) (Wagner, Torgesen, Rashotte, & Pearson, 2013) and were administered at Time 1 and Time 2:

CTOPP Elision

A 34-item test which measures a student's ability to say a word and then say what remains of the word when designated sounds are omitted. This task can be administered to students aged 4-24 years. The test-retest reliability coefficient for Elision is .82.

CTOPP Blending

This 33-item test was used to measure a student's ability to blend given sounds together to form words. The sounds were provided via audio recording to ensure consistency of delivery. This task can be administered with individuals aged 4 – 24 years. The reliability is reported as .75.

CTOPP Sound Matching

This task consisted of 26 items. The student was required to match sounds at the start or end of a word. To administer this assessment, a Picture Book was required. The examiner said a word and pointed to a picture of that item. Then three further words were said, and those images were also indicated. The student was asked to point to the word which began (or ended) with the same sound as the first word. This assessment gave 13 initial-sound tasks and 13 final-sound matching tasks. The assessment is suitable for use with students aged 4-6 years only. The reliability coefficient for Sound Matching is .78.

Phonological Awareness Total Score

In order to create a single measure to represent a participant's awareness of the phonological structure of oral language, the scores for the three phonological awareness tasks *Elision*, *Blending* and *Matching*, were added together. The maximum scores for each of those tasks were: Elision = 34, Blending = 33, Matching = 26; therefore, the combined phonological awareness (PA) score had a possible total of 93.

Oral Language

The British Picture Vocabulary Scale (BPVS III) (Dunn, Dunn, & Styles, 2009) was used as the oral language measure in this study. The BPVS is a one-to-one, picture-based test that measures a student's receptive (heard) vocabulary. For every item in the test, the assessor said a word and the student was asked point to the picture that best illustrated that word. The BPVS can be used with students aged 3 to 16 years of age, though only items of the appropriate age range are administered. According to the examiner manual, the median reliability is 0.90 which shows it to be a very reliable test. The BPVS was administered at Time 1 but not at Time 2 because it the BPVS manual suggests it is poor practice to repeat the assessment within 6 months (Dunn et al., 2009).

Each variable was analysed for links with RR selection and outcomes. The variables were also analysed for correlations between pairs of assessments, and for combinations of multiple assessments. In order to identify which combination of assessment measures could indicate the greatest difference between the SD and RO groups, the following variables were also calculated:

- A. The sum of all T1 scores (LN, LS, Mat, Eli, Ble, & BPVS)
- B. The sum of all T1 scores with no BPVS (because oral language was not found to be predictive in some studies)
- C. The sum of all T1 scores with no BPVS or LS (because so many students scored 0 for LS at T1 that it may be confounding the data)
- D. The sum of T1 LN & Matching only (because these two variables showed the most separation of scores between the two groups)
- E. The sum of all T2 scores (LN, LS, Mat, Eli, Ble)
- F. The sum of all T2 scores without Matching (i.e. LN, LS, Eli, Ble)
- G. The sum of all T2 scores without Elision (i.e. LN, LS, Ble, Mat)

Combinations of scores taken at different time points were not considered because results from a single time point would be most useful in a realistic classroom assessment situation.

Results

This thesis investigated the ability of early literacy assessments to predict later reading outcomes. First, a description of the analysis is given. Then the chapter gives the results for the first question: Which early literacy assessments predicted student selection for Reading Recovery for students at RR schools? Following this, the exit data of RR students is analysed to identify the results of which of the early assessments were predictive of being successful or unsuccessful in RR. The analysis of questions one and two includes comparisons of pairs of variables by scatterplot and Pearson correlation tables, and is followed by binomial logistic regression. The results chapter ends with an examination of which combinations of literacy assessments predicted the reading levels of students after two years at school to confirm whether the measures were reliable with this data set.

Analysis

This thesis analysed ELP data (Chapman et al., 2016) in order to identify early literacy assessments which were predictive of Reading Recovery selection and outcomes. IBM SPSS data analysis software was used for all analyses. See the Appendix for a table of means and standard deviations for all variables used in this thesis. First, the initial assessment data (Time 1) for 'students at schools that offer Reading Recovery' (RRS) was compared to the data for 'students selected for Reading Recovery' (PRR). The data were analysed for assessments that linked school entry skills with the likelihood of selection for Reading Recovery. Using the sample of students from schools that offered RR (n=224),

the Time 1 assessment scores were analysed by a correlation matrix and binary logistic regression techniques to identify any tasks, or combinations of tasks, which were predictive of Reading Recovery selection. This was repeated with the assessment scores for the Time 2 variables to check whether the later assessments (i.e. assessments after six months at school) were more predictive than those taken at school entry (Time 1).

T1 Assessment Variables:

Letter Names (LN), Letter Sounds (LS), Elision (Eli), Blending (Ble), Matching (Mat), Phonological Awareness combined score (PA), and the British Picture Vocabulary Scale (BPVS).

T2 Assessment Variables:

Letter Names (LN), Letter Sounds (LS), Elision (Eli), Blending (Ble), Matching (Mat), and Phonological Awareness combined score (PA).

For Question 1 (*Which early literacy assessments predict selection for Reading Recovery?*), the analyses of all measures were based on n = 224 (students at schools that offer RR). Because the outcome variable is dichotomous (students are either selected for Reading Recovery or they're not), a binomial logistic regression was run. This model predicts the probability of an outcome occurring (i.e. the likelihood of being selected for Reading Recovery). Logistic regression can work with one or more independent variables; in this study, there are seven possible variables at Time 1 that could be included (LN, LS, Eli, Ble,

Mat, PA and BPVS) and six at Time 2 (LN, LS, Eli, Ble, Mat and PA). The sample size of students selected for RR is small and this limits the number of variables that can be run in a logistic regression. Sample size calculations for logistic regression are complex, but Peduzzi, Concato, Kemper, Holford, and Feinstein (1996) give guidelines for a minimum number of cases to include in a study; if p is the smallest of the proportions of negative or positive cases and k is the number of independent variables, then the minimum number of cases to include is $n=10k/p$. In this study, the proportion of children in RR compared to those not in RR can be shown as $p=34/224=0.15$. There are seven possible independent variables that could be run in the logistic regression (LN, LS, Eli, Ble, Mat, PA and BPVS) and Table 2 shows the minimum number of cases required to run the logistical regression in this study.

Table 2 - Minimum Number of Cases Required to Run Logistic Regression

| Number of variables | Number of cases required to run the regression |
|---------------------|--|
| 1 | 67 |
| 2 | 133 |
| 3 | 200 |
| 4 | 267 |

As there were 224 cases in this study, it can be seen in Table 2 that a maximum of three variables could be run at a time if the advice of Peduzzi et al. (1996) is followed.

However, there are four key variables that need to be considered: LN, LS, PA and BPVS (Eli, Ble, and Mat are all subsets of PA). Fortunately, Vittinghoff and McCulloch (2006) advocate for relaxing the rule of ten events per variable (EPV) in logistic regression. Their

findings showed that running regression analysis with only 5-9 EPV has a similar accuracy to running the analysis with 10 EVP. Therefore, logistic regression was a valid option for analysing the data in this thesis, and was run using the four key variables of LN, LS, PA and BPVS at T1, and LN, LS and PA at T2.

Once the PRR group exited RR, their assessment data for all T1 and T2 variables was analysed to find any school entry assessments that were predictive of being successfully discontinued or referred on (i.e. Successful or Unsuccessful RR students). Exit outcomes from Reading Recovery were only reported for 17 of the cases, so analysis for questions two and three were based on n=17. Due to the small sample size, it was not possible to run a regression analysis to answer this question, so dot plots were used to compare the scores for both groups (SD and RO) for each variable that was assessed. Furthermore, combinations of assessment scores were calculated as new variables to investigate whether multivariate screening was more predictive than individual variables. The analysis of the combination scores are displayed in dot plots to highlight any areas of difference between the Successful and Unsuccessful students.

Finally, T1 and T2 assessment scores for all participants at RR schools were compared with their T5 reading levels (n=137 students who attended a RR school and had a T5 reading level reported). The findings of the literature review indicated early assessments in letter naming, phonemic awareness, phonological awareness and receptive vocabulary should be able to predict outcomes after two years at school (see Chapman et al., 2001) so this

additional step in the analysis was valuable. It was important to establish whether these T1 and T2 assessment measures were able to predict reading outcomes to confirm the accuracy of the data and the assessment measures.

Predicting selection for Reading Recovery

Firstly, in order to identify which early literacy assessments predicted selection for Reading Recovery, analysis was carried out on the assessment data for Time 1 (February 2015 when students first entered school) and Time 2 (June/July 2015 after approximately 6 months of schooling). The student scores for each assessment task were compared with the scores for the other tasks assessed at that time, to ascertain whether there were strong relationships between any pairs of variables.

A Pearson correlation coefficient was computed (see Table 3) to assess the relationships between pairs of T1 assessments. There was a positive correlation between all pairs of variables. Letter Names and Letter Sounds showed a very strong, positive relationship at T1 which is commensurate with the findings of Arrow (2012) that there is a high correlation between letter name and letter sound knowledge in the early years. In contrast, BPVS had a weak relationship with all variables. Torgesen (1998) suggests Matching is an appropriate skill for measurement at age five while Blending should be used from age six onward and this is reflected in the T1 data; Matching showed stronger relationships with other variables than was shown by Blending.

Table 3 - Pearson Correlation Matrix of Time 1 Assessment Tasks

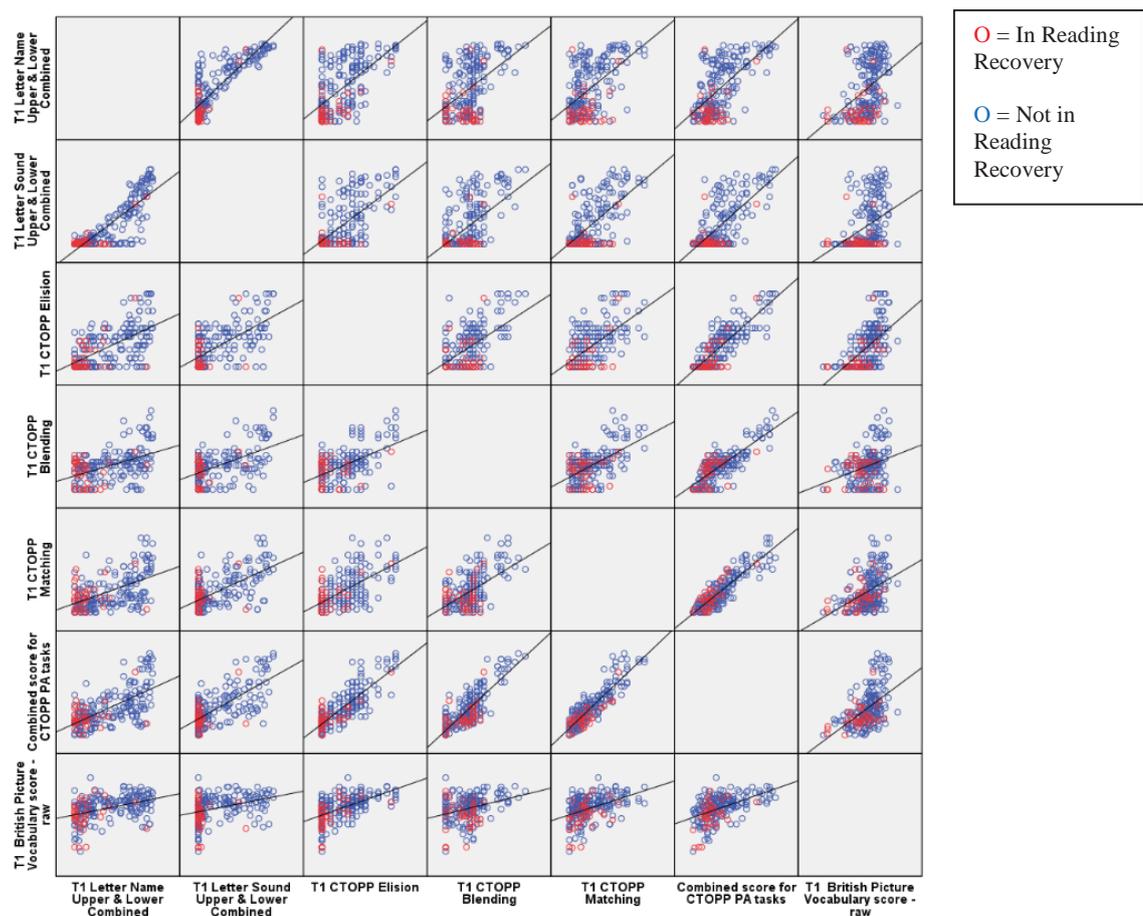
| | 1 | 2 | 3 | 4 | 5 |
|--|----------|----------|----------|----------|----------|
| 1. Letter Names | | | | | |
| 2. Letter Sounds | .849 | | | | |
| 3. Elision | .616 | .645 | | | |
| 4. Blending | .469 | .520 | .523 | | |
| 5. Matching | .539 | .629 | .614 | .553 | |
| 6. British Picture Vocabulary Scale | .415 | .363 | .556 | .312 | .438 |

Note. all correlations are significant at $p < .01$ level (2-tailed).

To identify whether the T1 assessments identified which students were selected for Reading Recovery, the correlations were shown in a scatterplot matrix as shown in Figure 1. For every variable in the matrix, the results for the Reading Recovery students (shown in red) overlap that of the students not in Reading Recovery (shown in blue). For example, when scores for Elision and Matching are compared, a widely-dispersed cloud of scores can be seen, and the red RR scores are scattered throughout. The cloud shape (rather than a tight line) shows that the scores of those two variables are not closely correlated. In addition, if RR selection could be predicted by that pair of assessments, the scatterplot would show a clear separation of red and blue within the data, but the RR scores are scattered throughout. Furthermore, there appear to be Reading Recovery outliers in most

assessments (as shown by the red scores that are far away from the main group of red scores). The lack of distinct groups shows that these pairs of variables are not useful for predicting Reading Recovery selection.

Figure 1 - Correlation Between Pairs of Time 1 Assessment Tasks



In Table 4, the Pearson correlation coefficient was computed to see if the T2 assessments were able to predict RR selection more accurately than T1. As can be seen in the table, there was a positive relationship between all pairs of variables at T2. Again, Letter Names

and Letter Sounds showed a strong positive relationship. At T2 the participants were approximately five- and-a-half-years-old and the scores for Matching showed stronger relationships to other variables than the scores for Blending, as seen at T1.

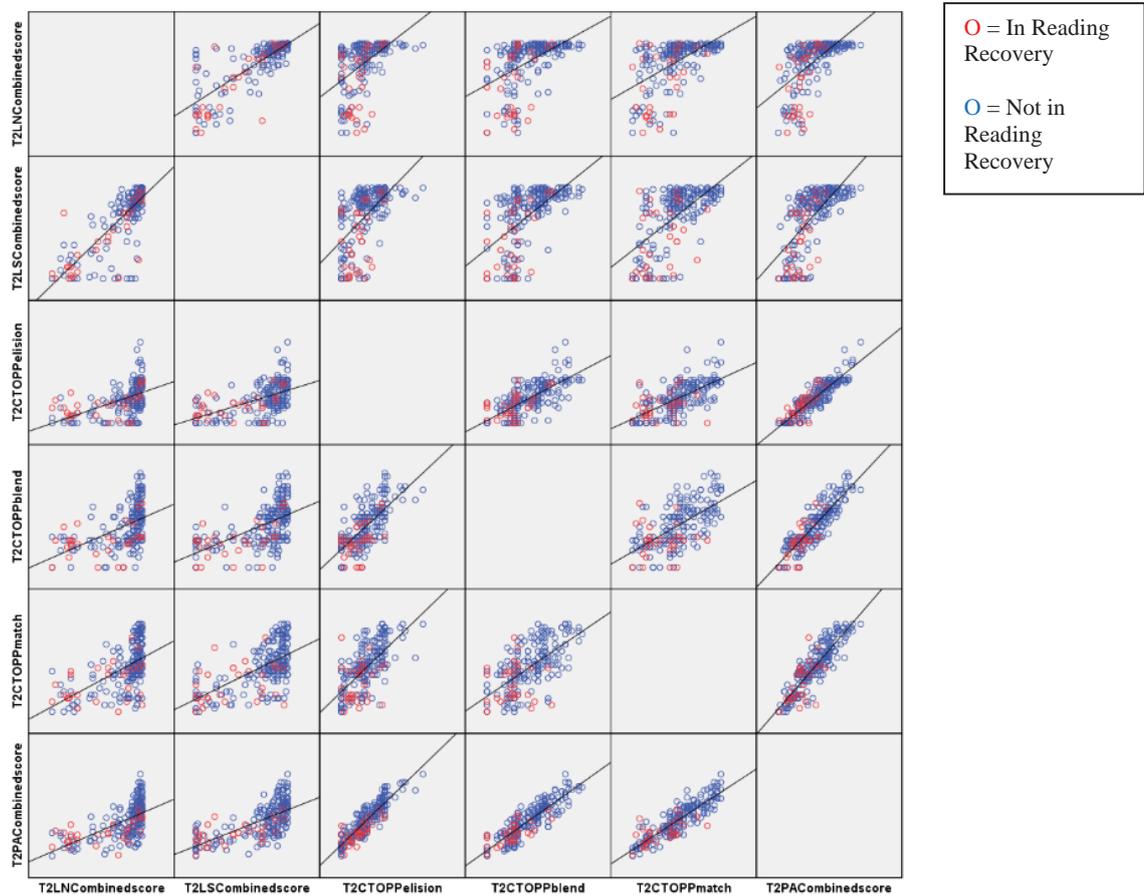
Table 4 - Pearson Correlation Matrix of Time 2 Assessment Tasks

| | 1 | 2 | 3 | 4 |
|-------------------------|----------|----------|----------|----------|
| 1. Letter Names | | | | |
| 2. Letter Sounds | 0.796 | | | |
| 3. Elision | 0.518 | 0.568 | | |
| 4. Blending | 0.498 | 0.586 | 0.707 | |
| 5. Matching | 0.565 | 0.629 | 0.669 | 0.633 |

Note. all correlations are significant at $p < .01$ level (2-tailed).

A scatterplot matrix summarises the results and compares the PRR group to the those not in RR (see Figure 2). When the T2 assessments were compared in the matrix, a similar pattern was seen to that shown at T1; for every variable, there was significant overlap between those who were selected for Reading Recovery and those who were not. In addition, outliers could be found in every task. This lack of separation of the two groups illustrates the point that both groups are similar in these pairs of assessments at T2 and that RR selection cannot be predicted by these pairs of assessments.

Figure 2 - Correlation Between Pairs of Time 2 Assessment Tasks



Having ascertained that pairs of assessments were not able to identify who would be selected for RR, further analysis was required to investigate whether a combination of multiple assessments would be more predictive. A binomial logistic regression was performed to ascertain the effects of T1 scores for letter name knowledge, letter sound knowledge, phonological awareness, and vocabulary (LN, LS, PA and BPV) on the likelihood of being selected for Reading Recovery (see Table 5). Linearity of the continuous variables with respect to the logit of the dependent variable was assessed via the Box-Tidwell procedure (Box & Tidwell, 1962, as cited in Laerd Statistics, 2015). A

Bonferroni correction was applied using all four terms in the model (and the constant) resulting in statistical significance being accepted when $p < .01$ (Tabachnick & Fidell, 2007). Based on this assessment, all continuous independent variables were found to be linearly related to the logit of the dependent variable, which shows the data is suitable for use with linear regression.

Table 5 - Logistic Regression Predicting Likelihood of Reading Recovery Selection Based on Knowledge of Letter Names, Letter Sounds, Phonological Awareness and Vocabulary at Time 1

| | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I. for EXP(B) | |
|---------------------------|-------|------|-------|----|------|--------|------------------------|-------|
| | | | | | | | Lower | Upper |
| Letter Names | -.065 | .028 | 5.208 | 1 | .022 | .937 | .886 | .991 |
| Letter Sounds | -.024 | .044 | .292 | 1 | .589 | .977 | .897 | 1.064 |
| Phonological Awareness | .024 | .031 | .613 | 1 | .434 | 1.024 | .965 | 1.087 |
| Vocabulary | -.019 | .013 | 2.235 | 1 | .135 | .981 | .957 | 1.006 |
| Constant | .398 | .721 | .304 | 1 | .581 | 1.489 | | |

In this regression, there were 6 observations having studentised (Z) residuals with a value of 2.5 standard deviations or greater (i.e. outliers), which were still all kept in the analysis due to the small sample size. The logistic regression model was statistically significant, $\chi^2(4) = 34.783$, $p < .0005$, though it explained only 26.0% (Nagelkerke R^2) of the variance in

Reading Recovery selection; this tells us other factors outside these aspects of literacy explain most of RR selection. Using the model, 82.4% of cases were correctly classified, however sensitivity was only 2.9% which tells us the model is only able to predict a very small percentage of students who will be selected for Reading Recovery. Of the four predictor variables used in this model, only one was statistically significant: Letter Name knowledge (as shown in Table 5). This is likely due to the fact that Letter Name knowledge is the only assessment from this battery that is similar to one used in the OSELA to select RR students. Finally, the regression analysis showed that increasing scores in letter names, letter sounds or vocabulary marginally decreased the likelihood of selection for Reading Recovery.

A binomial logistic regression was repeated using the T2 scores for letter names, letter sounds, and phonological awareness (LN, LS and PA – there was no measure of vocabulary used at T2) as shown in Table 6. Linearity of the variables was confirmed via the Box-Tidwell procedure; with the Bonferroni correction applied, all three terms in the model had $p < .0125$. There were 11 cases with studentised residuals with a value of 2 standard deviations or greater which were all kept in the analysis. The logistic regression model was statistically significant, $X^2(3) = 24.358, p < .0005$. The model explained 20% (Nagelkerke R^2) of the variance in Reading Recovery selection and correctly classified 86.7% of cases. Sensitivity was 3.6%, specificity was 99.5%, positive predictive value was 50%, and negative predictive value was 46.53%. Of the three predictor variables, none were statistically significant. Increasing scores in letter names, letters sounds or

phonological awareness by one unit marginally decreased the likelihood of selection for Reading Recovery.

Table 6 - Logistic Regression Predicting Likelihood of Reading Recovery Selection Based on Knowledge of Letter Names, Letter Sounds and Phonological Awareness at Time 2

| | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I. for EXP(B) | |
|---------------------------|-------|------|-------|----|------|--------|------------------------|-------|
| | | | | | | | Lower | Upper |
| Letter Names | -.029 | .020 | 2.214 | 1 | .137 | .971 | .934 | 1.009 |
| Letter Sounds | -.015 | .019 | .637 | 1 | .425 | .985 | .948 | 1.023 |
| Phonological Awareness | -.023 | .020 | 1.382 | 1 | .240 | .977 | .939 | 1.016 |
| Constant | .444 | .522 | .726 | 1 | .394 | 1.560 | | |

Overall the analysis of Question One showed none of the T1 and T2 variables were able to accurately predict who will be selected or Reading Recovery. Letter Names was the only significant variable linked with Reading Recovery selection but it was only able to account for a small part of the selection process. This lack of predictability may be due to the wide variation in selection process used by each school. In order to explore the variability of the selection process, a comparison of reading levels from the time students began entering RR was undertaken. The ELP data did not collect the reading levels at entry to

RR, therefore the T3 reading levels of the students was used as an alternative indicator of the level each student was reading at when they were selected for RR (see Figure 3).

Figure 3 - Reading Levels at Time 3 (T3) When Students Began Entering Reading Recovery (RR)

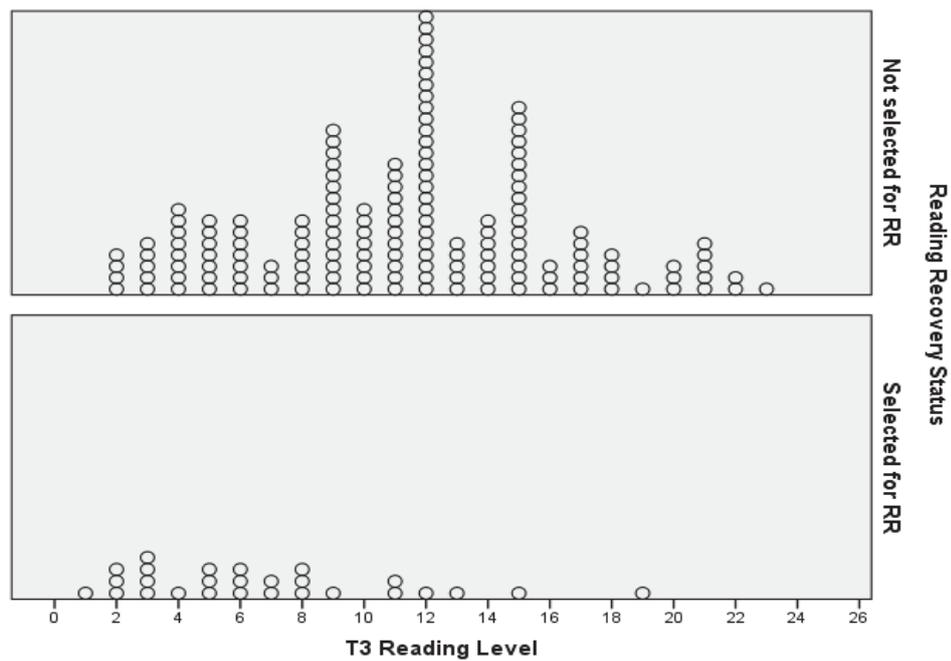


Figure 3 illustrates two points: Firstly, row 2 shows, when students are selected for Reading Recovery, schools are selecting students with a wide range of reading levels. While most students entered RR reading at levels 1 – 11 (i.e. working below the National Standard for a child at the end of one year at school), other students appear to have entered RR reading at levels 12 and above (i.e. already attaining the National Standard). While this area requires further exploration with accurate RR entry data, it certainly highlights the disparity between selected students. The second point illustrated in Figure

3 is that, at RR schools, there are many children who are reading at the lower end (i.e. Level 11 and lower) yet only a small portion received RR support. This also highlights the inequality of the current RR provision; the lowest readers are not necessarily the ones receiving the support. It is unknown whether this is due to schools not following the RR selection guidance, or whether other factors are impacting the selection process, and indicates this would be a valuable area for future exploration.

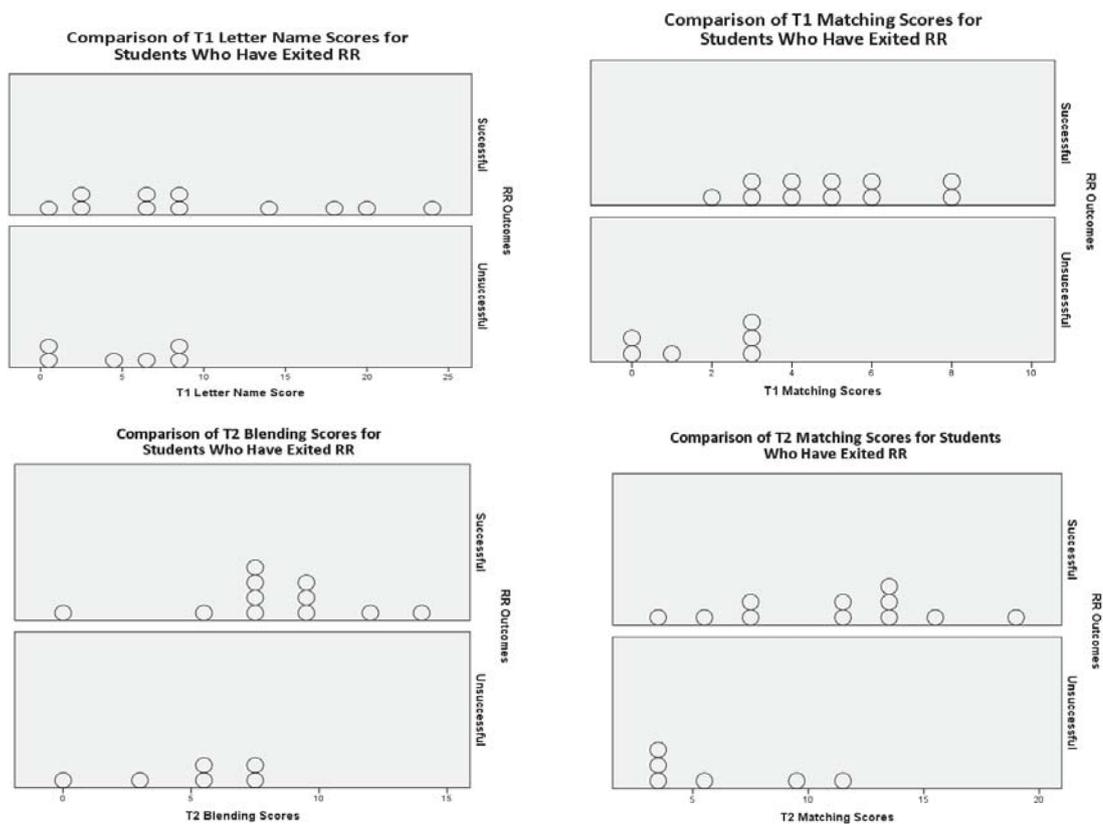
Predicting Reading Recovery outcomes

Having answered Question One, the analysis then aimed to identify which early literacy assessments predict success in Reading Recovery and which early literacy assessments predict the students who will be referred on from Reading Recovery for further support.

In order to identify which assessments were predictive of RR outcomes, data were collected once the students exited RR and compared to the assessment scores at T1 and T2. Schools reported exit data for 17 students; of this group, 11 students were successfully discontinued (SD) and six students were referred on (RO) for further support. The T1 and T2 assessment data of the 17 students were analysed to find any school entry assessments that linked to the Reading Recovery outcomes. First, all variables assessed at T1 and T2 were analysed by simple dot plots comparing the scores for the SD and RO students. If a variable was highly predictive, it was expected the results for each group would be distributed with clear separation between each group. However, if a variable did not link to RR outcomes, it is likely there would be substantial overlapping of both

groups. For example, if both groups had T1 Letter Name scores ranging from 0 to 30, then it would be unlikely that T1 Letter Names played a significant role in the RR outcomes for those students because they all began with similar knowledge in that aspect.

Figure 4 - Comparison of Scores for the Successful and Unsuccessful RR students for Time 1 (T1) and Time 2 (T2) Variables That Showed the Greatest Difference Between the Groups



In many variables, both the SD and RO groups had a similar range of scores so those assessments were not useful for indicating RR outcomes. However, Figure 4 shows four variables that had some distinction between the two groups: T1 Letter Names, T1 Matching, T2 Blending and T2 Matching. As can be seen in Figure 4, the variable showing

the greatest separation between groups was T1 Matching. This suggests that Reading Recovery outcomes could be indicated by Matching scores taken at T1; the RO students all had a score of three or less. This is also important because it implies that effective Matching (an early phonological awareness skill) is necessary for reading success.

In order to check whether a combination of scores was more predictive than the score of single variables, it was necessary to do further analysis. Due to the small sample size, it was not possible to run regression analysis to identify which combination of scores contributed most to the outcomes. Therefore, it was necessary to find other ways to compare the data from the two groups of RR students. Any combination of variables that could show separation in the scores of the successful and unsuccessful RR students would be an indicator that the combination was predictive of RR outcomes. In order to identify which assessment measures highlighted the greatest difference between the SD and RO students, the following combinations of variables were analysed:

- A. The sum of all T1 scores (LN, LS, Mat, Eli, Ble, & BPVS)
- B. The sum of all T1 scores with no BPVS
- C. The sum of all T1 scores with no BPVS or LS
- D. The sum of T1 LN & Matching only
- E. The sum of all T2 scores (LN, LS, Mat, Eli, Ble)
- F. The sum of all T2 scores without Matching
- G. The sum of all T2 scores without Elision

The combined scores were analysed by range, to identify which combined score resulted in the greatest differences in the range of the SD and RO groups. Of the seven combination scores that were analysed, (B) *T1 Combined Score with no BPVS* showed the greatest separation between the SD and RO students. At this point, it was identified that the unsuccessful RR students all have (B) scores that are 20 points or lower at T1 (for the sum of LN, LS, Elision, Matching & Blending). Furthermore, the scores of the RO students only partially overlaps the scores of the SD students (see Table 7). This finding shows an important indicator; students who score 20 or less on the combined score (B) at T1 appear to be less likely to succeed in RR.

Table 7 - Summary of Time 1 (T1) Combined Scores for Students Who Have Exited Reading Recovery

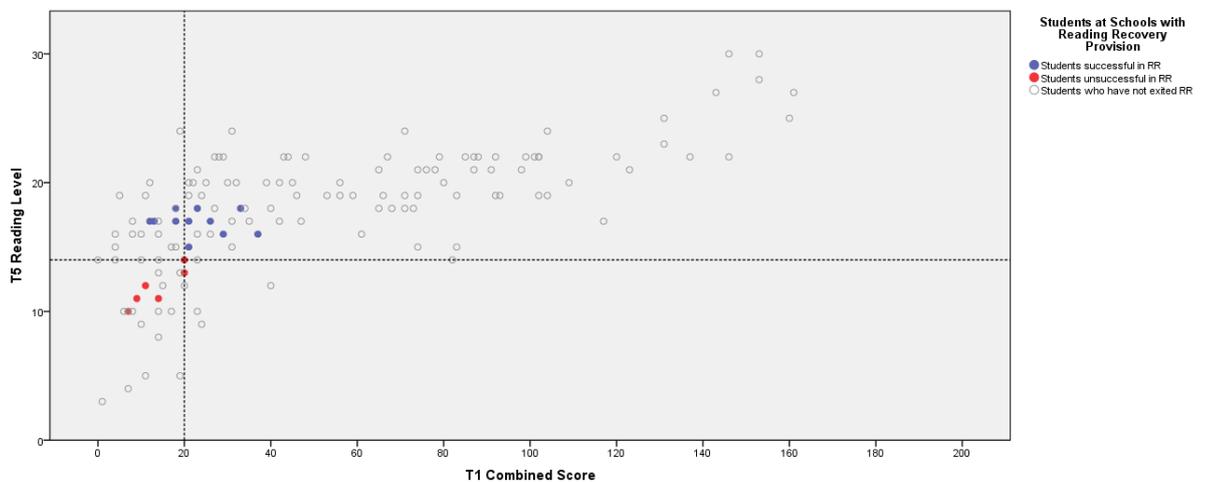
| | T1 Combined Score (B) | |
|-----------|-----------------------|----------|
| | At or Below 20 | Above 20 |
| SD (n=11) | 4 | 7 |
| RO (n=6) | 6 | 0 |

Note. The T1 Combined Score is comprised of the sum of the scores for Elision, Blending and Matching, From Letter Names (LN), Letter Sounds (LS).

In order to establish whether this T1 Combined Score was able to indicate who would have difficulty in the wider population, the T1 Combined score was compared to T5 Reading Levels for the whole group of students at schools that had provision for RR (see Figure 5). In NZ, the National Standard for Reading after two years at school is Turquoise

(Levels 17-18), and after one year at school students must aim to be reading at Green (Levels 12-14). When the reading levels were collected at T5, the students were 6 $\frac{3}{4}$ - 7 years old which means they have had approximately two years at school. Therefore, children who are reading Level 14 or below at T5 are reading at least one year below the National Standard. When the RR scores are highlighted in a scatter plot of the T5 reading levels (see Figure 5), we see both the SD and the RO groups have low *T1 Combined* scores (RR range = 7-37) while much of the cohort had a wide spread of scores (range = 0-161). Furthermore, it can be clearly seen that the RO students are all represented at the lower left quadrant of the scores.

Figure 5 - Time 1 (T1) Combined Assessment Scores Compared with Time 5 (T5) Reading Levels



In Figure 5, the vertical line indicates a score of 20 points at T1, and the horizontal line indicates the student is reading at level 14 at T5. These results from the ELP data suggest that students can be assessed at age five, using a combination of LN, LS, Elision, Blending

and Matching tasks, and that the *T1 Combined* score may indicate who will have reading difficulty after two years at school. Table 8 shows the percentage of students accurately accounted for using the Level 14 cut-off and the 20-point maximum (B) combined score.

Table 8 - Students Predicted as Having Reading Difficulty by the End of Year 2 Using the Time 1 Combined Score

| n=137 | | T1 Combined Score <= 20 | T1 Combined Score >= 21 | |
|------------------------------|--|---|-------------------------|-------------------------|
| Reading Level 15 or above | n=18 13.14% False Positives Type I Error | n=88 64.23% True Negatives | | Specificity = 83.02% |
| Reading Level 14 or below | n=25 18.25% True Positives | n=6 4.38% False Negatives Type II Error | | Sensitivity = 80.65% |
| | Positive predictive value = 58.14% | Negative predictive value = 93.62% | | |

Sensitivity refers to the T1 Combined Score's ability to correctly identify students who will have reading difficulty by the end of Year 2 (i.e. those reading at level 14 or below). From Table 8, it can be seen that the Combined Score at T1 has a fairly high sensitivity because it is able to predict 80.65% of the participants with reading difficulty. The result implies that of all the students reading at level 14 or below by the end of the second year at school, 80.65% could be predicted by the combined set of LN, LS and PA assessments

taken at age 5. In addition, the specificity is also fairly high (83.02) which tells us few participants were incorrectly identified as being at risk of reading difficulty. It is important to note, the group of false positives in this data set may be slightly inflated because it includes 4 students who received RR support and these students would likely have been working below the National Standard had they not received such an intensive intervention. Furthermore, if this group of false positive students were provided with early reading interventions along with the true positives, is it likely the false positives would progress quickly and it would soon become clear these students would not require ongoing support. Therefore, combination score (B) appears to be a potentially useful measure of later reading outcomes, and an indicator of where early literacy support could be allocated.

Summary

This chapter investigated the ability of early literacy assessments to predict later reading outcomes. First, the analysis explored which early literacy assessments predicted student selection for Reading Recovery for students attending RR schools. Assessments were analysed for correlations with RR selection; at T1, Letter Naming was found to be the most significant variable, and at T2, no variables were found to be significantly linked with selection.

Following this, the exit data of RR students was analysed to identify which of the early assessments were predictive of being successful or unsuccessful in RR. Of the 17 students

with reported RR outcomes, T1 Matching was the test showing the greatest difference in performance between the successful and unsuccessful students. This finding suggests the students' scores for Matching, when taken at T1, will indicate each student's likelihood for success in RR. This implies a minimum level of phonological awareness is required for success in the RR intervention. When combinations of assessments were analysed to find predictive links with RR outcomes, T1Combined Score (no BPVS) was shown to identify the greatest level of separation between the Successful and Unsuccessful students.

The analysis ended with an examination of which literacy assessments predicted the reading levels of students after two years at school to confirm whether the measures are useful for predicting RR outcomes only or for predicting reading outcomes for all students. When the T1 Combined Score was compared to the T5 reading Levels, it was identified that a T1 score of 20 points or less indicated students were likely to have reading difficulties by the end of Year 2.

Due to the small sample size, the reported findings can only be taken as indications of what may be possible to predict with the wider population. Further research will need to be carried out with a larger sample in order to confirm whether any of these findings are accurate with the wider population.

Discussion

This study aimed to identify which early literacy assessments were predictive of student selection for Reading Recovery and which assessments were predictive of Reading Recovery outcomes. This chapter begins with a discussion of the results, then the limitations of the study are discussed. Following this, there is an exploration of the implications for future research and the implications for the assessment and teaching of literacy for New Zealand students. The chapter ends with the conclusions from this study.

This study had two hypotheses. It was first hypothesised that children who started school with low phonological awareness (PA), low alphabet knowledge, and low oral language, would be most at risk of reading difficulty, and therefore were most likely to be selected for Reading Recovery. The second hypothesis was that, of the students who were selected for Reading Recovery, the higher their letter-naming and PA scores were, the more likely they were to be successfully discontinued, and therefore, those students who entered RR with the lowest PA and lowest alphabet knowledge were most likely to be referred on for further support. The first hypothesis was not supported, while the second hypothesis was moderately supported by the findings.

Predicting Reading Recovery selection

The first hypothesis of this study was that children who started school with low attainment in PA, letter names (LN), letter sounds (LS), and low oral language (measured

by BPVS) were most at risk of reading difficulty, so were most likely to be selected for Reading Recovery. The findings showed, only LN assessed at T1 was significantly linked to subsequent RR selection (see Table 3), which aligns with previous findings that letter-name recognition is one of the best predictors in learning to read (Ehri & McCormick, 1998; Foulin, 2005). Furthermore, it was likely that LN would be linked with RR selection because LN is very similar to the Letter Identification (LI) task which is part of the assessment battery used for RR selection. However, LN could only explain a small amount of the variance in RR selection. This suggests there are factors outside the assessed variables that explain most of RR selection. This was an unexpected finding. The hypothesis assumed that students who were selected for RR would also be the lowest readers in the cohort; this assumption was not supported (as indicated by the wide range of reading levels in Figure 3). Therefore, the first hypothesis not supported.

As discussed in the Methodology chapter, the variables (LN, LS, Eli, Ble, Mat, and BPVS) were chosen for this study because they were associated with predicting reading outcomes and identifying students at risk of reading difficulty. As Serry et al. (2014) explained, RR has a non-exclusion policy, so all children with the lowest scores in the screening tasks should be eligible for support. However, in this study many of the students selected for RR were not those with the lowest reading levels in the cohort (see Figure 3), which indicates RR selection was not representative of those most at risk of reading difficulty. This is commensurate with the findings of the American RR study by May et al. (2016), where it was reported that up to 22% of schools did not take the lowest

achieving students into Reading Recovery (see also Belgrave, 2009). Reasons for excluding the lowest achieving students included special education status and poor attendance. It can be seen that school variations in RR selection decisions mean some of the lowest readers miss out on support.

A further reason the RR selection could not be predicted, was the reading level of students who entered RR varied considerably between schools. As shown in Figure 3, the teacher-reported *T3 Reading Level* data suggests that some of the students entered RR already reading at a level close to the National Standard. The literature review tells us, in 2013, the range of reading levels on entry to Reading Recovery was Level 1 – Level 15 (Ministry of Education, 2014) and this the same as the range of reported T3 reading levels for the students selected for RR¹. Therefore, it appears the range of T3 reading levels in this study, are congruent with national RR entry levels. This wide variation in entry reading-levels highlights the inequitable provision of literacy support for New Zealand’s most vulnerable readers. While some schools have many students reading at low levels, other schools have very few, therefore perhaps the funding for RR should not be spread evenly through all schools; it should be allocated to the students with the greatest need.

¹ It is noted that the data set also showed one student reading at Level 19 on entry to RR; this is likely to be an error of data entry or teacher reporting - a student reading at that level would already be working at the National Standard and therefore would not require Reading Recovery support.

Predicting Reading Recovery outcomes

The second hypothesis was that, of the students that were selected for RR, the higher their letter-naming and PA awareness scores were, the more likely they were to be successfully discontinued. This hypothesis was moderately supported by the findings of this study but the finding was not as strong as was expected. The scores for the successful students and the unsuccessful students overlap on every variable that was assessed at T1 and T2, yet four variables showed larger differences in the scores between the two groups (as shown in Figure 4) which indicates there is some correlation between these four variables and RR outcomes. However, the small sample size makes it difficult to measure the significance of the correlation between the variables and the RR outcomes.

At T1, scores for *Letter Names* and *Matching* (a sub-set of phonological awareness) were the strongest indicators of RR outcomes. At T2, *Blending* and *Matching* (both sub-sets of Phonological awareness) were the strongest indicators. However, at T1 and T2, scores for the successful and unsuccessful groups showed considerable overlapping so these variables are only partially able to indicate who will be successfully discontinued. Of the 17 students with reported RR outcomes, T1 *Matching* was the test showing the greatest difference in performance between the successful and unsuccessful students. The data showed that the students with T1 *Matching* scores of 0 or 1 were the most likely to be unsuccessful in RR. This finding suggests students with weaknesses in phonological awareness at T1 are likely to have difficulty in the RR intervention. The finding implies a minimum level of phonological awareness is required for success in the RR intervention,

which is commensurate with the findings of Snow et al. (1998) and Tunmer and Chapman (2003), that children who do not respond to Reading Recovery have a weakness in phonological awareness. This finding also makes sense in the context of the Hoover and Gough (1990) model where reading success is the product of strong comprehension and strong decoding. In the model, secure phonological awareness is required for strong decoding, yet the group of readers who are struggling after two years at school had low phonological awareness scores on entry to school. Further evidence that phonological awareness is required for proficient reading is seen in Spear-Swerling and Sternberg's (1994) model; when students are working at the *Phonetic Cue Word Recognition* stage, they must have phonemic awareness in place in order to move to the next stage. If these students do not develop adequate phonological skills, they are less likely to become proficient readers and may require explicit teaching in this area in order to progress.

Reading outcomes for the cohort

The focus of this study was on predicting RR selection and outcomes, and the skills hypothesised as predicting RR selection were PA, LS, LN and Oral language. In this study, the oral language measure was not found to contribute to the prediction of reading outcomes. However, as shown in Figure 5, the combination (B) (of PA, LN and LS assessed at T1) was very predictive of reading book levels for the cohort. The accuracy of this assessment combination is commensurate with the findings of Scarborough (1998) that a combination of tasks is more accurate in identifying at-risk students. The combined score (B) was able to indicate future reading outcomes for the cohort of students at RR schools

(n=224). Those children with a combined (B) score of 20 or less were at greater risk of not succeeding in RR than children who entered RR with a combined score of 21 or more. This suggests, a combined score (made up of PA, LN and LS assessments), that is administered when children enter school, may predict reading difficulties after two years at school, and therefore would allow for early identification of those most at risk. This would warrant further research because early identification of at-risk readers could help schools target early literacy support from the first term of child's schooling.

Limitations of this study

As with any research, there are limitations which affected the success of this study and the validity of the results. These include limitations of data collection, availability of data, skewed data, selection of assessment tools, and the criteria for selection of RR students. Bias in data can skew the results and conclusions. In this study, any potential bias from those running the Early Literacy Project was minimised because the data used in this thesis is the unpublished data before it had been analysed or shaped by the project leaders.

Collecting data from such a diverse range of schools and classrooms was challenging due to student transience, staff changes, illness, absence, and school timetabling. In this study, student transience has led to attrition from the original 354 students in the cohort, to the current number of 300 that was used as the *Main Group* in the data set, so this limited the data set available for analysis. In addition, the small sample size of students

entering RR (n=34) and exiting RR (N=17) could not generate statistically significant results and limited both the analysis and the interpretations that can be made. Therefore, the findings related to RR outcomes can only be taken as indications of what may be true for the wider population and further research with a larger sample will be required to confirm these findings.

During this study, there were also limitations due to the availability of data. For example, due to staggered entry dates to Reading Recovery, the exit dates were later than initially expected, which delayed the timing of the analysis. In addition, some data sets were not complete, but no explanation was available at the time of printing. For example, when the cohort of children from RR schools (n=224) is compared with students from this cohort who also had T5 reading levels reported (n=171), we see there is a large group with unreported data. This may have skewed the data and therefore affected the findings.

Skewed data can also affect the reliability of research findings. When the T5 reading levels for students at RR schools were analysed, it is important to acknowledge that the cohort contained students who received intensive RR support, therefore that group of low readers have already raised their attainment, leaving fewer students reading at low levels, which skews the data. However, to remove the RR students from the data set would also skew the data because that would remove too many of the low achieving readers from the sample. Therefore, in this instance it was prudent to leave the RR data in the set. The data analysis was also affected by the number of outliers in the data. Figures 1 and 2

show outliers in every assessment task analysed in the scatterplot matrix. The large quantity of outliers makes it difficult for a regression model to accurately predict outcomes.

Another limitation was the selection of assessment tools. This study had no influence on the selection of assessment tools used in the Early Literacy Project but fortunately there was a wide range of assessment data gathered which included measures that were relevant for this study. The assessments used in this study (LN, LS, Elision, Blending, Matching and BPVS) were selected from the ELP data set because the literature review identified them as useful measures for predicting reading outcomes. The literature also identified that low RAN scores are frequently linked with predicting poor reading outcomes, though there is also some controversy over the accuracy of RAN data for predicting outcomes when used with school-aged children. This is an aspect that may warrant further investigation.

The final limitation was unexpected; the amount of variation in Reading Recovery selection criteria used by schools. Reading Recovery aims to support the needs of the children making the slowest progress in literacy learning (Ministry of Education, 2014). However, there are many students reading at level 14 or below, which is more than a year below the National Standard for *After two years at school*. It is clear that there are many students working at a low level who were not selected for Reading Recovery. Therefore,

this lack of consistently applied criteria for RR selection limited the predictability of the data.

Conclusions and Implications

The aim of this research was to identify which early literacy assessments could predict student selection for Reading Recovery and which assessments could predict Reading Recovery outcomes. This study adds to the body of knowledge of early literacy in New Zealand primary schools. A strength of this study is that the data includes measurements of phonological awareness, letter sound knowledge and oral language that are not often used in the school setting. Furthermore, the assessment data were collected over multiple time points, enabling the analysis to indicate the ideal timing of early literacy assessments. Understanding variables such as useful literacy assessment tools and the timing of assessments may help with the effective administration of early literacy assessment and efficient early literacy support. This in turn may contribute to the advancement of reading achievement in the early years of school.

The first hypothesis of this study is that children who started school with low attainment in phonological awareness (PA), letter names (LN), letter sounds (LS), and oral language (measured by BPVS) are most likely to be selected for Reading Recovery. While PA, LN LS and oral language are identified in international research as predictors of reading outcomes, this study showed only LN is significant in the prediction of selection for RR. Furthermore, LN can only explain a small amount of the variance which means factors

outside the assessed variables contribute to RR selection. In the course of the study it was identified that some schools were not selecting their lowest readers for RR. Variations in RR selection processes mean it is not possible to predict who will be selected for RR. This is an unexpected finding; it was assumed that the students with the lowest reading attainment would be selected for RR, but this assumption was proven wrong, so the first hypothesis was incorrect.

The second hypothesis is that RR outcomes can be predicted using a combination of LN and PA scores; the higher the scores the more likely the students were to be successfully discontinued (SD), and the lower the scores, the more likely they were to be referred on (RO) for further support. To test this hypothesis, the scores for the SD and RO groups were compared for all 11 variables. While both groups had a similar range of scores for many variables, four of the variables showed a lower range of scores for the students who were referred on. This indicates there was a correlation between these assessments and RR outcomes. Of these assessments, the lower the scores on *T1 Matching*, *T1 LN*, *T2 Blending* and *T2 Matching*, the more likely a student was to be referred on from RR. This analysis showed the second hypothesis was moderately supported by the findings, but the correlation was not as strong as was expected. Due to the small sample size, it is difficult to measure the significance of the correlation between the variables and the RR outcomes, so these findings can only be used as indications of what may be true for the wider population.

Whilst testing the two hypotheses, it was discovered that a combination of T1 assessment scores (combination (B) made up of Matching, Blending, Elision, LN and LS) showed a strong correlation with student reading levels after two years at school. This suggests, an assessment administered in a child's first term of school, using a combination of PA, LN, and LS, may indicate if the child is at risk of reading difficulties after two years at school. This study has shown, by using effective early literacy assessment tools, it may be possible to begin early identification of students at risk of literacy difficulty. This is important because the earlier at-risk readers can be identified, the earlier their support can begin.

Implications for future research

The current study included a small sample of students selected for Reading Recovery, and a smaller sample of students exiting RR. Further research, involving larger sample sizes, would allow more accurate conclusions regarding the predictability of RR selection and outcomes. Furthermore, a longitudinal study over several years could track the RR selection and outcomes from participating schools. The current study found the students selected for RR did not include all the lowest readers in the cohort. A longitudinal study of RR selection would identify whether the variations in RR selection were present across the wider sample, or were restricted to certain schools within the population.

In this study, it was indicated that a combination of assessments, administered in the first term of primary school may indicate reading outcomes after two years at school. In a study with a larger sample size, it would be useful to confirm which combination of early

literacy assessments was most predictive of later outcomes. Research could then investigate which interventions were most effective in reducing the number of at-risk students. In addition, it would be very useful to know whether these interventions increased the rates of successful discontinued children in RR. Controlled studies with a large sample would need to be carried out to identify which intervention would be most appropriate with the at-risk students, but this appears to be a worthwhile avenue to explore to achieve higher success rates with RR.

Once students are identified as being at-risk of later reading difficulties, further research could explore the characteristics of treatment resisters. Research is needed to identify the ideal intervention content, intensity and duration, so that treatment resisters can be minimised. Current research about treatment resisters lacks clarity about the level of intervention the participants have received prior to participation in the studies.

Therefore, it will be important to ensure there is clarity about intervention levels in future research.

The final area suggested for future research is oral language. The literature review showed oral language is often found to be predictive of later reading difficulties, yet in this study, the *British Picture Vocabulary Scale* did not show any influence on the predictability of selection for RR or of RR outcomes. In addition, when assessment scores were combined to predict reading levels for the cohort, BPVS was not found to contribute. However, the other variables from this study (LN, LS and PA) were all useful in indicating future reading

outcomes. Obviously, the sample size in this study was small and therefore the findings cannot be generalised, so the influence of oral language warrants further investigation in future research.

Implications for New Zealand educators

The results from this study indicate changes to the assessment battery for selection of students for RR could be considered. The current RR selection process is largely based on the results of the Observation Survey of Literacy Achievement (OSELA) which is administered when students are aged six. If this selection process continues, the results of this study indicate the addition of a measure of phonological awareness may provide greater accuracy in identifying those with the greatest need and also indicating RR outcomes.

Establishing new criteria for the allocation of RR funding is another area for consideration. As shown in Figure 3, there is currently a wide variation in reading levels when students enter RR, because schools select the students with the lowest reading levels in the school, not the lowest reading levels of all 6-year-olds in New Zealand. While at one school, the lowest readers enter RR already reading at level 15, at another school there may be many students reading at levels 1 and 2 but only the lowest 20% in each school are entitled to receive RR support. Therefore, the current selection process may be considered inequitable if there are children with low levels of skill who cannot be given RR support because their school has already filled their 20%. This normative approach means some

children in the 'lowest readers in the school' group may actually be achieving adequate levels of literacy skills and may not require such intensive intervention, and this prevents the intervention funding reaching those who are most vulnerable (Torgesen, 2000). In order to ensure the students with the lowest literacy attainment receive the appropriate support, a change in the selection criteria could be considered, from an arbitrary cut-off point to a minimum literacy attainment level which is applied to all students across the entire cohort. The findings of this study showed the importance of phonological awareness for indicating RR outcomes for RR and for indicating future reading levels, so perhaps a RR selection measure could be derived from a combination of assessments that include a PA assessment with the OSELA, or instead, a use combination of PA, LN and LS as found in this study.

A further area of consideration is the ideal time to administer early literacy assessments. In this study, early literacy assessments were administered at school entry (T1), and mid-way through the first year of schooling (T2). Screening in the second term of schooling has been recommended as the ideal time to administer early literacy assessments (Torgesen, 1998). However, Figure 5 indicated that administering assessments for LN, LS and PA in the first term of school is the timing that is the most predictive of later reading outcomes. Furthermore, the variables this study are able to be administered by the classroom teacher and would only take approximately 10 minutes per child.

A final implication for educators, is provision of appropriate literacy intervention to readers at risk of reading difficulty. The findings of this study suggest, students who score 20 points or less as a T1 combined score, are at greatest risk of reading difficulty, therefore, intervention should begin as quickly as possible to prevent further difficulty occurring. It is important to acknowledge that some students will experience reading success even when they score 20 points or less (i.e. they are false positives), but the risk is greater for low scoring students, so it is best to provide support to the whole group of low-scoring students to ensure everyone has the best opportunity to succeed. When considering which interventions would be most appropriate, intensive instruction in phonemic awareness and alphabetic coding has been found effective (Tunmer & Prochnow, 2009), while others advocate the use of programmes that target phonological deficits for remediation (McKenna & Violato, 2003; Torgesen et al., 1997). It has been identified that a short, intensive programme of phonological awareness instruction can be effective in raising the literacy attainment of 5-year-olds (Carson et al., 2013). These areas definitely warrant further consideration to ensure those children who really need intensive literacy support receive the right support at the right time.

Summary

This study investigated whether it was possible to predict RR selection and outcomes. It was found it was not possible to accurately predict which students would be selected for Reading Recovery due to the variations in RR selection processes. Furthermore, it was found that the higher RR children's phonological awareness scores were, the more likely

they were to be successfully discontinued. It was also found that if a child scored 20 points or less, in a combination of T1 assessments (letter names, letter sounds and three measures of phonological awareness), they were likely to have a body of literacy abilities that meant they would be working at least a year below the National Standard by the end of their second year at school. These findings indicate that standardising the selection of students for RR may mean more students with the lowest literacy attainment get support. In addition, early literacy assessments, including measures of phonological awareness, should be administered early in a child's schooling and those identified as being at risk of reading difficulty should receive literacy support without delay. Addressing students' low levels of phonological awareness in the first year of schooling may lead to better outcomes for students who participate in RR. Therefore, it can be seen that further research is required, so we can ensure our funding for early literacy interventions is used in a way that brings about the best outcomes for at-risk readers.

APPENDIX

Means and standard deviations for all groups and all variables

| | Students in RR Schools | | Students selected for RR | | Successfully Discontinued from RR | | Referred on from RR | |
|------------------------------------|---------------------------|-----------|-----------------------------|-----------|---|-----------|------------------------|-----------|
| | n=224 | | n=34 | | n=11 | | n=6 | |
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| T1 Letter Names (54) | 22.80 | 17.88 | 8.27 | 11.08 | 9.91 | 7.98 | 4.50 | 3.45 |
| T1 Letter Sounds (54) | 12.66 | 16.03 | 2.41 | 7.00 | .55 | 1.03 | .50 | .83 |
| T1 Elision (34) | 4.53 | 4.56 | 1.95 | 3.39 | 2.00 | 2.53 | 1.33 | 1.96 |
| T1 Blending (33) | 6.81 | 4.63 | 4.89 | 3.45 | 5.45 | 3.70 | 5.50 | 2.81 |
| T1 Matching (26) | 7.15 | 5.88 | 4.35 | 3.64 | 4.91 | 1.97 | 1.67 | 1.51 |
| T1 PA Total (93) | 18.49 | 12.73 | 11.19 | 7.39 | 12.36 | 3.64 | 8.50 | 3.73 |
| T1 BPVS (168) | 70.05 | 17.66 | 58.46 | 20.13 | 62.73 | 16.64 | 58.00 | 22.59 |
| T2 Letter Names (54) | 42.83 | 14.33 | 30.72 | 17.36 | 35.09 | 16.67 | 24.00 | 20.86 |
| T2 Letter Sounds (54) | 37.04 | 17.66 | 20.58 | 18.05 | 27.27 | 13.37 | 13.67 | 18.50 |
| T2 Elision (34) | 9.65 | 6.25 | 5.88 | 4.58 | 7.45 | 4.78 | 5.17 | 2.79 |
| T2 Blending (33) | 12.23 | 6.30 | 7.55 | 4.38 | 8.00 | 3.61 | 5.17 | 3.13 |
| T2 Matching (26) | 12.97 | 6.88 | 8.13 | 5.45 | 10.91 | 4.72 | 6.00 | 3.23 |
| T2 PA Total (93) | 34.88 | 17.16 | 21.09 | 10.73 | 26.36 | 7.71 | 16.33 | 7.97 |
| T3 Reading Book Level | 10.47 | 5.08 | 6.83 | 3.56 | 7.30 | 2.63 | 6.50 | 3.78 |
| T5 Reading Book level | 17.73 | 4.93 | 13.82 | 3.87 | 16.91 | .94 | 11.83 | 1.47 |
| T1 Combined Total Score (A) | 123.59 | 53.04 | 80.32 | 33.15 | 85.55 | 20.22 | 71.50 | 24.38 |
| T2 Combined Total Score | 115.59 | 43.22 | 70.13 | 40.15 | 88.73 | 28.31 | 54.00 | 42.21 |
| T1 LN, LS, Eli, Ble, Mat Total (B) | 53.95 | 42.51 | 21.86 | 21.69 | 22.82 | 7.87 | 13.50 | 5.54 |
| T1 LN, Eli, Ble, Mat Total | 41.29 | 27.81 | 19.46 | 15.90 | 22.27 | 7.98 | 13.00 | 5.22 |
| T1 LN and Mat Total | 29.94 | 21.61 | 12.62 | 12.93 | 14.82 | 8.11 | 6.17 | 4.22 |
| T2 LN, LS, Ble and Eli Total | 102.71 | 38.07 | 63.97 | 38.19 | 77.82 | 27.46 | 48.00 | 39.14 |
| T2 LN, LS, Ble and Mat Total | 99.84 | 37.31 | 62.64 | 37.81 | 75.27 | 26.70 | 44.50 | 38.86 |

Note: T1 = Time 1, T2 = Time 2, T3 = Time 3, T6 = Time 6, LN = Letter Names, LS = Letter

Sounds, Eli = Elision, Ble = Blending, Mat = Matching, PA = Phonological Awareness, BPVS =

British Picture Vocabulary Scale, RR = Reading Recovery

References

- Adams, M. J. (1990). *Beginning to read: Learning and thinking about print*. Cambridge, MA: MIT Press.
- Adams, M. J., Foorman, B. R., Lundberg, I., & Beeler, T. (1998). *Phonemic awareness in young children: A classroom curriculum*. Baltimore, MD: Paul Brookes pub.
- Al Otaiba, S. (2003). Identification of non-responders: Are the children "left behind" by early literacy intervention the "truly" reading disabled? *Advances in Learning and Behavioral Disabilities, 16*, 51-82.
- Al Otaiba, S., & Fuchs, D. (2006). Who are the young children for whom best practices in reading are ineffective? An experimental and longitudinal study. *Journal of Learning Disabilities, 39*(5), 414-431.
- Anand, V., & Bennie, N. (2005). *Annual monitoring of Reading Recovery: 2003 data*. Wellington, NZ: Ministry of Education.
- Arrow, A. W. (2012). *Alphabet knowledge in the acquisition of early orthographic representation*. Paper presented at the Scientific Studies in Reading Annual Meeting, Montreal, Canada.
- Belgrave, J. (2009). *The characteristics of children who are referred on from Reading Recovery*. (Master of Education), University of Waikato, Hamilton, NZ.
- Blachman, B. A. (1984). Relationship of rapid naming ability and language analysis skills to kindergarten and first-grade reading achievement. *Journal of Educational Psychology, 76*(4), 610.
- Blachman, B. A. (1994). What we have learned from longitudinal studies of phonological processing and reading, and some unanswered questions: A response to Torgesen, Wagner, and Rashotte. *Journal of Learning Disabilities, 27*(5), 287-291.
- Blachman, B. A. (2000). Phonological awareness. *Handbook of Reading Research, 3*, 483-502.
- Carson, K. L., Gillon, G. T., & Boustead, T. M. (2013). Classroom Phonological Awareness Instruction and Literacy Outcomes in the First Year of School. *Language, Speech & Hearing Services in Schools, 44*(2), 147-160. doi:0161-1461(2012/11-0061)
- Castle, J. M., Riach, J., & Nicholson, T. (1994). Getting off to a better start in reading and spelling: The effects of phonemic awareness instruction within a whole language program. *Journal of Educational Psychology, 86*(3), 350.

- Catts, H. W., Compton, D., Tomblin, J. B., & Bridges, M. S. (2012). Prevalence and nature of late-emerging poor readers. *Journal of Educational Psychology, 104*(1), 166.
- Catts, H. W., & Kamhi, A. (Eds.). (2005). *The connections between language and reading disabilities*. Mahway, NJ: Erlbaum.
- Center, Y., Wheldall, K., Freeman, L., Outhred, L., & McNaught, M. (1995). An evaluation of reading recovery. *Reading Research Quarterly, 240-263*.
- Chapman, J. W., Arrow, A., Tunmer, W. E., & Braid, C. (2016). *Early Literacy Research Project*. Unpublished data set. Massey University. New Zealand.
- Chapman, J. W., Tunmer, W. E., & Prochnow, J. E. (2001). Does Success in the Reading Recovery Program Depend on Developing Proficiency in Phonological-Processing Skills? A Longitudinal Study in a Whole Language Instructional Context. *Scientific Studies of Reading, 5*(2), 141-176. doi:10.1207/S1532799Xssr0502_2
- Clay, M. M. (1991). *Becoming literate: The construction of inner control*. Portsmouth: Heinemann Educational Books.
- Clay, M. M. (1993). *An observation survey of early literacy achievement*. Portsmouth: Heinemann.
- Clay, M. M. (2005). *An observation survey of early literacy achievement* (rev. 2nd ed.): Portsmouth, NH : Heinemann,.
- D'Agostino, J. V., & Murphy, J. A. (2004). A meta-analysis of Reading Recovery in United States schools. *Educational Evaluation and Policy Analysis, 26*(1), 23-28.
- Dunn, L., Dunn, D., & Styles, B. (2009). *British Picture Vocabulary Scale* (3rd ed.). London: GL Assessment.
- Ehri, L. (2005). Development of sight word reading: phases and findings. In M. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 135-154). Maiden, MA: Blackwell Publishing Ltd.
- Ehri, L., & McCormick, S. (1998). Phases of word learning: Implications for instruction with delayed and disabled readers. *Reading & Writing Quarterly, 14*(2), 135-163. doi:10.1080/1057356980140202
- Ehri, L., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of Educational Research, 71*(3), 393-447.

- Fazio, B. B. (1997). Learning a new poem memory for connected speech and phonological awareness in low-income children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research, 40*(6), 1285-1297.
- Foorman, B. R., Francis, D. J., Winikates, D., Mehta, P., Schatschneider, C., & Fletcher, J. M. (1997). Early interventions for children with reading disabilities. *Scientific Studies of Reading, 1*(3), 255-276.
- Foulin, J. N. (2005). Why is letter-name knowledge such a good predictor of learning to read? *Reading and Writing, 18*(2), 129-155.
- Gilmore, A., Croft, C., & Reid, N. A. (1981). *Burt Word Reading Test: Teachers Manual*: New Zealand Council for Educational Research.
- Gough, P. B., & Juel, C. (1991). The first stages of word recognition. In L. Rieben & C. Perfetti (Eds.), *Learning to read: Basic research and its implications* (pp. 47-56). Hillsdale, NJ: Lawrence Erlbaum.
- Greaney, K., & Arrow, A. (2012). Phonological-based assessment and teaching within a first year reading program in New Zealand. *Australian Journal of Language and Literacy, The, 35*(1), 9.
- Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing, 2*(2), 127-160.
- Iversen, S., & Tunmer, W. E. (1993). Phonological processing skills and the Reading Recovery Program. *Journal of Educational Psychology, 85*(1), 112.
- Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80*(4), 437.
- Juel, C., Griffith, P. L., & Gough, P. B. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology, 78*(4), 243.
- Laerd Statistics. (2015). Binomial logistic regression using SPSS statistics. Statistical tutorials and software guides. Retrieved from <https://statistics.laerd.com/>
- Lesaux, N. K., & Kieffer, M. J. (2010). Exploring sources of reading comprehension difficulties among language minority learners and their classmates in early adolescence. *American Educational Research Journal, 47*(3), 596-632.
- Liberman, I. Y., Shankweiler, D., & Liberman, A. M. (1989). The alphabetic principle and learning to read. In D. Shankweiler & I. Y. Liberman (Eds.), *Phonology and reading disability: Solving the reading puzzle* (pp. 1-33). Ann Arbor, MI: University of Michigan Press.

- May, H., Sirinides, P., Gray, A., & Goldsworthy, H. (2016). *Reading Recovery: An evaluation of the four-year i3 scale-up*. CPRE Research Reports. Retrieved from http://repository.upenn.edu/cpre_researchreports/81.
- McDowall, S., Boyd, S., Hodgen, E., & Vliet, T. (2005). *Reading Recovery in New Zealand: Uptake, implementation, and outcomes, especially in relation to Maori and Pasifika students*. Wellington: New Zealand Council for Educational Research.
- McKenna, G. S., & VioLato, C. (2003). A combined phonological awareness, whole language, and metacognitive approach to teach reading to reading resisters. *Canadian Journal of School Psychology, 18*(1-2), 115-128.
- Ministry of Education. (2010). *The literacy learning progressions: Meeting the reading and writing demands of the curriculum*. Wellington, New Zealand: Learning Media Limited.
- Ministry of Education. (2012). Student support: Reading Recovery. Retrieved from <https://nzcurriculum.tki.org.nz/content/download/24470/233689/file/Reading%20Recovery%20Supplementary%20Supports%20Inventory%20for%20Schools.doc>
- Ministry of Education. (2014). *Annual monitoring of Reading Recovery: The data for 2013*. Wellington: Ministry of Education.
- Muter, V., Hulme, C., Snowling, M., & Taylor, S. (1997). Segmentation, not rhyming, predicts early progress in learning to read. *Journal of Experimental Child Psychology, 65*(3), 370-396.
- National Reading Panel. (2000). *Report of the National Reading Panel: Teaching children to read: an evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: reports of the subgroups*. Washington, D.C.: National Institute of Child Health and Human Development.
- Nelson, J. R., Benner, G. J., & Gonzalez, J. (2003). Learner characteristics that influence the treatment effectiveness of early literacy interventions: A meta-analytic review. *Learning Disabilities Research & Practice (Wiley-Blackwell), 18*(4), 255. doi:10.1111/1540-5826.00080
- Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology, 49*(12), 1373-1379.
- Richardson, E., & DiBenedetto, B. (1985). *Decoding skills test*. Parkton, MD: York Press.

- Scarborough, H. S. (1998). Predicting the future achievement of second graders with reading disabilities: Contributions of phonemic awareness, verbal memory, rapid naming, and IQ. *Annals of Dyslexia*, 48(1), 115-136. doi:10.1007/s11881-998-0006-5
- Scarborough, H. S. (2005). Developmental relationships between language and reading: Reconciling a beautiful hypothesis with some ugly facts. In H. Catts & A. Kamhi (Eds.), *The connections between language and reading disabilities* (pp. 3-24). Mahwah, NJ: Erlbaum.
- Serry, T., Rose, M., & Liamputtong, P. (2014). Reading Recovery teachers discuss Reading Recovery: a qualitative investigation. *Australian Journal of Learning Difficulties*, 19(1), 61-73. doi:10.1080/19404158.2014.909862
- Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children: Report of the committee on the prevention of reading difficulties in young children* Washington D.C.: National Research Council.
- Spear-Swerling, L. (2013). A road map for understanding reading disabilities and other reading problems, redux. In D. Alvermn, N. Unrau, & R. Ruddell (Eds.), *Theoretical models and processes of reading* (6th ed., pp. 412-436). Newark, USA: International Reading Association.
- Spear-Swerling, L., & Sternberg, R. J. (1994). The Road Not Taken An Integrative Theoretical Model of Reading Disability. *Journal of Learning Disabilities*, 27(2), 91-103.
- Stahl, S. A., & Murray, B. A. (1994). Defining phonological awareness and its relationship to early reading. *Journal of Educational Psychology*, 86(2), 221.
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 360-407.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson/Allyn & Bacon.
- Torgesen, J. K. (1998). Catch them before they fall. *American Educator*, 22, 32-41.
- Torgesen, J. K. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research & Practice*, 15(1), 55-64.
- Torgesen, J. K., & Davis, C. (1996). Individual difference variables that predict response to training in phonological awareness. *Journal of Experimental Child Psychology*, 63(1), 1-21.
- Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1997). Prevention and remediation of severe reading disabilities: Keeping the end in mind. *Scientific Studies of Reading*, 1(3), 217-234.

- Tunmer, W., & Chapman, J. (2003). The Reading Recovery approach to preventive early intervention: As good as it gets? *Reading Psychology, 24*(3-4), 337-360.
doi:10.1080/02702710390227378
- Tunmer, W., Herriman, M., & Nesdale, A. (1988). Metalinguistic abilities and beginning reading. *Reading Research Quarterly, 134*-158.
- Tunmer, W., & Prochnow, J. (2009). Cultural relativism and literacy education: Explicit teaching based on specific learning needs is not deficit theory. In R. Openshaw & E. Rata (Eds.), *The politics of conformity in New Zealand* (pp. 154-190). Auckland: Pearson.
- Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R., & Denckla, M. B. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology, 88*(4), 601.
- Vittinghoff, E., & McCulloch, C. E. (2006). Relaxing the rule of ten events per variable in logistic and cox regression. *American Journal of Epidemiology, 165*(6), 710-718.
doi:10.1093/aje/kwk052
- Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2013). *Comprehensive Test of Phonological Processing: CTOPP2* (2nd ed.). Austin, Texas: Pro-Ed.
- What Works Clearinghouse. (2007). Reading Recovery intervention report. Retrieved from <https://readingrecovery.org/reading-recovery/research/what-works-clearinghouse>