EFFICIENT AND EFFECTIVE CLASSROOM PHONOLOGICAL AWARENESS PRACTICES TO IMPROVE READING ACHIEVEMENT

Karyn Louise Carson

BSLT (first class honours)

A thesis submitted for the degree of

Doctor of Philosophy

College of Education

University of Canterbury

New Zealand

April 2012

Declaration of Originality

I certify that this thesis does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any university; and that to the best of my knowledge and belief it does not contain any material previously published or written by another person except where due reference is made in the text.

I also cerfity that the research reported in this thesis has been approved by the University of Canterbury Human Ethics Committee.

Signed: ______________________ On: _____/____/_____

Acknowledgements

'All of life is an experiment. The more experiments you make the better.'

– Ralph Waldo Emerson

In July 2008, I approached the University of Canterbury and my primary supervisor enquiring about the possibility of commencing a PhD in reading acquisition for school-aged children. My most vivid memory of this encounter was my soon-to-be primary supervisor stating that, *"… a PhD is a **significant** piece of original research"* – almost four years on my understanding of this phrase, appreciation for the research process and awareness of how evidence-based practice can enhance children's quality of life has grown in leaps and bounds.

The road to preparing this thesis has been characterised by many highs and lows, throughout which numerous individuals have offered support and encouragement in helping me get to this final stage. First, I would like to extent my utmost thanks to my primary supervisor Professor Gail Gillon and to my secondary supervisor Dr Therese Boustead for their mentorship, expertise and direction in preparing this thesis. Without their guidance, I would not have reached the accomplishment of thesis submission.

I would like to express sincere thanks to the children, parents, teachers and wider school communities who so generously offered their time to participate in the various experiments that comprise this document. Special thanks are extended to Pat Coope for her support with statistical analysis, to Alexandre Heitz for his support in computer programming, and to Ellen Nijhof for her support in data collection, analysis and reliability work. Thanks are also expressed to the New Zealand Tertiary Education

Commission for financial support during my time as a doctoral candidate through their award of the *New Zealand Top Achiever Doctoral Scholarship*.

I owe my deepest gratitude to a number of friends who have relentless supported me with not only words of encouragement but also a listening ear during the preparation of this thesis. Special thanks to Meagan Boakes, Megan Carpenter, Aimee Schicker and Jane Carroll for constant reminders that *"… the race is long and in the end it's only with yourself"*. To my ironman coach John Ellis and running inspiration Brett Tingay for teaching me how to train my mind to overcome the impossible because *"… the very word itself says, I'm possible"*. I owe much thanks to my parents (Catherine and Tommy Long) and siblings (Julie, Michael and Richard) for their support not only in this PhD journey, but also with whatever life has thrown my way.

Finally and most importantly, it is an honor to offer my most heart-felt thanks to my husband, Robert Carson. His patience, love and constant reminders that "…a ***PhD is an ultra-ironman brain marathon ….you must remember to pace yourself"*** have been truly valued.

~ *Karyn Carson*

Abstract

International studies of reading achievement demonstrate that significant inequalities in reading outcomes continue to exist among some of the world's wealthiest countries, despite strong investment in initiatives directed towards raising literacy achievement for all children (United Nations Educational, Scientific and Cultural Organisation—UNESCO, 2009; United Nations Children's Fund—UNICEF, 2010). One approach towards the elevation of reading achievement is to investigate how key predictors of reading success are incorporated into everyday classroom literacy practices. Phonological awareness (PA) is widely recognised as a powerful predictor and underlying precursor to early reading success for both typically developing and at-risk readers (Al Otaiba, Kosanovich, & Torgesen, 2012; Blachman, Ball, Black, & Tangel, 2000; Goswami, 2001; Pressley, 2006). A majority of research demonstrating the benefits of PA to literacy growth has been conducted under controlled research settings outside of the classroom environment (Ehri, Nunes, Willows, Schuster, Yaghoub-Zadeh, & Shanahan, 2001; Gillon, 2000a, 2005; Gillon & McNeill, 2009), and thus less is known about whether such benefits hold true when integrated into the heterogeneous classroom setting. For this reason, four experiments reported in this thesis investigated whether PA can be efficiently and effectively integrated into the classroom literacy programme with the overarching aim of raising reading achievement and equalising reading outcomes for the majority of children in the first year of formal education.

In the first experiment (reported in Chapter 3), time-efficiency and congruency of scores between a computer-based PA screening and monitoring tool (described in Chapter 2) and a paper-based equivalent were examined. Thirty-three children aged between four years 10 months and five years zero months participated in the study, 12 of

whom presented with moderate-severe speech delay (MSD). Participants were randomly allocated to either Group A or Group B experimental assessment conditions. A crossover research design was employed where Group A received the paper-based version of the PA assessment followed two weeks later by the equivalent computer-based assessment (CBA). Group B received the same assessments but in the reverse order of delivery. That is, the computer-based PA assessment first followed two weeks later by the paper-based counterpart. Results demonstrated that: 1) the CBA generated comparable scores to the paper-based equivalent for both children with typical development and children with MSD, and 2) CBA took 31 per cent less time than paper-based administration. These results demonstrate that CBA can provide educators with a time-efficient approach to the screening and monitoring of PA development in the classroom while maintaining equivalency of scores with paper-based testing. Having established the time-efficiency of CBA, the next step was to investigate the use of the computer-based PA screening and monitoring tool as part of the beginning classroom reading programme.

In the second experiment (reported in Chapter 4), the influence of a short and intensive period of teacher-implemented classroom PA instruction on reading outcomes in the first year of education was investigated. One-hundred and twenty-nine children aged five-years participated in the study. Using a quasi-experimental design, thirty-four children in two classrooms received 10 weeks of PA instruction from their teachers, as an adjunct to the 'usual' reading programme. Ninety-five children from 10 classrooms continued with the 'usual' reading programme, which included phonics instruction but did not target PA. Results demonstrated that children exposed to classroom PA instruction performed significantly higher on reading and spelling measures compared to children who received the 'usual' reading programme only. Of importance, the number of children experiencing word decoding difficulties after one year of schooling reduced from 26 per cent among children who followed the ‘usual’ reading programme to 6 per cent among children who received classroom PA instruction. These results provide evidence that a short and intensive period of classroom-wide PA instruction in the first year of schooling can have a positive influence on raising reading achievement.

In the third experiment (reported in Chapter 5), the effect of classroom PA instruction on raising reading achievement and reducing inequality in literacy outcomes for children with spoken language impairment (SLI) was examined. The data from 129 five-year-old children who participated in the second experiment were extracted and analysed. End-of-year reading outcomes between children with SLI who received classroom PA instruction (n = 7) was compared to: 1) children with typical language development (TD) who received classroom PA instruction (n = 27), 2) children with SLI who followed the ‘usual’ reading programme (n = 21), and 3) children with TD who followed the ‘usual’ reading programme (n = 74). Children with SLI who received classroom PA instruction showed significant improvements in PA, reading and spelling acquisition immediately and up to six months following PA instruction. However, this cohort, in comparison to children with TD, appeared less able to transfer their enhanced PA knowledge to reading and writing tasks. Of importance, children with SLI who received PA instruction performed significantly higher than children with SLI who followed the ‘usual’ reading curriculum; and on par with children with TD who followed the ‘usual’ reading programme. Children with TD who received classroom PA instruction significantly outperformed all other cohorts in this experiment on end-of year reading measures. These results indicate that both children with TD and children with risk for reading difficulties can benefit from classroom-wide teacher-directed PA instruction. These findings have positive implications for elevating reading achievement and reducing inequality between good and poor readers.

In the fourth experiment (reported in Chapter 6), the validity and reliability of the computer-based PA screening and monitoring tool was investigated and established. Using a longitudinal research design, the responses of 95 children to test items in the CBA at the start, middle and end of the first year at school were collated and analysed to provide evidence of content, construct and criterion validity, in addition to test-retest and internal consistency reliability. A number of statistical analyses were employed including Rasch Model analysis, exploratory factor analysis and multiple regression analysis. Results demonstrated that the majority of test items were appropriate for five-year-old children in the first year of school and sampled a spectrum of ability levels that would be present in a typical classroom environment. Rhyme oddity, initial phoneme identity and letter-knowledge tasks were most appropriate at school-entry while tasks of final phoneme identity, phoneme blending and phoneme segmentation became more suitable by the middle and end stages of the first year at school. Importantly, performance on the CBA predicted end-of-year reading status with 94 per cent accuracy, and in conjunction with language abilities accounted for 68.9 per cent of the variance in end-of-year reading performance. These findings indicate that the computer-based PA screening and monitoring tool developed and applied in this thesis has sufficient validity and reliability to be used confidently as a time-efficient assessment tool in the classroom.

The results from the experiments reported in this thesis provide evidence that PA can be efficiently and effectively integrated into the beginning classroom reading programme from two complementary perspectives: 1) through use of computer-based screening and monitoring of PA skills, and 2) through implementation of a short and intensive period of teacher-directed classroom-wide PA instruction. The results reported in this thesis demonstrate that the evidenced-based integration of key predictors of literacy success, such as PA, into existing classroom programmes can support national and international initiatives that seek to raise reading achievement and reduce inequalities in literacy outcomes for all children.

Publications arising from this thesis

Carson, K., Gillon, G., & Boustead, T. (in second stage review process). Classroom phonological awareness instruction and literacy outcomes in the first year of school. *Language, Speech, and Hearing Services in Schools.*

Carson, K., Gillon, G., & Boustead, T. (2011). Computer administered versus paper-based assessment of school-entry phonological awareness ability. *Asia Pacific Journal of Speech, Language, and Hearing 14*(2), 85-101.

Presentations with published abstracts

Carson, K., Gillon, G., & Boustead, T. (2011). *Classroom-based phonological awareness intervention effectiveness for children with language delay*. Oral Presentation at the 2011 American Speech, Language and Hearing Association Convention: Beacons of Inspiration – Innovation to Action, 17th – 19th November, San Diego, United States of America.

Carson, K., Gillon, G., & Boustead, T. (2011). *Predicting literacy outcomes using computer-based phonological awareness assessment at school entry*. Oral Presentation at the 2011 American Speech, Language and Hearing Association Convention: Beacons of Inspiration – Innovation to Action, 17th-19th November, San Diego, United States of America.

Gillon, G., McNeill, B., van Bysterveldt, A., Carroll, J., & Carson, K. (2011). *Delivering phonological awareness within the classroom context.* Seminar at the 2011 American Speech, Language and Hearing Association Convention: Beacons of Inspiration – Innovation to Action, 17th-19th November, San Diego, United States of America.

Carson, K., Gillon, G., & Boustead, T. (2011). *Teacher-implemented phonological awareness instruction and literacy outcomes in the first year of school*. Poster Presentation at the 2011 Society for the Scientific Study of Reading Conference, 13th-16th July, Florida, United States of America.

Carson, K., Gillon, G., & Boustead, T. (2011). *Introducing phonological awareness into the classroom reading programme.* Oral Presentation at the 2011 Speech Pathology Australia National Conference: Diversity & Development, 27th- 29th June, Darwin, Australia.

Carson, K., Gillon, G., & Boustead, T. (2010). *Relationships between home literacy environments and school-entry phonological awareness in children with and without speech delay*. Oral Presentation at the International Association of Logopedics and Phoniatrics, 22nd-26th August, Athens, Greece (Full Conference Paper).

Carson, K., Gillon, G., & Boustead, T. (2010). *Computer-based phonological awareness assessment at school-entry*. Oral Presentation at the International Association of Logopedics and Phoniatrics, 22nd-26th August, Athens, Greece (Full Conference Paper).

Carson, K., Gillon, G., & Boustead, T. (2010). *Classroom-based phonological awareness screening: The use of computer versus conventional assessment methods.* Oral Presentation at the New Zealand Speech-Language Therapists' Association Conference: A Practice Worth Spreading, Wellington, New Zealand.

Carson, K., Gillon, G., & Boustead, T. (2009). *The use of technology in phonological awareness assessment: A pilot study.* Poster Presentation at the American Speech, Language, and Hearing Association Convention, New Orleans, United States of America.

Carson, K., Gillon, G., & Boustead, T. (2009). *Computer-based technology in classroom reading assessment*. Poster Presentation at the Literacy Research Symposium, College of Education, University of Canterbury, Christchurch, New Zealand.

Contents

List of Figures

List of Tables

List of Abbreviations

ANOVA	analysis of variance
BPS	British Psychology Society
CBA	computer-based assessment
CBM	Curriculum-Based Measurement
CELF-P2	Clinical Evaluations of Language Fundamentals, Preschool, Second Edition
CoPS	Cognitive Profiling System
CTOPP	Comprehensive Test of Phonological Processing
CTT	Classical Test Theory
DIBELS	Dynamic Indicators of Basic Early Literacy Skills
IPA	International Phonetic Alphabet
IQ	intelligence quotient
IRT	Item Response Theory
LAC	Lindamood Auditory Conceptualization Test, Revised Edition
MOE	Ministry of Education
MSD	moderate-severe speech delay
NAEP	National Assessment of Educational Progress
NARA	Neale Analysis of Reading Ability, Third Edition
NPV	Negative Prediction Value
NZAT	New Zealand Articulation Test
OECD	Organization for Economic Co-operation and Development
OS	Observational Survey of Early Literacy

P1	Phase 1
P2	Phase 2
PA	phonological awareness
PALS	Phonological Awareness Literacy Screening
PAT	Gillon Phonological Awareness Training Programme
PCC	percentage of consonants correct
PIPA	Primary Inventory of Phonological Awareness
PIPS	Performance Indicators in Primary School On-Entry Baseline Assessment
PIRLS	Progress in International Reading Literacy Study
PISA	Programme of International Student Assessment
PPV	Positive Prediction Value
PROPH	Profiling of Phonology
PTONI	Primary Test of Non-verbal Intelligence
RIA	Read It Again
RTI	Response to Intervention
SEA	School-Entry Assessment
SLI	spoken language impairment
SPSS	Statistical Package for the Social Sciences
TCC	test characteristic curves
TD	typical spoken language development
TERA	Test of Early Reading Ability
TOPA	Test of Phonological Awareness
UNESCO	United Nations Educational, Scientific and Cultural Organization

UNICEF	United Nations Children’s Fund
VIF	variance inflation factor

Chapter 1: Literature Review

1.1 Introduction

Developing skilled reading is a significant milestone in the early years of schooling (Kamhi & Catts, 2012; Kern & Friedman, 2008). Many children will learn to read with ease and will enjoy the satisfaction that comes from interacting with a written world. Some children will approach beginning-reading instruction having had limited experience with print. However, with sufficient classroom instruction, they will progress on a trajectory towards reading competency. Conversely, other children will experience significant difficulties in learning to read, which will affect their academic, social and personal development (Nelson, 2010). These children are at serious risk for falling behind their typically developing counterparts in reading acquisition and for experiencing significant inequalities in educational outcomes (Morgan, Farakas, & Hibel, 2008; Stanovich, 1986).

Ensuring that children become proficient readers in their own classrooms is a critical issue in education. Reading programmes that are inclusive of skills that are highly predictive of early reading success are most effective in generating competent readers (Ehri et al., 2001). Phonological awareness (PA) is one important prerequisite for reading proficiency because it helps initiate word-recognition development, which in turns supports reading comprehension (Al Otaiba et al., 2012; Justice, 2006; Nelson, 2010). The pivotal role of PA in the early stages of learning to read means it has become widely known as a powerful predictor and prognostic marker for identifying risk for reading problems in the early school years (Blachman et al., 2000; Ehri et al., 2001; Goswami, 2001; Pressley, 2006). Many children classified as poor readers often present with deficiencies in PA knowledge (Catts, Fey, Zhang, & Tomblin, 1999). Over the last 30 years, a great deal of research examining how to measure and instruct children in PA has been conducted in controlled clinical settings under the guidance of university researchers or specialised professionals outside the classroom environment (Bradley & Bryant 1983; Byrne, Fielding-Barnsley, & Ashley, 2000; Ehri et al., 2001; Gillon, 2000a, 2005; Gillon & McNeill, 2009).

It has been reported that PA has a greater effect on student achievement compared to other classroom variables including financial resources, class size and whole language programmes (Hattie, 2005). Despite this positive effect, knowledge of how to best integrate PA assessment and instruction into everyday classroom environments is still unclear. This information is critical for developing programmes that maximise the quality of classroom PA instruction and will support educators in their efforts to raise reading achievement and reduce inequalities in learning outcomes. The current thesis investigates two methods of integrating PA into beginning classroom literacy programmes by: 1) examining the use of a computer-based modality to screen and monitor PA development and 2) investigating the benefits of a short and intensive period of teacher-directed PA instruction for raising achievement and equalising equality in reading outcomes.

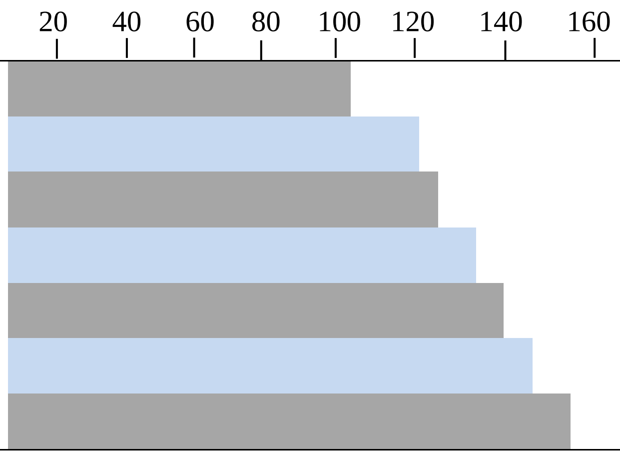
To understand the rationale underpinning the research presented in this thesis, the ensuing literature review is divided into three main areas that cover: 1) the international need to address inequalities in reading outcomes, 2) the theoretical context of PA in learning to read and 3) the importance of bridging this theoretical context into the practical environment of the classroom. The literature review concludes with a synopsis and three broad hypotheses that will be successively addressed to achieve the overarching aim of this thesis; that is, to contribute towards efforts aimed at raising reading achievement and reducing inequalities in school-aged literacy outcomes.

1.2 Reducing Inequality in School-Aged Reading Outcomes

The development of skilled reading and writing from the outset of formal schooling is critical for achieving life-long academic, vocational and social success (UNESCO, 2009). However, international prevalence statistics suggest that as many as one in three children may struggle with the acquisition of basic reading and writing skills (National Assessment of Educational Progress —NAEP, 2003; Nicholson, 2009). In New Zealand, the country in which this doctoral research was undertaken, it is estimated that approximately one in five children struggle with early literacy development (Nicholson, 2009). Recently, international studies of literacy achievement have examined the extent to which vulnerable readers from the Organization for Economic Co-operation and Development's (OECD) wealthiest countries are falling behind their typically developing counterparts in reading acquisition (UNICEF, 2010). These studies include the UNICEF Report Card 9: *The Children Left Behind* (2010), the *Programme of International Student Assessment* (PISA) (OECD, 2010) and the *Progress in International Reading Literacy Study* (PIRLS) (Martin, Mullis, & Kennedy, 2007). Large inequalities between the abilities of good and poor readers have been identified. Countries such as Finland, Denmark, Ireland and Canada demonstrate less variability between high- and low-ability readers, whereas countries such as the United States of America, the United Kingdom, Italy, France and New Zealand present with much larger gaps between the literacy outcomes of good and poor readers (UNICEF, 2010). As an example, Table 1.1 profiles the results from the 2006 PISA study for the top four countries presenting with the smallest gap between average and low-achieving readers (i.e., Finland, Denmark, Ireland and Canada) and for a selection of countries generating larger discrepancies in reading performance (i.e., United Kingdom, New Zealand and France). The PISA study collects data from over 40 countries worldwide on the ability of 15-year-old students to decode, comprehend, interpret and reflect on written information. Table 1.1 has been adapted from information available in the UNICEF Report Card 9: *The Children Left Behind* (2010, p. 9).

Table 1.1

Inequality in reading outcomes between average and low-achieving 15-year-old students

Country	Inequality of literacy outcomes between children in the 50th percentile and children in the 10th percentile 	50th percentile (median) score	10th percentile (median) score
Finland		550	441
Denmark		499	378
Ireland		522	395
Canada		534	402
UK		501	359
New Zealand		528	381
France		499	346

Note. The bar chart represents the size of the gap between the 50th percentile median score and the 10th percentile median score. Statistics taken from the UNICEF Report Card 9: *The Children Left Behind* (2010, p.9).

The PIRLS project, which evaluates trends in reading achievement for 10-year-old children in over 40 countries at five-yearly intervals, also provides evidence of inequalities in reading outcomes (Martin et al., 2007). A similar pattern to that reported in the UNICEF Report Card 9: *The Children Left Behind* (2010) is highlighted in which OECD countries such as New Zealand, England and the United States of America present with a large spread of scores between high- and low-ability students. Children who are more likely to be over-represented as poor readers include those with a disability (e.g., such as spoken language impairment), from minority and indigenous backgrounds, from rural or remote communities, and who are boys (Martin et al., 2007; UNESCO, 2009; UNICEF, 2010). International statistics from the UNICEF Report Card 9: *The Children Left Behind*, the PISA study and the

PIRLS project demonstrate that inequalities in reading outcomes between high risk groups and good readers is not closing, despite initiatives over the last decade to improve reading standards (Morgan et al., 2008). These findings have led international associations to call for a global commitment towards raising achievement and reducing reading inequalities for all children (UNESCO, 2009; UNICEF, 2010). The research reported in this thesis addresses this issue by investigating how one key predictor of literacy success, namely PA, can be effectively and efficiently included as part of a comprehensive classroom programme to help raise equality in reading achievement.

1.2.1 Key pillars of early reading success.

A wide range of options are currently under investigation to help address inequalities in reading achievement. One method of addressing reading inequality is to evaluate how key predictors of literacy success are measured and taught in the classroom. Reading is a linguistic skill that is reliant on the integration of sufficient phonological, semantic, syntactic and pragmatic spoken language abilities (Kamhi & Catts, 2012; Lonigan, Schatschneider, & Westberg, 2008). According to the National Reading Panel meta-analysis (Ehri et al., 2001), the five key pillars of reading success are:

1) phonological awareness
2) letter-sound knowledge
3) vocabulary development
4) reading fluency
5) comprehension strategies.

Each of these pillars is interrelated. Strong PA and letter-knowledge support the development of reading fluency, which in turn enables the use of vocabulary and comprehension strategies to access the meaning of written text (Ehri et al., 2001; Kamhi & Catts, 2012). This thesis focuses on examining the development of one of these five pillars—

PA—in addition to the link this has with letter-knowledge (i.e., alphabetic knowledge) when applying a decoding strategy to learning to read. PA is defined as the purposeful ability to attend to and manipulate the sound structure of spoken words at the syllable, rhyme and phoneme levels (Gillon, 2004). The more sensitive children are to the sound structure of spoken words, the more likely they are to become stronger readers, as this awareness allows children to link phonemes (i.e., sounds) to graphemes (i.e., letters) to decode printed words and access meaning (Al Otaiba et al., 2012). Torgesen, Wagner, and Rashotte (1994) found that children who began the first grade (i.e., at approximately six years of age) with PA abilities below the 20^{th} percentile performed on average three grade levels behind their peers in word-reading tasks by the fifth grade (i.e., at approximately 10 years of age). It has been widely reported that without adequate instruction in PA, the gap between good and poor readers widens over time in a phenomenon termed the 'Matthew Effect' (Stanovich, 1986); an effect similar to that reported in international studies of literacy achievement. In fact, researchers suggest PA plays a stronger role in predicting reading outcomes than educational measures such as intelligence, vocabulary, listening comprehension and socio-economic status (Lundberg, Olofsson, & Wall, 1980; MacDonald & Cornwell, 1995; Torgesen et al., 1994). Of particular concern is the development of literacy abilities in young children with spoken language impairment. These children are four to five times more likely to struggle with reading acquisition due to deficits in underlying skills, such as PA, that support written language development (Catts, Fey, Zhang, & Tomblin, 2001).

Over the last three decades, a large body of research surrounding the benefits of PA assessment and instruction has been generated (Anthony & Francis, 2005; Ehri et al., 2001). The majority of this research has been conducted under controlled research conditions with specialised professionals in one-to-one or small group frameworks outside the classroom environment (Bradley & Bryant 1983; Chard & Dickson, 1999; Ehri et al., 2001; Gillon, 2000a, 2005; Gillon & McNeill, 2009). More recently, efforts have been directed towards transitioning effective PA practices into the heterogeneous classroom environment in which teachers are programme implementers (e.g., Fuchs, Fuchs, Thompson, Al Otaiba, Yen, Yang, Braun, & O'Connor, 2001; Justice, McGinty, Cabell, Kilday, Knighton, and Huffman, 2010; Shapiro & Solity, 2008). Investigating how PA can be effectively utilised in the classroom environment may contribute to the elevation of reading achievement and the reduction of disparities at national and international levels.

Importantly, a number of practical barriers in the classroom may hinder the sustained inclusion of PA as part of the beginning literacy curriculum. These barriers include time availability, paucity of educators' own PA knowledge, difficulty in accessing resources, high material costs and systemic changes, such as teacher turnover rates (Barnett, 2003; Bryant et al., 2002; Carroll, Gillon, & McNeill, 2011). Although each of these barriers requires in-depth investigation, the specific barrier addressed in this doctoral research is that of time-efficiency. According to McLeod, Fisher, and Hoover (2003), the time required to implement a particular activity plays a critical role in determining whether it can be successfully implemented as part of classroom practice; for example, activities that are too time consuming may be omitted by teachers in an attempt to balance a busy classroom schedule. Given the importance of PA to the prediction and development of proficient reading (Ehri et al., 2001), it appears critical that educators are equipped with time-efficient methods of monitoring and providing instruction in this area of development. Ensuring teachers can easily include PA as part of the core beginning-reading programme may help reduce inequalities in reading achievement by ensuring children at risk for reading difficulties are identified early.

This thesis investigates classroom PA practices from two complementary perspectives:

1) First, this thesis explores the use of a computer-based assessment modality to allow teachers to screen and monitor the PA development of five-year-old children in the first year of formal education.
2) Second, this thesis investigates the effect of a time-efficient and highly intensive teacher-directed classroom PA programme on raising literacy achievement and reducing inequality before the end of the first year of formal schooling.

The experimental studies reported in this thesis contribute to educational research by providing insight into how effective and efficient assessment and instructional practices in PA can be implemented by classroom teachers. Further, this thesis contributes to reading research by advancing current understanding of the role PA skills play in supporting early literacy success.

To begin addressing the need to reduce inequalities in reading outcomes, the theoretical importance of including PA as part of effective classroom reading practices must be discussed. Thus, the following section outlines relevant theories of how children learn to read, frameworks for classifying reading difficulties and instructional models that can be applied to the teaching of reading.

1.3 Theories of Reading and Phonological Awareness

1.3.1 Theories of learning to read.

Learning to read fluently and effortlessly is a complex process that relies on the development and integration of a number of linguistic and cognitive processes (Invernizzi & Hayes, 2011; Kamhi & Catts, 2012). PA is critical in the early stages of reading development through its influence on the acquisition of strong word-recognition skills (Al Otaiba et al., 2012). Such skills enable children to read fluently and to achieve the ultimate purpose of reading—the comprehension of written material (Gillon, 2004; Gough & Tunmer, 1986; Perfetti, 1985; Stanovich, Nathan, & Zolman, 1988). Understanding how children learn to read is essential for teachers who want to implement effective instructional strategies in the classroom and to support struggling readers. Thus, this section discusses five models for understanding the development of word recognition: the dual-route model, the modified dual-route model, the connectionist model, developmental models and the interactive model of word recognition. This is followed by a review of two models for classifying reading difficulties; namely, the phonological-core variable-difference model and the simple view of reading. Three approaches to reading instruction are then addressed: whole language instruction, phonics instruction and balanced literacy instruction. These theoretical models highlight the importance of PA and letter-knowledge for learning to read. They also provide a strong rationale for ensuring classroom teachers are equipped with effective and efficient methods for assessing and instructing in PA as part of classroom practice.

1.3.1.1 Dual-route model of word recognition.

One of the early models for describing how children recognise words in print is referred to as the dual-route model of word recognition. This model proposes that readers access the meaning of a printed word using either a phonological or a visual route (Coltheart, 1978; Morton & Patterson, 1980). The phonological route involves segmenting a printed word into letters or groups of letters, which are then linked to corresponding phonemes (i.e., sounds). These phonemes are then assembled (i.e., phonological recoding) to form a phonological representation of the printed word to allow access to word meaning. To use the phonological route, the reader must have an understanding of how graphemes (i.e., letters) are represented by phonemes, as well as the ability to segment and blend individual phonemes (e.g., PA) to construct accurate phonological representations. A phonics approach to reading instruction in which letter-sound relationships are prioritised supports the phonological route to word recognition (Invernizzi & Hayes, 2011). More recently, with the understanding of the role of PA in learning to read, best practice has shifted towards teaching children how to segment and blend phonemes and how to apply letter-sound correspondences in the word-recognition process (Johnson & Watson, 2005; Rose, 2006).

The phonological route proposed in this model does not allow for the fact that many English words are phonetically irregular (i.e., do not confirm to simple phoneme–grapheme matching) and, as such, an alternative visual route that is independent from the phonological route is required. The visual route proposes that the reader makes connections between a printed word's shape and orthographic representation to access word meaning. Recognising words using the visual route relies on the readers' previous experiences seeing the word and does not involve the use of phonological skills, such as PA and letter-sound knowledge (Coltheart, 1978; Coltheart, Curtis, Atkins, & Haller, 1993). Sight word teaching strategies in which children are presented with flash cards containing printed words is an example of an approach that supports the visual route. According to the dual-route model, readers use both phonological and visual routes interchangeably, depending on the type of material that is being read (Invernizzi & Hayes, 2011). For example, initially the phonological route is used to access the meaning of unfamiliar and low frequency words, but as words become increasingly familiar, they are accessed more directly using the visual route. This highlights the importance of teaching children how to use phonological information to decode words that are, in the first instance, unfamiliar, and lays a foundation for more rapid access to word meaning as skilled reading develops.

1.3.1.2 Modified dual-route model of word recognition.

Ehri (1992) proposed a modified dual-route model to acknowledge that phonological skills are also required when readers use the visual route to recognise printed words. Ehri

(1992) argued that irregular words are only partially irregular, and therefore phonological knowledge can be used to decode parts of irregular words that are regular. For example, in the word *'island'* the *'land'* component of the word is regular, the *'s'* is not pronounced, and the *'i'* is produced as the letter-name as opposed to the letter-sound. In the modified dual-route model, the phonological route is preserved but the visual route is modified to become the visual-phonological route. The visual-phonological route uses phonological skills to decode parts of irregular words that are regular, while visual cues (e.g., word shape) are used to recognise parts of the word that are irregular. That is, once a printed word is processed from the phonological and visual routes, information is combined to access word meaning. The visual-phonological route reduces memory demands because readers do not have to recall all irregular words by orthographic shape alone. Rather, a definitive number of phoneme–grapheme relationships can be used to support this process. In essence, children first learn to read a word via the phonological route, using knowledge of phoneme–grapheme correspondences and PA. With increased exposure, children begin to recognise printed words more quickly by combining visual and phonological information. This model highlights the importance of phonological information for recognising both regular and irregular words in print.

Researchers have identified several limitations to the dual-route model of word recognition. First, this theory evolved from studies of adults with sudden brain injury resulting in an acquired reading disability rather than from developmental studies involving young struggling readers (Invernizzi & Hayes, 2011). Second, investigations show that children who have trouble with word recognition have deficits in both phonological and orthographic routes, as opposed to an either-or framework as proposed by the dual-route model (Stanovich, Siegel, & Gottordo, 1997). Third, Zabell and Everatt (2002) demonstrated that adults with primary difficulties in the phonological or orthographic route did not perform significantly different to each other on a number of phonological processing tasks. Finally, researchers argue that phonological and orthographic information work in unison to support word recognition. Therefore, they cannot be viewed as two separate routes to learning to read (Vellutino, Scanlon, & Chen, 1995). This has led researchers to consider multiple connections between skill areas, rather than separate processes, as a pathway to skilled word recognition.

1.3.1.3 Connectionist model of word recognition.

A connectionist or parallel-distributed processing model of word recognition proposes that readers integrate phonological, orthographic and semantic knowledge to access word meaning (Invernizzi & Hayes, 2011; Plaut, 2007; Seidenberg, 1995; Seidenberg & McCelland, 1989). Using a computer program, researchers have modelled the roles of phonological, orthographic and semantic knowledge in reading development by imposing inhibitory or facilitator effects on each of these language areas and examining the effect this has on word-recognition ability (e.g., Harm & Seidenberg, 1999). In this model, skilled phonological knowledge is necessary to process unfamiliar words (e.g., non-words), phonetically regular words and phonetically irregular words that may be partially regular. When the computer program inhibits phonological skills, Harm and Seidenberg (1999) reported significant difficulties in the reading of non-words, irregular words and generalising phonological skills to the decoding of untaught words. This model suggests that phonological knowledge is necessary to enable beginning readers to decode new words, and that competent reading requires the integration of phonological, orthographic and semantic information.

In contrast to the dual-route model of word recognition, a connectionist model considers the interplay of several skills required in learning to read printed words, as opposed to separate processing routes (Seidenberg, 1995). Teachers need to know about the various layers of phonological, orthographic and semantic knowledge that contribute to skilled word recognition to identify which areas are in deficit. Such knowledge is critical in informing instructional strategies that can be implemented under the guidance of developmental models of word recognition.

1.3.1.4 Developmental models of word recognition.

According to developmental models, the ability to recognise printed words develops over time, in unison with increased awareness of the relationships between oral and written language (Ehri, 1991; Frith, 1985). Therefore, a deep understanding of the developmental stages or phases through which children pass on the way to skilled word recognition is essential for assessment and instruction in the classroom. Different cognitive skills and strategies predominate at each stage of development. However, strategies from previous or emerging stages may also be present (Treiman & Bourassa, 2000). Stage models of word recognition, such as those proposed by Ehri (1991) and Frith (1985), involve three key stages: logographic, alphabetic and orthographic.

During the logographic stage, children recognise whole words based on the shape of that word—just as they would recognise pictures. Phonological or phoneme–grapheme knowledge is not present. During the alphabetic stage, children learn to use PA knowledge and letter-sound correspondences to decode printed words (Ehri, 1985; Firth, 1985). A key feature of this stage is the ability to recognise words based on phonological and letter-sound knowledge. In Ehri's view (1998), children need to learn increasingly sophisticated strategies to recognise words, including segmenting words into individual phonemes, understanding letter-sound correspondences and then applying both segmentation and letter-sound skills to the decoding of printed words. During the orthographic stage, experience and exposure to written language enables children to quickly recognise morphological units within printed words (e.g., *ing*, *ed*). These units are recognised as a whole, without the need for specific letter-sound decoding. This ultimately fosters speed and efficiency in the word-recognition process.

An important component of the alphabetic stage is the development of self-teaching, which leads to independence in learning to read. According to Share's (1995) self-teaching hypothesis, previous exposure and practice at converting letters to sounds (i.e., phonological recoding) helps beginning readers to engage in self-teaching of new letter-sound correspondences. For example, a beginning reader can apply previous knowledge of letter-sound correspondences to estimate an unfamiliar word's spoken pronunciation. Initial successful decoding attempts require only a small number of repeated exposures before the word becomes more familiar and quickly recognisable. This ability to self-teach increases the number of words in a reader's orthographic lexicon, which in turn increases reading fluency and proficiency. The self-teaching hypothesis has important repercussions for reducing inequalities in reading achievement and supporting the transition to reading fluency in the orthographic stage. This is because children who approach reading instruction with a poor ability to translate letters into sounds and to construct accurate phonological representations are more limited in their ability to self-teach than are children with more proficient knowledge. This highlights the importance of ensuring that children have opportunities to develop phonological and orthographic knowledge as part of the beginning classroom reading programme.

Stage models highlight the importance of understanding how children progress towards accuracy and competency in word recognition. The importance of phonological knowledge, in particular PA and phoneme–grapheme correspondence, is critical in the alphabetic stages of learning to read. However, the purpose of reading is not to read written words in isolation. Rather, word recognition needs to be fluent within the context of connected text. Thus, theories of connected text reading are of interest to teachers because of their implications for classroom assessment and instructional practice.

1.3.1.5 Interactive model of reading.

The aforementioned models of word recognition illustrate how readers recognise printed words in isolation. However, the context afforded by connected text provides additional information to support young readers in recognising printed words. This additional information includes semantic relationships within a sentence and paragraph, context to deduce possible word meaning when a word has more than one meaning, and sentence and narrative structure (Kim & Goetz, 1994; Stanovich, 1984). Researchers have proposed an interactive model of reading (Kim & Goetz, 1994; Rumelhart, 1977; Stanovich, 1980, 1984) in which both word-level and text-level strategies support word recognition and subsequent reading comprehension. A word-level (or bottom-up) approach involves a series of gradual operations through which letters are converted into sounds, strings of which are then used to access word meaning from memory. A text-level (or top-down) approach emphasises the use of semantic and syntactic processes to access meaning. Simultaneous use and synthesis of information derived from both processes enables reading fluency and comprehension to take place. However, Share (1995) indicated that reading errors could be doubled if children rely on context alone to deduce the meaning of unfamiliar words. This highlights the importance of ensuring children have strategies to use phonological information when decoding printed words in connected text. These successful decoding attempts can then be paired with text-level information to achieve proficient reading comprehension.

Stanovich (1980, 1984) proposed an interactive-compensatory model of reading. A key feature of the original interactive model of reading is the synthesis of phonological, orthographic, syntactic, semantic and lexical processes to extract meaning from connected text (Rumelhart, 1977). Stanovich (1980, 1984) added a compensatory hypothesis to this model to account for the fact that deficiencies in any of these processes can be compensated for by another skill area. Support for this adapted model comes from evidence that poor readers with deficient decoding skills over-rely on contextual cues to recognise printed words (Catts & Kamhi, 2012). The attention required to deduce the meaning of single words using context, as opposed to phonological and orthographic skills, reduces the cognitive resources available to assimilate text-level information to achieve comprehension (Kamhi & Catts, 2012). The limitations this places on attaining skilled reading fluency and comprehension necessitates that instruction in word-level skills (e.g., PA and letter-knowledge) is included in classroom reading programmes.

1.3.1.6 Summarising theories of learning to read.

The aforementioned models of word recognition provide a template for understanding how reading develops and the importance of phonological information within this process. Difficulties in recognising printed words are a predominant feature of reading difficulties (Invernizzi & Hayes, 2011), and can severely limit reading fluency and comprehension. This is because reading fluency relies on strong word-recognition skills that become increasingly efficient through cognitive processes of automaticity and rapid naming (Fletcher, Lyon, Fuchs, & Barnes, 2007). Reading comprehension requires words to be recognised efficiently so that cognitive resources can be allocated towards the integration and comprehension of information at the text level (Stanovich, 1994). Deficiencies in phonological skills and word recognition can adversely affect learning to read, and profile strongly in models used to classify reading difficulties.

1.3.2 Models for classifying reading difficulties.

Models used to classify reading difficulties provide classroom teachers with a framework for identifying deficits that may restrict a child's reading development. This information is critical for informing curriculum design. Two models commonly reported in the literature include the phonological-core variable-difference model and the simple view of reading.

1.3.2.1 Phonological-core variable-difference model.

A traditional method of classifying reading difficulties is the phonological-core variable-difference model proposed by Stanovich (1988). In this model, children are classified as having a ‘specific’ reading disability if they demonstrate a discrepancy between average to above-average intelligence (IQ) and impaired reading performance that has an underlying phonological cause. Children with more generalised language difficulties affecting their reading development are classified as ‘garden variety’ poor readers. This latter group would often fail to demonstrate an IQ-achievement discrepancy due to global deficits in spoken language skills and a lower IQ. Researchers have challenged this method of classifying reading disability. Catts, Hogan, and Fey (2003) reported that there is very little clinical or theoretical evidence to support the use of an IQ-achievement discrepancy in the classification of reading difficulties or consequent interventions. In fact, there is meagre qualitative difference in children’s reading behaviours as a function of an IQ-discrepancy (Hatcher & Hulme, 1999; Stage, Abbott, Jenkins, & Berninger, 2003; Vellutino, Scanlon, & Lyon, 2000). Research clearly demonstrates that both children with a specific deficit in the phonological language domain (i.e., dyslexic profile) and children with more general language-based reading problems respond equally well to instruction in word recognition. This suggests that children with both broad and narrow difficulties in spoken language that affect written language development may benefit equally from classroom instruction that includes a focus on PA.

1.3.2.2 Simple view of reading.

A second approach to classifying subgroups of children with reading difficulties is the simple view of reading (Catts, Hogan, Adlof, & Barth, 2003; Gough & Tunmer, 1986; Hoover & Gough, 1990). The goal of reading is to comprehend the meaning of printed material. The simple view of reading proposes that reading comprehension is a result of word recognition ability and listening comprehension skills. This view classifies children based on difficulties in word recognition (i.e., dyslexia), deficits in listening comprehension (i.e., specific comprehension deficit) or a mixture of both (i.e., mixed reading disability). In support of this view, Hoover and Gough (1990) conducted a longitudinal study demonstrating that word recognition and listening comprehension accounted for 72 to 85 per cent of the variance in reading comprehension from first to fourth grade for bilingual speakers of English and Spanish. This result was supported by Adolf, Catts, and Little (2006), who found that in the second, fourth and eighth grades, measures of word recognition and listening comprehension collectively accounted for almost 100 per cent of the variance in reading comprehension performance.

1.3.2.3 Summarising models for classifying reading difficulties.

The phonological-core variable-difference model and the simple view of reading provide two frameworks for classifying reading difficulties. Understanding how deficits in linguistic or cognitive processes affect different parts of reading development is essential for classroom teachers. This knowledge informs teachers about how to best guide struggling readers towards reading competency throughout classroom reading instruction.

1.3.3 Models of reading instruction.

Models of reading instruction have been the centre of much debate over the last 30 years, with contention existing between two alternative approaches: whole language instruction and phonics instruction. Recently, a balance of both approaches has been recognised as important for addressing a range of literacy needs in the classroom.

1.3.3.1 Whole language instruction.

A whole language approach to reading instruction focuses on the meaning of written text and was the predominant approach to literacy instruction throughout the 1980s and 1990s (Moats, 2000; Pressley, 2006). Whole language is a constructivist approach to teaching reading and includes a number of key features such as: 1) an emphasis on immersing children in real literacy experiences, 2) child-centred instruction, 3) a focus on meaning, 4) connecting reading and writing processes and 5) the use of context, including syntactic and semantic cues, to deduce meaning. Children are encouraged to use strategies such as reading to the end of the sentence, drawing on prior knowledge, and using text structure to extract meaning (Moats, 2000; Pressley, 2006). This is in line with the top-down approach or text-level processing outlined in the interactive model of reading (see Section 1.3.1.5).

A whole language approach does not include explicit and systematic teaching of letter-sound relationships (i.e., phonics) or PA. Rather, attention is drawn to the links between letters and sounds as the need arises, and only in the context of connected text (i.e., embedded phonics) (Pressley, 2006; Walker, 2008). Underlying this approach is the assumption that children will learn to read 'naturally', as they did with spoken language, and that children learn to read independently with minimal explicit or direct instruction (Pressley, 2006; Tunmer, Chapman, & Prochnow, 2004).

A whole language approach to reading instruction has received criticism in the last decade due to a lack of focus on word-level skills. For example, Pressley (2006) highlighted two core features of this approach that lack empirical validation: 1) sentence context cues as the main strategy for decoding unfamiliar words and 2) reading as a skill that will develop 'naturally'. Further, two large-scale studies entitled the *Commission on Preventing Reading*

Difficulties in Young Children in 2008 (Snow, Burns, & Griffin, 1998) and the *National Reading Panel* in 2000 (Ehri et al., 2001) showed that explicit PA and phonics instruction (i.e., teaching children letter-sound correspondences) positively contributed to the ability to read words in isolation. Moreover, embedded phonics or no phonics instruction was associated with lower rates of reading achievement. Researchers suggest that a whole language approach may work well for children who have a strong understanding of PA and the alphabetic principle, but could be detrimental to children with inexperience in these foundation word-level skills (Tunmer, Chapman, & Prochnow, 2006). As a result, this research has contributed to a shift towards the inclusion of phonics instruction as part of the classroom reading programme in recent decades.

1.3.3.2 Phonics instruction.

A phonics-based method of reading instruction focuses on teaching children the relationship between letters and sounds, and how this knowledge can be applied to decode printed words (Walker, 2008). Phonics can be viewed as a bottom-up or word-level process through which individuals use phonological and orthographic features of written words to deduce meaning. Several phonics-based methods have been reported in the literature, including synthetic phonics, analytic phonics, analogy phonics and embedded phonics. In synthetic phonics, each letter-sound correspondence in a word is deduced and then blended together to form a word. In analytic phonics, letter-sound correspondences are taught without instruction in how to blend letter-sound links to encode or decode printed words and in which consonant blends are taught as whole units (e.g., ‘pl’ as opposed to ‘p’ and ‘l’). Analogy phonics requires children to be taught to memorise banks of phonograms (i.e., riming units such as ‘at’ or ‘ap’) and to use these analogously to read and spell unknown words. Finally, embedded phonics teaches letter-sound relationships in the context of connected text (Walker, 2008). Phonics differs from PA in that the focus is on teaching letter-sound patterns and, in the case of synthetic phonics, blending those letter-sound patterns together to decode or encode a printed word. Conversely, PA refers to an awareness of the sound structure of spoken words (i.e., no reference to letters but its teaching is paired with letter knowledge to bridge speech to print) and involves instructing children on how to identify, blend, segment, delete and manipulate syllables, rhymes and individual phonemes in words (see Section 1.3.4) (Gillon, 2004).

Support for phonics-based instruction (and PA) comes from the aforementioned reports from the *National Reading Panel* (Ehri et al., 2001) and the *Commission on Preventing Reading Difficulties in Young Children* (Snow et al., 1998). In addition, reports from the United Kingdom (e.g., *Independent Review of the Teaching of Early Reading*) (Rose, 2006), Australia (e.g., *National Inquiry into the Teaching of Literacy*) (Australian Government, 2005) and Scotland (e.g., *The Effects of Synthetic Phonics Teaching on Reading and Spelling Attainment*) (Johnson & Watson, 2005) assert the importance of using synthetic phonics as the teaching method of choice in the classroom. Although reference to the importance of PA is mentioned in these reports from the United Kingdom, Australia and Scotland; phonics teaching is given priority. This suggests that a gap exists between research evidence demonstrating the key role played by PA in learning to read and recommendations made in education reports on reading instruction.

1.3.3.3 Balanced literacy instruction.

More recently, a balanced literacy approach that aggregates the best features of whole language and phonics instruction into a core reading programme has been emphasised (Moats, 2000; Pressley, 2006). In teaching phonics, letter-sound correspondences can be taught in isolation and then applied to the decoding of unfamiliar words in connected text (Gaskin, 2011). In teaching whole language, setting aside time to instruct in letter-sound relationships removed from connected text ensures that children can devote complete attention to learning letter-sound patterns. This minimises disruptions and helps children learn that they can use letter-sound links as the primary strategy for decoding unfamiliar words, rather than relying solely, or largely, on context to decipher meaning (Gaskin, 2011). Support for both approaches comes from research on supplementary reading programmes in which Hatch et al. (2006) and Mathes et al. (2005) posit that highly structured and direct models of teaching are as effective as less structured and indirect models in raising reading achievement. Moreover, a balance of both approaches enables teachers to adapt instructional strategies to meet the individual learning needs of a diverse group of young readers in the classroom.

1.3.3.4 Summarising models of reading instruction.

Methods of teaching children to read have shifted over the last 30 years. In the 1980s and 1990s, a whole language philosophy dominated classroom reading instruction. However, over the last decade, word-level instruction using phonics has gained popularity (Johnson & Watson, 2005; Rose, 2006). More recently, a balanced approach to reading instruction that encompasses the best features of both whole language and phonics has been receiving attention as a more effective method for teaching young children to read (Pressley, 2006; Moats, 2000). Understanding different approaches to reading instruction is critical, as this allows educators to adapt instructional strategies to meet the individual learning needs of children in the classroom. This is important given that children arrive at school with differences in essential reading-related skills, and will require varying degrees of instruction in PA and letter-sound knowledge to begin reading independently (Nelson, 2010). Some children will only require a few explicit exposures to letter-sound links before 'getting' how to break the alphabetic code (Snow & Juel, 2005). Other children will require highly structured and teacher-supported guidance in learning to read (Kamhi & Catts, 2012). Children who struggle to perceive the links between speech and print intuitively in the context of whole language instruction will require direct and explicit instruction in PA and letter-knowledge (Moats, 2000; Torgesen, 2004, 2005). Thus, knowledge of different instructional approaches, and ensuring beginning-reading programmes include a balanced focus on PA and code-focused reading strategies, is critical when considering methods for reducing inequalities in reading achievement for school-aged children.

1.3.3.5 Section summary.

Understanding models of word recognition, models for classifying reading difficulties and different approaches to reading instruction provide a theoretical context against which to begin considering how best to address inequalities in reading outcomes. While learning to read is a complex process, the importance of phonological skills to word recognition and subsequent reading fluency and comprehension cannot be ignored in a classroom programme. A number of phonological processing skills are important in learning to read, of which PA has proven to be a powerful predictor of early reading success (Gillon, 2004). The next section, therefore, discusses the role of PA in reading acquisition and outlines why it is important for teachers to consider the use of this skill in the early identification and prevention of reading difficulties in the classroom.

1.3.4 Phonological awareness and learning to read.

Models of word recognition demonstrate that the acquisition of skilled reading relies on the integration of a complex tapestry of knowledge and experience (Invernizzi & Hayes, 2011; Konza, 2006). Establishing strong foundational skills that support early word-recognition abilities, such as PA and alphabetic knowledge, ensures children have the necessary building blocks to access the meaning of written information (Justice, 2006; Kamhi & Catts, 2012). Foundation skills that are weak in depth and width not only limit children's progress towards skilled reading, but also increase the risk for long-term literacy failure and academic underachievement (Bishops & Adams, 1990; Catts, Bridges, Little, & Tomblin, 2008; Conti-Ramsden & Durkin, 2007). Consequently, integrating PA into the classroom reading programme to support early word-recognition development is essential for raising reading achievement and reducing inequality.

1.3.4.1 Defining phonological awareness as a construct.

A deficit in processing the phonological component of language and applying this to written language acquisition is one potential trigger for word-recognition difficulties (Wagner & Torgesen, 1987). Phonological processing refers to the use of phonological information during oral and written language tasks (Catts, Kamhi, & Adolf, 2012). There are three types of phonological processing abilities: PA, phonological memory and phonological naming. PA involves an understanding of and ability to manipulate the sound structure of spoken words (Gillon, 2004; Kamhi & Catts, 2012; Liberman, Shankweiler, Fischer, & Carter, 1974). Phonological memory involves recoding printed information (i.e., letters) into phonological representations (i.e., sequence of phonemes) and holding this information in verbal working memory to attach meaning to a printed word (Baddeley, 1986), while phonological naming refers to the rapid retrieval of phonological representations from long-term memory (Wagner & Torgesen, 1987). Of these three kinds of phonological processing skills, PA has demonstrated the most significant relationship to early literacy success (Wagner et al., 1997).

PA can be defined as the conscious ability to attend to and manipulate the sound structure of spoken words at the syllable, rhyme and phoneme levels (Gillon, 2004). These three levels are often termed syllable awareness, rhyme awareness and phoneme awareness. Knowledge of the sound structure of spoken words contributes to early word recognition and spelling by providing a phonological map upon which the orthographic features of written language can be plotted. In the majority of cases, beginning readers must first learn how to decode printed words before they can access the meaning of written text (Pratt & Brady, 1988; Wagner & Torgesen, 1987). Once a child is able to decode printed words efficiently, a focus on comprehending a wide range of printed materials through the integration of semantic, syntactic and text-level processes can begin to take priority (Konza, 2006). Inefficient PA and alphabetic knowledge has a negative effect on the development of word-recognition skills, which has a secondary effect on reading comprehension (Catts, Kamhi, & Adlof, 2012).

A number of frameworks have been proposed to define PA. These frameworks range from a narrow definition of PA to a much broader conceptualisation of this construct. Researchers aligned with a narrow definition suggest that PA is represented by separate abilities that differ as a function of the linguistic complexity (e.g., syllables, rhyme and phonemes) and cognitive operations (e.g., identifying, blending, segmenting, deletion and manipulating) required to complete a task (Hulme, Hatcher, Nation, Brown, Adams, & Stuart, 2002; Yopp, 1988). According to Yopp (1988), phoneme awareness consists of two related factors: a simple phonemic awareness factor and a compound phonemic awareness factor. Tasks that require only one operation, such as identifying initial or final sounds or blending sounds together, represent a simple phonemic awareness factor. Tasks that require two operations where the first operation is held in the memory while the second operation is being performed, such as phoneme manipulation, represent a compound phonemic awareness factor. In line with the concept of separate abilities, Carroll, Snowling, Hulme, and Stevenson (2003) found that rhyme and phoneme awareness represented two distinct abilities, whereas Hoien, Lundberg, Stanovich, and Bjaalid (1995) reported that a syllable factor, a rhyming factor and a phoneme factor provided the best conceptualisation of PA.

Researchers aligned with a broader definition suggest that various tasks used to measure PA represent a single underlying trait (e.g., Anthony & Lonigan, 2004; Stanovich, 1992). For example, Stanovich (1992) reported that PA appears to develop along a continuum from awareness of ‘shallow’ phonological units (e.g., syllables and rhyme) to awareness of ‘deeper’ phonological units (e.g., phonemes). Similarly, Anthony and Lonigan (2004) demonstrated that tasks requiring different PA skills (e.g., syllable segmentation and blending, rhyme oddity, phoneme segmentation and blending) administered to children between two and seven years of age revealed a single underlying construct.

The view that will be applied in this thesis is as follows. It is accepted that PA consists of three levels of awareness: syllables, rhyme and phonemes. Children appear to become aware of larger phonological units before smaller phonological units, and thus awareness of syllables and rhymes is considered easier to acquire than awareness of phonemes (Liberman, Liberman, Mattingly, & Shankweiler, 1980). At each level of awareness, a number of cognitive operations such as identification, blending, segmenting, deleting and manipulation can be employed. Cognitive processes vary in difficulty, with tasks such as identifying, blending and segmenting requiring only one cognitive operation, while tasks that are more difficult, such as deletion or manipulation, require two operations (Yopp, 1988). The difficulty of a PA task can also be influenced by linguistic features of spoken words, including syllable structure (e.g., consonant and vowel shape), sound class (i.e., manner of articulation) and number of sounds (Al Otaiba et al., 2012). For example, compound words are easier to segment that non-compound words (e.g., ‘cowboy’ versus ‘happy’) (Sterling-Orth, 2004), rime units are easier to detect when the final sound is visible (e.g., ‘p**up**, c**up**’ versus ‘d**og**, f**og**’), or when the rime unit ends in a consonant rather than a vowel sound (e.g., ‘c**at**, m**at**’ versus ‘k**ey**, s**ea**’) (Schuele & Boudreau, 2008). Further, identifying, deleting and manipulating continuant sounds (e.g., m, s, n, f, v) is easier than for non-continuant sounds (e.g., k, g, t, d, p, b) (Hubbard & Mahanna-Boden, 2000; Snider, 1995). Understanding these layers of PA is critical for the measurement and instruction of PA in the classroom.

1.3.4.2 The development of phonological awareness.

There appears to be a universal sequence of PA development across alphabetic languages (Anthony & Francis, 2005; Ziegler & Goswami, 2005). In general, children become aware of larger sound units (e.g., syllables and rhyme) before they become aware of smaller sound units (e.g., phonemes) (Schuele & Boudreau, 2008). Developmental studies suggest that an awareness of syllables in words (e.g., by clapping or tapping syllables) emerges between three and four years of age, and an awareness of rhyme emerges between four and five years of age (Schreiber, 2008). Early phoneme-level skills, such as identifying the first sound in a spoken word, emerge between four and five years of age, with further development at the phoneme level occurring in unison with the onset of beginning literacy instruction (Dodd & Gillon, 2001; Lonigan, Burgess, Anthony, & Barker, 1998; Lonigan et al., 2008). Consistency in PA ability appears at around four years of age for the majority of children (Gillon, 2004). This is important for screening and monitoring to take place.

An awareness of phoneme-level skills is considered more complex to master than syllable or rhyme awareness because phonemes are acoustically more difficult to perceive (Liberman et al., 1980). Despite this, research shows that an awareness of individual phonemes in spoken words is strongly related to early reading success, over and above the ability to manipulate syllables or rhyme (Schreiber, 2008; Gillon, 2004). Muter and Snowling (1998) demonstrated this in their longitudinal study of the rhyme and phoneme-level skills of 34 British children at four, five, six and nine years of age. Measurements of rhyme detection at four, five and six years did not predict reading accuracy at nine years of age. However, measures of phoneme deletion at five and six years strongly predicted reading achievement at nine years. These findings are consistent with a body of literature suggesting that rhyming activities decline in their prognostic value as children move through the education system (Muter, Hulme, Snowling, & Taylor, 1997). Further, instruction at the phoneme level is believed to support development at the syllable and rhyme levels, whereas the reverse is less likely to occur (Brown, 1998; Yeh, 2003). Due to this relationship, classroom programmes should ideally place a high-priority focus on PA at the phoneme level.

Although PA skills appear to develop in a universal sequence, mastery at one level of awareness is not a prerequisite for development at the next level. Rather, there appears to be an overlapping continuum of development. For example, Anthony, Lonigan, Driscoll, Phillips, & Burgess (2003) conducted a hierarchical log-linear analysis to demonstrate that the PA skills of children between two and five years of age appear to emerge in what is termed a ‘quasi-parallel manner’, whereby children do not necessarily show mastery at an easier level of PA before demonstrating development at a more complex level. For example, children need not master rhyme awareness before beginning to develop awareness at the phoneme level—acquisition in these two areas of PA knowledge can overlap. This indicates that classroom programmes do not necessarily need to wait for children to master ‘shallow’ PA skills (e.g., syllable or rhyme awareness) before teaching skills at the critical phoneme level. Moreover, a specific focus on developing skills at the phoneme level may be more time-efficient than a broad focus on skills at all three levels of awareness.

1.3.4.3 The role of phonological awareness in reading development.

According to the relevant literature, the relationship between PA and learning to read can be viewed from three different perspectives (Castles & Coltheart, 2004; Elbro & Pallesen, 2002; Hatcher et al., 2006; Lukatela, Carello, Shankweiler, & Liberman, 1995; Muter, Hulme, Snowling, & Stevenson, 2004; Troia, 1999; Wagner, Torgesen, & Rashotte, 1994). The first perspective is that the amount of PA ability a child possesses affects their reading development. Evidence to support this view comes from longitudinal studies in which significant correlations between early PA ability and later reading development have been identified, and from training studies showing that exposure to PA instruction can have a significant effect on reading growth (Bradley & Bryant, 1985; Hatcher et al., 2006; Wagner et al., 1994).

The second perspective proposes that a child's PA ability develops because of learning to read. Evidence to support this view comes from research studies showing that individuals' who read a non-alphabetic script, or adults who are illiterate, have a meagre or no awareness of the sound structure of spoken words (Morais, 1991). The third perspective suggests that the relationship between PA and learning to read is bidirectional, in that early PA skills encourage the development of early word recognition, which in turn promotes the acquisition of more complex PA knowledge (Burgess & Lonigan, 1998; Castles & Coltheart, 2004; Cataldo & Ellis, 1988; Perfetti, Beck, Ball, & Hughes, 1987; Stuart & Coltheart, 1988). Each of these perspectives highlights the importance of PA in learning to read, and the importance of including PA in a beginning-reading program.

PA skills at the phoneme level, or what is often referred to as 'phoneme awareness' or 'phonemic awareness', are the most critical for the development of skilled word recognition. According to Al Otaiba et al. (2012), phoneme awareness plays three key roles in learning to read:

1) Phonemic awareness provides children with a platform to understand that the sounds in spoken words can be represented in print using alphabetic letters.
2) Phonemic awareness enhances children's ability to recognise regular phoneme–grapheme relationships, which consolidates the development of phonological representations that support word-recognition fluency.
3) Phonemic awareness helps children to decode words that are partially irregular, by sounding-out the regular phoneme–grapheme components within the word and deducing possible word meanings.

The role of PA, in particular phonemic awareness, in learning to read provides a strong rationale for researchers, educators and policy makers to ensure its inclusion as part of classroom assessment and instructional practice in the early stages of reading development.

1.3.4.4 Section summary.

Evidence of the critical relationship between PA and learning to read has been extensively demonstrated over the last 30 years through longitudinal studies, correlational research and training studies (Bishops & Adams, 1990; Bus & Van Ijzendoorn, 1999; Catts, Fey, Tomblin, & Zhang, 2002; Ehri et al., 2001; Lonigan et al., 2008). Given that PA is a powerful predictor, potential underlying cause and effective instructional approach for identifying and remediating early reading difficulties, it makes sense that efforts are directed towards ensuring its inclusion as part of core classroom practice at a national and international level, and that barriers to its implementation are reduced. Such a pursuit may support initiatives aimed at achieving equality in reading outcomes for school-aged children.

1.4 Bridging Theory to Classroom Practice

1.4.1 Identifying reading problems in the classroom: National and international needs.

Currently, researchers and educators are investigating how best to integrate research-based evidence into heterogeneous classroom environments to support literacy development (Adams, Foorman, Lundberg, & Beeler, 1997; Justice et al., 2010; O'Connor, Notari-Syverson, & Vadasy, 2005; Shapiro & Solity, 2008). This is an important area of study, given that the classroom is where the majority of student learning takes place. Further, the classroom provides an environment within which inequalities in reading achievement can be addressed. With this in mind, the intention of this doctoral research is to contribute to curriculum development at both national and international levels, as follows:

1. **Nationally:** This doctoral research will evaluate whether the integration of PA assessment and instruction into the New Zealand literacy context is able to exert a significant influence on raising reading achievement and equalising reading outcomes. The New Zealand literacy context is considered a challenging one because New Zealand produces one of the largest spread of scores between high- and low-ability readers (Martin et al., 2007) and has a predominant focus on using whole language methods for teaching reading (Connelly, Johnston, & Thompson, 2001; Tunmer & Chapman, 1999, 2002, 2004; Nicholson, 2000; Smith & Elley, 1994). Of particular importance is raising reading achievement for an indigenous Māori population (MOE, 2008).
2. **Internationally:** Researchers have called for improved screening instruments and detailed investigation into the optimal amount of classroom PA instruction in an effort to prevent reading disability in the framework of Response to Intervention (RTI) (Mastropieri & Scruggs, 2005; Scruggs & Mastropieri, 2002). This thesis will evaluate the use of computer-based test administration and the benefits of a short-duration classroom-wide and teacher-led PA programme to address these two areas of need.
3. **Practical considerations nationally and internationally:** The importance of including PA in classroom practice means that it is imperative that barriers to its adoption (e.g., time-efficiency) are addressed. This thesis will examine the effect of time-efficient assessment and instruction on reading outcomes, with the condition that these could be implemented in classrooms nationally and internationally.

1.4.1.1 Identifying reading difficulties in New Zealand classrooms: A national objective.

In New Zealand, children begin formal schooling on or as close to their fifth birthday as possible (Ministry of Education—MOE, 2012b). This means that children begin school on a rolling-basis throughout the academic year, as opposed to starting at pre-determined intake dates. Most parents select this option even though they are not legally required to have their child or children enrolled at a school until six years of age (MOE, 2012b). *The New Zealand Curriculum* is the overarching document that governs the national education system (MOE, 2012b). Within this document, there are eight learning areas: English, the arts, health and physical education, learning languages, mathematics and statistics, social science and technology. English is the learning area within which oral and written language learning objectives are detailed. At school-entry (i.e., Level 1), English learning objectives focus on teaching children the relationships between oral, written and visual language. Objectives aim to ensure children can integrate many sources of information (e.g., semantics, syntax, visual and letter-sound relationships) to comprehend a variety of texts. Understanding the relationships between letters and sounds (i.e., phonics), the ability to use comprehension strategies (i.e., whole language) and learning to self-monitor and think critically about printed materials is emphasised. Specific reference to teaching PA, in particular phoneme awareness (i.e., identifying, blending, deleting, segmenting and manipulating sounds in words), is not documented as a key learning objective (MOE, 2012b).

Classroom assessment: A significant barrier to widespread PA screening and monitoring in New Zealand classrooms is the lack of country-specific PA instruments or a database of developmental norms. Instruments available in New Zealand schools to assess reading-related skills either do not assess PA or provide information on only shallow levels of PA (e.g., syllable or rhyme awareness). For example, the School-Entry Assessment (SEA) is one of the most commonly reported instruments used to gather baseline information on the early literacy, numeracy and oral language abilities of children as they begin formal schooling (Dewar & Telford, 2003). The SEA does not assess PA ability, and as indicated by a recent review, ‘the range of literacy and numeracy skills addressed by the SEA is not as comprehensive as the range of skills that recent research suggests are critical, in the early years of schooling’ (Anderson, Lindsey, Schulz, Monseur, & Meiers, 2004, p. 1).

The SEA is often benchmarked against the *Observational Survey of Early Literacy* (OS) (Clay, 2002), which is administered at six years of age to gauge knowledge of letter recognition, concepts about print, vocabulary and text reading (Clay, 2002). Within the OS, a dictation task called ‘Hearing and Recording Sounds in Words’ measures phoneme awareness by having children write words and sentences that an examiner speaks aloud. This is not consistent with commonly reported tasks used to measure PA. Research into the predictive validity of the SEA is sparse (Anderson et al., 2004), and research into the utility of the OS indicates inadequate floor and ceiling effects, with further work required in setting benchmarks for the predictive validity of the instrument (Denton, Ciancio, & Fletcher, 2006).

The last instrument discussed here, the *Performance Indicators in Primary School On-Entry Baseline Assessment* (PIPS) (Tymms, 1999), is a computerised assessment used in some New Zealand primary schools to measure early reading, mathematics, PA, social and emotional development, behaviour and attitudes as children enter formal education. The PIPS includes a task that evaluates rhyme awareness, but does not include evaluation at the critical phoneme level (Tymms, 1999).

The professional group most likely to assess PA are speech-language therapists. In a recent survey investigating the types of assessment instruments used by New Zealand speech-language therapists, it was identified that there was a lack of PA instruments designed and standardised for New Zealand school-aged children (Klee & Tillard, 2010). Given the lack of normative data and assessment instruments to screen and monitor PA development in New Zealand school-aged children, it appears necessary for researchers and educators to collaborate to construct methods for effectively identifying children at risk for reading problems before they fall behind in classroom reading instruction. This area of need will be addressed in this thesis.

Classroom reading instruction: In New Zealand classrooms, a whole language approach to reading instruction was the instructional method of choice throughout the 1980s and 1990s (see Section 1.3.3.1). Although this approach continues to dominate classroom practice, a shift towards a more balanced approach that includes the use of phonics instruction is becoming more widely accepted (Connelly et al., 2001; Tunmer & Chapman, 1999, 2002, 2004; Nicholson, 2000; Smith & Elley, 1994). However, the explicit and systematic teaching of PA is still rare. Research has shown that PA and letter-knowledge provide the best indication of how well children will learn to read during their first years at school, and that the pairing of these two skill areas produces the greatest effect on elevating reading achievement (Ehri et al., 2001). This thesis will provide evidence that the inclusion of teacher-led PA instruction as part of a classroom programme that already includes phonics can have a much greater influence on raising achievement and reducing reading inequalities than do current classroom practices.

Early identification methods: In New Zealand, a nationally applied early intervention initiative to reduce the prevalence of reading difficulties is the Reading Recovery programme (Clay, 2002, 2003). Reading Recovery has been in operation since the mid-1980s; it is offered to six-year-old children performing in the bottom 10–20 per cent of readers after one year of formal schooling. Reading Recovery employs similar strategies to whole language instruction, but in a more intensive one-to-one format lasting up to 20 weeks. There is research evidence to support the use of Reading Recovery (Clay, 2002, 2003).

However, in some respects, Reading Recovery can be viewed as a ‘wait and see’ approach that allows children at risk for reading disability to struggle for up to one year of schooling prior to receiving supplementary support. Over the last decade, New Zealand has consistently presented with one of the largest gaps in the performance of good and poor readers (Martin et al., 2007), suggesting that existing classroom curriculums and the supplementary Reading Recovery programme require adaptation to meet the needs of New Zealand’s most vulnerable school-aged readers. This thesis aims to address this need by investigating the benefits of integrating PA into existing classroom programmes.

Raising reading achievement for Māori students: Internationally, research indicates that children who belong to indigenous or minority groups are often poorer readers compared to children who come from the majority group within a country’s education system (Green, 2001; Haycock, 2001; Hedges & Nowell, 1999; Jeynes, 2007). The New Zealand education system is distinguished from schooling systems worldwide through its protection of indigenous Māori culture, values and languages and its expanding multi-cultural student-base. An estimated 30, 000 New Zealand children access the national curriculum to varying degrees through Māori-medium educational settings such as Kura Kapa Māori (full immersion), bilingual schools (partial immersion), schools with full immersion classes, or schools with bilingual classes (MOE, 2008). Inequalities in school entry PA knowledge and subsequent reading and spelling performance between Māori and Pakeha (i.e., of European descent) students have been reported. For example, Tunmer, Chapman, & Prochnow (2002) revealed that the phonological processing skills, including PA (e.g., onset-rime segmentation, sound matching, phoneme segmentation), that young Māori children bring to beginning literacy instruction are significantly inferior to those of young non-Māori children. Further, National Certificate of Educational Achievement (NCEA) results demonstrate that significantly fewer Māori students (65 per cent) achieve the required benchmark for literacy standards by 15-years of age compared to their non-Māori peers (80 per cent) (MOE, 2008). This suggests that early differences in reading and spelling abilities at school entry may well persist into the adolescent years.

The New Zealand Ministry of Education strategy to raise Māori achievement is entitled, *'Ka Hikitia - Managing for Success: The Māori Education Strategy 2008 – 2012'* (MOE, 2012a). This strategy focuses on four areas in which the Ministry of Education aims to improve the education system for and with the Māori population. These areas include:

1) the *'foundation years'* in which developing a strong foundation for learning during the early childhood and schooling years is emphasised,
2) *'young people engaged in learning'* where the focus is on ensuring 14- to 18-year-old Māori students are achieving their learning potential,
3) *'Māori language in education'* which highlights the importance of Te Reo Māori in mainstream and Māori-immersion education settings as an official language of New Zealand, and
4) *'organisational success'* where the roles of the New Zealand Ministry of Education and the national education sector at large are highlighted with regards to raising educational performance for Māori.

The research reported in this thesis will address the area of the *'foundation years'* in the *'Ka Hikitia - Managing for Success: The Māori Education Strategy'* (MOE, 2012a) by providing an example of how classroom reading instruction that includes a focus on PA can successfully raise the literacy achievement for a group of school-aged children that is inclusive of a Māori cohort.

1.4.1.2 Enhancing early identification of reading difficulties: An international objective.

In the field of education, RTI is one framework within which the experiments in this thesis can be viewed. RTI aims to prevent academic failure through research-based classroom instruction, regular screening and monitoring of student progress, and provision of increasingly intensive supplementary levels of support for children experiencing difficulty (Ehren & Nelson, 2005; Griffiths, Parson, Burns, Van Der Heyden, & Tilly, 2007). Underlying RTI is the use of scientifically based instructional strategies and approaches that have proven effective in randomised controlled trials (Nelson, 2010). A key objective of RTI is to ensure all children are provided with a scientifically driven core classroom curriculum, and that children at risk for academic underachievement are identified and supported in a timely fashion. The assumptions of RTI include (Griffiths et al., 2007; Nelson, 2010):

1) All children can be taught effectively in an education system.
2) Early intervention is important for preventing difficulties becoming unnecessarily severe.
3) Scientific evidence should drive instruction and intervention as much as possible.
4) Successive levels of support should be provided for children not showing progress as expected.
5) Monitoring student progress is essential for informing instructional practice.

RTI is intended to be a supplement to the ‘usual’ classroom curriculum, as opposed to a replacement, and is designed to be implemented at three levels (or tiers) of support. Tier 1—the first level of support—specifically focuses on ensuring all children in the classroom receive scientifically based instruction. Screening and progress monitoring evaluates academic growth over time and identifies children who are not progressing as expected. Screening usually takes place three times a year using criterion referenced or normreferenced measures, whereas progress monitoring is more frequent and may involve using Curriculum-Based Measurement. In Tier 1, it is estimated that 80 to 85 per cent of children will perform at an age-expected level. Children not progressing as expected are offered Tier 2 support. This involves small group instruction inside or outside the classroom environment that may be implemented for 30-minute sessions two to four times a week for at least nine weeks. This support is offered in conjunction with Tier 1 support. Frequent progress monitoring at Tier 2 enables educators to determine which children have responded to supplementary instruction and can return to Tier1, and which children will require further support at Tier 3. It is estimated that approximately three to six per cent of children will require support at Tier 3. This tier involves intensive one-to-one instruction, probably over two 30-minute sessions each school day for nine to 12 weeks. Tier 3 is often seen as 'special education'. This framework is proactive and aims to identify children at risk before they fall behind in academic achievement (Griffiths et al., 2007; Nelson, 2010).

The RTI framework is primarily used in the United States of America. However, its application to reading education in classrooms globally may provide a framework to help reduce inequalities in literacy achievement. Researchers contend that for the RTI framework to be more effective, improved screening instruments need to be made available to educators. In addition, a more in-depth understanding regarding the duration over which students should receive instruction at each tier of support is required (Mastropieri & Scruggs, 2005; Scruggs & Mastropieri, 2002). This thesis contributes to these areas of need in two complementary ways. First, by designing, trialling and evaluating the use of a time-efficient and user-friendly computer-based assessment to screen and monitor PA development and qualify risk for reading disorder in the classroom, teachers will be assisted in identifying which children may require advanced support at Tier 2. Second, by examining and drawing conclusions about the duration and intensity of classroom phonological training required to raise achievement and reduce disparities in reading skill acquisition, an optimal combination will be proposed.

1.4.1.3 Practical considerations for time-efficient measurement and instruction in the classroom.

Ensuring teachers can integrate PA assessment and instructional strategies easily and effectively into the classroom requires consideration of practicalities such as time. Attempts to bridge approaches that have proven effective under controlled research settings into the heterogeneous classroom environment may be unsuccessful if teachers perceive them as too time consuming to implement. McLeod et al. (2003) proposed three key elements that are critical for managing the day-to-day running of a classroom. These elements are: 1) managing time and space, 2) managing student behaviour and 3) aligning instructional strategies to curriculum content and student needs. Among these key elements, McLeod et al. (2003) state that:

> time constraints make a critical difference in whether a strategy can be used successfully. In situations where there are large numbers of curriculum objectives to cover in a short time span, teachers may have to forgo strategies that are time-consuming (p. 129).

In addition to quantity of teacher time, measuring PA during the timeframe within which it is most predictive of reading outcomes and teaching PA early so that children can take advantage of beginning-reading instruction, are two further methods of viewing time-efficiency. In general, the importance of time-efficiency can be viewed from three perspectives:

1) **Quantity of assessment and instructional time:** Given the large number of competing priorities in the classroom, inclusion of PA in the curriculum should

ideally be time-efficient to achieve manageability for teachers. Time-efficiency is particularly important because assessment procedures that are time consuming may go unused by educators. In a comprehensive review of the SEA (Dewar & Telford, 2003), the New Zealand Ministry of Education identified that 66.7 per cent of new entrant teachers (i.e., teachers of children in the first year of formal schooling) that had previously used the SEA no longer did so because it was considered too time consuming to administer. Further, approximately 10 per cent of respondents stated that they did not have enough time to assess every child in the classroom due to other classroom or school commitments. Internationally, it has been estimated that assessment activities can consume approximately one third of a typical teacher's classroom time (McTighe & Ferrara, 1998). Therefore, to ensure key skills, such as PA, are included in the classroom programme, consideration must be given to the time-efficiency of different approaches to assessment and instruction, with the aim of alleviating demand on teacher time.

2) **Using PA when it is most predictive of reading outcomes:** Another way of promoting time-efficiency is by ensuring that PA is measured when at its most powerful as a predictor of later reading outcomes. According to the literature, the optimal time for using PA as a predictor of reading success is during the early stages of reading education (Al Otaiba et al., 2012; Torgesen, Wagner, & Rashotte, 1997). In a longitudinal study, Catts and Hogan (2002) identified that the predictive relationship between PA and reading altered from kindergarten through to second and fourth grades. Results showed that measures of phonemic awareness in kindergarten provided predictive information, beyond that accounted for by word-decoding measures, regarding how well children would perform in reading by the second grade. However, when measured in the second grade,

phonemic awareness added little additional prognostic information regarding future reading outcomes beyond that already provided by measures of second grade word decoding. Similarly, Torgesen and colleagues (1997) identified that phonemic awareness tasks, administered to children in the second- and third-grade (i.e., approximately seven to nine years of age) added little precision to the identification of reading difficulties over and above that already identified by reading measures. Therefore, ensuring that measurement of PA occurs during the early stages of reading education is important for capitalising on its predictive power.

3) **Ensuring PA is in place to take advantage of reading instruction:** Time-efficiency can also be enhanced by ensuring young children have the necessary foundational skills in place to make the most of beginning classroom reading instruction. If children know how to apply PA knowledge to learning to read, inequalities in reading may be reduced because they will:
    a) Have the skills to support early word recognition (Ehri et al., 2001).
    b) Be able to use knowledge of letter-sound patterns to begin self-teaching (Share, 1995).
    c) Be able take advantage of a bidirectional relationship between phoneme awareness and reading, in which improvement in one area leads to improvement in the other (Carroll & Snowling, 2004).
    d) Have a reduced chance of falling victim to the ‘Matthew Effect’, which shows that children rich in phoneme awareness and letter-knowledge become increasingly stronger readers and children with poorer knowledge become comparatively weaker readers (Morgan et al., 2008).

e) Be less likely to develop a negative self-perception as a reader, which would contribute to fewer attempts at interacting with print (Chapman & Tunmer, 1997).

Given the importance of PA for learning to read, it appears essential that researchers consider time-efficiency when designing and investigating assessment and instruction strategies. This will help ensure the successful implementation of PA in the classroom by everyday teachers.

1.4.1.4 Section summary.

Bridging the gap between theoretical perspectives on reading and classroom practice may support initiatives aimed at reducing inequalities in reading outcomes at national and international levels. Ensuring PA assessment and instructional strategies are implementable by time-poor teachers is an important practical consideration that must be addressed if widespread adoption is to occur. This thesis addresses this need by examining time-efficient methods of assessing and instructing children in PA during the first year of formal schooling.

1.4.2 Classroom phonological awareness assessment.

A number of instruments are available to teachers to measure the PA ability of children in their classrooms. These instruments vary in their intended purposes, ranging from the screening and monitoring of PA development, to the provision of diagnostic information on phonological deficit and reading disability. In their review of instruments commonly used to measure PA, Sodoro, Allinder and Rankin-Erickson (2002) listed the following: the *Comprehensive Test of Phonological Processing* (CTOPP) (Wagner, Torgesen, & Rashotte, 1999), the *Lindamood Auditory Conceptualization Test, Revised Edition* (LAC) (Lindamood & Lindamood, 1979), the *Test of Phonological Awareness* (TOPA) (Torgesen & Bryant, 1994), the *Phonological Awareness Profile* (Robertson & Salter, 1995) and the *Yopp-Singer Test of Phoneme Segmentation* (Yopp, 1995). Other examples of commonly reported measures include the *Dynamic Indicators of Basic Early Literacy Skills* (DIBELS) (Good & Kaminski, 2005), the *Phonological Awareness Literacy Screening-Kindergarten* (PALS-K) (Invernizzi, Juel, Meier, & Swank, 2005) and the *Preschool and Primary Inventory of Phonological Awareness* (PIPA) (Dodd, Crosbie, MacIntosh, Teitzel, & Ozanne, 2000).

Table 1.2 profiles a selection of instruments currently available to classroom teachers that measure the PA abilities of five-year-old children. This is not intended as an exhaustive list. These tools are compared based on the *administration time* required for the teacher and child (i.e., the time taken to administer all tasks within an assessment tool), the *modality* of administration (i.e., whether it is paper-based and requires the presence of a teacher, or whether it is to be delivered by computer) and the *content* of the assessment. In terms of content, assessments that have a high-priority focus on PA at the phoneme level and evaluate how this knowledge is transfered to print (e.g., letter-knowledge or non-word decoding) are considered narrow, whereas assessments that focus on a broader range of phonological abilities (e.g., syllable awareness, phonological memory or rapid automatic naming) or additional language skills (e.g., vocabulary) are considered to be broad. As the focus of this section is the time-efficiency of assessment, Table 1.2 includes a hypothetical example of the amount of teacher time required to assess all children in a classroom size of 12 using these assessments.

Table 1.2

Time-efficiency, administration modality and content of a selection of phonological awareness assessments available to classroom teachers

	Administration Time (in minutes)		Modality		Content		Classroom Example
	Teacher	*Child*	*Paper*	*Computer*	*Broad*	*Narrow*	*Teacher time required for a class of 12*
CTOPP*	30	30	+		+		6 hours
LAC*	20–30	20–30	+		+		4–6 hours
TOPA-2+ **	30–45	30–45	+			+	6–9 hours
DIBELS	7	7	+		+		1.5 hours
PALS-K**	20	20	+		+		4 hours
PIPA	25–30	25–30	+		+		5–6 hours
PA Profile	10–20	10–20			+		2–4 hours
Yopp-Singer	5–10	5–10	+			+	1–2 hours

Note. + indicates the modality and content within each assessment; * indicates that a specialised qualification or form of training is required for administration of the assessment; ** indicates that the assessment can be administered on an individual basis or to a small group of children; CTOPP = Comprehensive Test of Phonological Processing (Wagner et al., 1999); LAC = Lindamood Auditory Conceptualisation Test, Revised Edition (Lindamood & Lindamood, 1979); TOPA- 2+ = Test of PA, Second Edition: PLUS (Torgesen & Bryant, 2004); DIBELS = Dynamic Indicators of Basic Early Literacy Skills (Good & Kiminski, 2003); PALS-K = PA Literacy Screening—Kindergarten (Invernizzi et al., 2005); PIPA = Preschool and Primary Inventory of PA (Dodd et al., 2000); PA Profile = PA Profile (Robertson & Salter, 1995); Yopp-Singer = Yopp-Singer Test of Phoneme Segmentation (Yopp, 1995).

The breath of content in an assessment and the modality used to administer that content to children are two design features that influence time-efficiency. These are also reported in Table 1.2. According to the literature, a multivariate approach that combines PA with other important reading skills has proven to be the most accurate in predicting reading outcomes, but in some cases, this may not be time-efficient when considering the needs of the classroom (Compton, Fuchs, Fuchs, & Bryant, 2006; Fuchs, Fuchs, & Compton, 2004; O'Connor & Jenkins, 1999). In Table 1.2, tools that are multivariate (e.g., target vocabulary or reading comprehension in addition to PA) or sample a broad range of PA skills (e.g., syllables, rhymes and phonemes) include the CTOPP, DIBELS, LAC, PA Profile and PIPA. The CTOPP, LAC and PIPA are primarily used for diagnostic purposes and range from 20 to 30 minutes of administration time for teachers and children. In a class of 10 to 15 children, a teacher would need to set aside three hours and 20 minutes to seven hours and 30 minutes for administration time alone. This does not include the time required to set up the assessments, score responses and interpret results. The DIBELS, PALS-K and PA Profile tools assess PA and a range of related skill, often in the form of screening and monitoring instruments. The DIBELS is the most time-efficient instrument profiled in Table 1.2 and involves seven one-minute tasks that measure phoneme awareness, alphabet knowledge, reading accuracy and fluency, vocabulary and reading comprehension. Although time-efficient for teachers, research demonstrates that DIBELS measures can produce floor effects (i.e., when the majority of children score zero correct) when administered to children five years of age, consequently reducing the predictive validity of this screening tool at this stage (Catts, Petscher, Schatschneider, Bridges, & Mendoza, 2009).

The adoption of a univariate approach in which one skill or a small combination of skills is measured is becoming increasingly popular for time-efficient screening and monitoring in the classroom (Kamhi & Catts, 2012). A key consideration when designing univariate measures is the tendency to obtain a high rate of false positives. This is because tools that measure only one or a small number of skills tend to over-identify the number of children who are at risk for reading difficulties (Blachman et al., 2000). This places an added strain on school and classroom resources because children are provided with additional support although they may not necessarily have required that support. Researchers suggest that high false-positive rates can be reduced through dynamic assessment or repeated administrations throughout the early school years (Good, Simmons, Kame'euni, Kaminski, & Wallin, 2002; Kamhi & Catts, 2012; Good, Simmons, & Kame'enui, 2001).

According to the National Reading Panel (Ehri et al., 2001), two measures that provide the best indication as to how well children will learn to read and spell during their formative years of schooling include phoneme awareness and letter-knowledge. The Yopp-Singer Test of Phoneme Segmentation reported in Table 1.2 provides this 'narrow' focus on phoneme awareness (i.e., univariate) but does not include a measure of letter-knowledge. Further, the TOPA-2+ provides a measure of initial and final phoneme identity, letter-knowledge and non-word spelling, but does not include the phoneme-level skills reported to provide the best indication of early reading outcomes; namely, phoneme blending and phoneme segmentation (Schuele & Boudreau, 2008). However, it is important to note that the TOPA can be administered in small groups, positively affecting administration time-efficiency.

In review of PA tools currently available to teachers, there appears to be scope for the development of a 'narrow' assessment tool with a high-priority focus on assessing and monitoring critical phoneme-level ability and letter-knowledge in the classroom. Focusing on this small combination of skills is one way of increasing the time-efficiency of assessment, which may be further enhanced through use of a computer-based administration modality. Although it is important for classroom teachers to assess a wide range of reading constructs, a time-efficient tool specifically focusing on key predictors of literacy success, and which does not require the teacher for its administration (e.g., uses computers), provides an alternative option for ensuring all children are monitored when time constraints in the classroom are an issue. This thesis aims to address this need by developing (see Chapter 2), applying (see Chapters 4 and 5) and establishing the validity and reliability (see Chapter 6) of a computer-administered classroom PA tool.

1.4.2.1 Computer-based phonological awareness assessment to increase time-efficiency.

Computer-based assessment (CBA) offers an excellent opportunity for teachers to overcome time-related barriers to the routine assessment and monitoring of PA in the classroom. CBA is defined as ‘any psychological assessment that involves the use of digital technology to collect, process and report the results of that assessment’ (British Psychology Society - BPS, 1999, p. 11). CBA can be further defined as non-adaptive or adaptive. Non-adaptive assessments present test items in a sequential manner, and are often the first step towards the construction of an adaptive CBA. Adaptive assessments adjust to the ability of the student by presenting test items that are within the student’s zone of development (Eggen & Straetmans, 2009). Adaptive CBA requires large item banks and extensive psychometric testing (Tymms, 2001).

A commonality among the instruments reported in Table 1.2 is the use of paper-based testing procedures whereby the teacher must set aside time to administer test stimuli, record responses and interpret results for each child in the classroom. Research indicates that computer-based administration of assessment content offers several advantages in terms of time-efficiency when compared to paper-based procedures (Martin, 2008; Tymms, 2001). This is because the computer can 1) present all test items (e.g., pictures and verbal instructions), 2) allow children to respond by using a computer mouse or touch screen

(Singleton, Horne, & Thomas, 1999) and 3) score and interpret responses into a database that is readily available to teachers (Bjornsson, 2008; Martin, 2008; Ripley, 2008). These features reduce the amount of time that teachers are required to be directly involved in the administration of the assessment, as well as the time required to set up assessment materials (e.g., paper cards), record responses and interpret the results that will inform classroom instruction (Bridgeman, 2009). Olsen (1990) provided evidence of time-efficiency when comparing the time taken to administer a paper-based school achievement test in comparison to a computer-based version of the same assessment. Results showed that a non-adaptive CBA took 25 to 50 per cent less time to administer in comparison to the paper-based administration format. Time-efficient computer-based measurement that places little demand on teacher time may therefore ensure that children at risk for reading disability are identified early and that inequalities in reading outcomes are reduced.

Additional benefits of using CBA in the classroom include: 1) increased consistency in presenting test items and improved objectivity of assessment, 2) use of technology already in place in schools and the classroom, 3) minimal training needed for classroom teachers and paraprofessionals, 4) low cost once established, due to the removal of ongoing expenses associated with purchasing test booklets and 5) the provision of incentives and motivation for children to participate, which is particularly important for ensuring the reliability of results and accurately identifying those children at risk of developing reading difficulties (Singleton, Thomas, & Leedale, 1996; Tymms, 2001).

CBA is becoming increasingly common in the field of education, and it is important to mention two fully computerised psychometric assessment systems that include a section on PA assessment. The *Cognitive Profiling System* (CoPS) (Singleton, Thomas, & Leedale, 1996; see http://www.lucid-research.com/sales/esales.htm?category_id=31&product_id=181) is a computer assessment administered to children as they begin school to identify risk for learning difficulties, including dyslexia. The assessment measures visual, spatial, verbal and auditory sequential memory, auditory-visual associative memory, visual-verbal associative learning, PA, auditory discrimination and colour discrimination. PA is measured using a rhyme-matching task and an initial phoneme-matching task. Similarly, PIPS (Tymms, 1999; also see http://www.cemcentre.org/pips/pips) is a fully computerised adaptive on-entry baseline assessment that measures the areas of reading and mathematics. PA is measured using a rhyming game. Although CoPS and PIPS are computer-based, the information they provide on PA is largely shallow (e.g., rhyming). In-depth detail on phoneme-level skills such as phoneme blending and phoneme segmentation is not collected. Thus, scope exists for the investigation of CBAs that specifically focus on phoneme-level skills, and are user-friendly and time-efficient for classroom teachers.

An integral step towards validating the use of CBA of PA in the classroom is to ensure that computer administration does not advantage or disadvantage performance. That is, teachers must be confident that a computer-administrated version of an assessment will produce comparable results to a paper-based counterpart (Csapo, Molnar, & Toth, 2009; Singleton, 2007). A variety of studies, ranging from meta-analyses of reading and mathematical ability (Wang, Jiao, Young, Brooks, & Olsen, 2007, 2008), to both larger and smaller scale studies of student achievement (Csapo et al., 2009; Maguire, Knobel, Knobel, & Sedlacek, 1991), generally find that student's scores are not significantly affected by the modality of assessment. For example, Maguire et al. (1991) found a high correlation (r = 0.76) and no significant difference between the results participants obtained on a computer-administrated version of the *Peabody Picture Vocabulary Test—Revised* and the paper-based counterpart. Similarly, Wilson, Thompson, and Wylie (1982) found no significant difference between a computerised and conventionally administered version of the *Mill Hill Vocabulary Test*. Researchers have also used methods of concurrent or predictive validity to help establish equivalency for tests that have been developed specifically for the computer and thus have no conventional counterpart. For example, Singleton, Horne, and Vincent (1995) demonstrated a strong correlation between the conventionally administered *Edinburgh Reading Test* and a trial edition of a computer-based reading comprehension test. Obtaining congruent results between a computer-administered and paper-based assessment enables educators to use CBA reliably as a time-efficient tool in the classroom (Tymms, 2001).

1.4.2.2 Psychometric considerations for computer-based assessment.

How to ensure that the technical adequacy of a screening and monitoring tool is preserved while achieving time-efficient measurement in the classroom is an important issue to address. To be confidently used in the classroom, CBA must present with adequate psychometric quality and sensitivity for the prediction of reading difficulty. Issues associated with the assessment of young children in general and those associated with the use of a computer-based test modality with a young cohort can influence the psychometric quality and sensitivity of an instrument.

Psychometric quality: Evidence suggests that the earlier an assessment tool is administered to young children, the more difficult it is to obtain reliable and valid results. This is because assessments for younger children are often shorter to help maintain interest and avoid fatigue. Specific issues relating to using the computer with young children must also be considered. For example, Barnes (2010) compared computer- and paper-based testing of the rhyme awareness subscale from the PALS assessment. Results showed that 12 per cent of kindergarten children (i.e., five years of age) required support to use the computer mouse or to follow instructions, whereas almost half of preschool children (i.e., under five years of age) required assistance. An administration-modality effect was identified for preschool children, in that paper-based testing was easier than CBA. This suggests that CBA for children under five years of age is less reliable than for children beginning formal schooling.

Although this thesis targets children who are five to six years of age, research by Barnes (2010) highlights that CBA may lack psychometric quality when used with young children. Therefore, establishing and reporting the psychometric calibre of the computer-based PA tool developed in this thesis is critical (see Chapter 6).

Sensitivity: To be effective, classroom assessment tools must be able to differentiate those children who are likely to become good readers from those children who are likely to become poor readers (Justice, Invernizzi, & Meier, 2002). Measurement tools that do not accurately identify children at risk for reading difficulties are ineffective in that they allow beginning readers to participate in classroom instruction with an unidentified risk for literacy impairment. Sensitivity refers to the percentage of poor readers correctly identified, while specificity refers to the percentage of good readers correctly identified (Spitalnic, 2004). High sensitivity is important in the classroom setting to ensure children at risk for reading disability are provided with adequate support (O'Connor & Jenkins, 1999; Scanlon & Vellutino, 1996). However, due to the desire to identify all children at risk for reading problems, measurement tools may produce a high number of false positives (Dickman, 2006). This means that children who are not at risk are identified as being at risk. As mentioned, this is problematic in that these children are provided with support that they do not necessarily require. The CBA designed in this thesis is used as a screening and monitoring tool, thus allowing false positives to be reduced over repeated administrations.

1.4.2.3 Section summary.

Computer administration alleviates demands on teacher time, thereby allowing educators to focus on how results can inform effective instructional practice. Such efficiency may contribute to the early identification of reading disorders and the minimisation of inequalities in literacy achievement in the classroom. This thesis investigates the benefits of computer-based PA assessment for the early identification of reading problems and the reduction of inequalities in reading outcomes from test development (see Chapter 2), to classroom application (see Chapter 4 and 5) and finally to evidence of test validity and reliability (see Chapter 6).

1.4.3 Classroom phonological awareness instruction.

In the early school years, classroom time is largely spent on developing literacy-based skills (Johnson & Watson, 2005; Rose, 2006). Beginning-reading instruction involves teaching children a number of important skills that will lay the foundations for future learning. It is widely recognised that PA is one critical component to mediating successful or unsuccessful reading outcomes (Carroll & Snowling, 2004; Catts et al., 2001; Ehri et al., 2001; Lonigan et al., 2008). Scientific evidence surrounding the benefits of PA instruction for literacy growth is well reported for children with typical development, as well as for children with elevated risk for reading disorder. However, the majority of PA intervention studies have been conducted under controlled research settings using small group or individual models of service delivery outside the classroom environment (Brady et al., 1994; Ehri et al., 2001; Gillon, 2000a, 2005; Gillon & McNeill, 2009). Therefore, less is known about the effectiveness of PA instruction—in particular, the optimal duration and intensity of instruction—when implemented in the heterogeneous environment of the mainstream classroom. This thesis addresses this gap.

1.4.3.1 Optimal duration and intensity of classroom phonological awareness instruction.

Developing programmes to be delivered by teachers to children in the classroom requires consideration and collaboration at a number of different levels if successful implementation is to occur. As indicated in Section 1.4.1.3, one key consideration is time-efficiency. Time-efficiency is important for ensuring programmes are easily integrated into an existing classroom schedule and for ensuring children have the necessary skills in place to take advantage of reading instruction. According to the literature, frequent and intensive sessions are an important component of effective PA instruction (Elbaum, Vaughn, Hughes, & Moody, 1999; Gillon, 2004). In controlled clinical settings, explicit and systematic instruction involving two one-hour individual sessions per week is considered high in intensity (i.e., 2 hours per week). Instruction of this intensity, focused at the phoneme level for 20 hours over a 10-week period, has proven sufficient for raising reading achievement in at-risk populations in individualised and small group therapy settings (Gillon, 2000a, 2005; Gillon & McNeill, 2009). Investigating whether this duration and intensity of instruction can be replicated in a classroom environment will provide valuable information regarding the amount of time teachers should devote to PA to help reduce reading inequality.

Research into the effectiveness of classroom-based literacy programmes that include a focus on PA have varied in the duration and intensity of teacher-directed instruction. In this thesis, the literature was reviwed to identify research programmes that included a focus on PA and were implemented by teachers in the classroom. Four studies met these criteria and were compared using the following classifications:

1) *Duration*: programmes implemented for more than one academic year (i.e., greater than 36 weeks) are considered long in duration, and programmes implemented for less than one academic year (i.e., less than 36 weeks) are considered short.

2) *Intensity*: programmes involving two hours or more of instruction per week are viewed as highly intensive, while programmes involving less than two hours of instruction per week are considered low in intensity. A cut-off of two hours per week was selected based on evidence that this intensity of PA instruction over 10 weeks is sufficient for raising reading achievement in children at-risk (Gillon,

2000, 2005), whereas less than 10 hours of instruction has proven less effective in accelerating reading development (Gillon & Dodd, 1997).

An additional area of variability among studies relates to the *content* of PA instruction. To compare content of instruction, studies that target PA at the phoneme-level (i.e., developing awareness of individual sounds in words) were classified as narrow, and studies targeting a wide range of PA skills (e.g., syllables, rhyme and phonemes) were considered broad. Table 1.3 compares the duration, intensity and content of PA instruction on reading outcomes from four recent investigations. The studies in this table do not represent an exhaustive list. Rather, the table presents a compilation of recent investigations in which teachers have acted as programme implementers in the classroom.

Table 1.3

Duration, intensity and content of classroom phonological awareness instruction on reading outcomes

	Duration		Intensity		Content		Reading Outcomes	
	Long	Short	High	Low	Broad	Narrow	Immediate	Sustained
Shapiro & Solity (2008)	+		+			+	+	+
McIntosh et al. (2007)		+	+		+		+	
Fuchs et al. (2001)		+		+	+		+	
Justice et al. (2010)		+		+	+		+	

Note. + indicates the type of duration, intensity and content included in each study; ‘Immediate reading outcomes’ refers to improvements demonstrated immediately after the programme’s conclusion; ‘Sustained reading outcomes’ refers to reading improvements still evident when measured up to 5 months post-instruction.

Shorter time-efficient periods of PA instruction may offer several advantages in the classroom. These include 1) ensuring that foundation skills are in place from the outset of beginning-reading instruction, 2) facilitating early identification and support for children not making expected progress, 3) a reduction in resource requirements, 4) higher teacher and child motivation and 5) a greater likelihood of being incorporated into a busy classroom schedule. Table 1.3 demonstrates that studies of shorter duration have varied in intensity, focused on a broad range of PA skills and struggled to demonstrate sustained benefits for reading outcomes.

In contrast, longer periods of instruction require 1) consistency of teaching staff, 2) persistence by children who may already be struggling with word recognition and 3) increased resources. Further, longer periods of instruction may also lean towards a 'wait and see' approach in the identification of reading difficulties. Table 1.3 demonstrates that Shapiro and Solity's (2008) study, which was long in duration, high in intensity, and focused on PA at the phoneme-level, produced sustained benefits for reading growth. However, this study required the adaptation of the entire school day and stability of teaching staff over time. Based on the data from Table 1.3, there is scope for investigation into whether short and intensive phoneme awareness instruction can also produce positive and sustained improvements for reading growth in the classroom.

1.4.3.2 Review of classroom phonological awareness studies.

Of the studies reviewed in Table 1.3, Shapiro and Solity's 2008 study demonstrated the most significant effect on reducing the prevalence of reading disorder, using a long duration and highly intensive classroom programme focused on PA at the phoneme level. For two years, 251 British school children received explicit instruction in phoneme blending and segmentation, high-frequency phoneme–grapheme correspondences and sight vocabulary (i.e., learning to recognise phonetically irregular words by sight) over three 12-minute sessions per day as part of the classroom reading programme. This equated to approximately 110 hours of instruction. Two years of highly intensive exposure to phoneme level instruction in the classroom contributed to a reduction in the prevalence of reading disorder, from 20 per cent among children who received the usual programme, to five per cent among children who received instruction in phoneme awareness, when measured in the third year of school. Investigating whether a similar reduction in the prevalence of reading disorder can be achieved through a shorter period of highly intensive PA instruction focused at the phoneme-level may contribute to the manageable integration of PA teaching into the classroom schedule, and may support time-efficient early identification of reading difficulties.

Studies that are short in duration (i.e., less than one academic year), low in intensity and have a broad PA focus often report improved reading outcomes immediately following instruction. However, these same studies have struggled to demonstrate sustained improvements beyond five months of the programme's completion. Using a short 20-week programme, Fuchs et al. (2001) compared the effectiveness of teacher-delivered PA instruction with and without instruction in decoding printed words. Four hundred and four five-year-old children received instruction in either, PA and decoding instruction, PA instruction or the usual literacy curriculum (i.e., control). Fifteen PA activities were taken from the *Ladders to Literacy Program*, targeting syllable, rhyme and phoneme awareness. Decoding instruction was based on Peer Assisted Learning Strategies (PALS) and involved children working in pairs on word-reading tasks. PA and word-decoding instruction involved three 15-minute sessions per week, totalling 15-hours of teaching—considered low in intensity. Children who received PA and word-decoding instruction outperformed children in the PA only and control classrooms on reading and spelling tasks immediately following instruction. Five months post-instruction, children who received PA and word-decoding instruction no longer demonstrated a statistically significant advantage in reading and spelling ability. This study demonstrated that teaching a broad range of PA skills with low intensity (i.e., 45-minutes per week) over a short period is less effective for achieving sustained improvements in reading outcomes. It is possible that a narrow focus on phoneme-level skills with high intensity over a short period would produce results that are more promising.

In a short, low intensity programme, focused on a wide range of PA skills, Justice and colleagues (2010) demonstrated the importance of including specific teaching in PA at the phoneme level for children vulnerable to reading disorder. Sixty-six children aged between three years and three months and five years and six months received literacy and language instruction using a programme called *Read It Again* (RIA). The programme involved two 20 to 30 minute classroom sessions per week for 30 weeks, targeting PA, print, vocabulary and narrative knowledge. PA instruction targeted syllable, rhyme and phoneme awareness and was taught at least once per week for 20 to 30 minutes, equating to 10 to 15 hours of instruction. Children who received RIA instruction performed significantly higher than did comparison children (n = 71) on measures of language and literacy immediately following instruction. For children with low language abilities, this programme did not advance phoneme awareness and alphabetic knowledge to the same extent that it did for children with average to high language abilities. It is important to note that additional risk factors beyond language capabilities (e.g., socio-economic status) may have also influenced PA and reading outcomes. Nonetheless, these results reinforce the importance of focusing on phoneme-level skills, particularly for children at risk for reading difficulties. Investigating the benefits for reading growth following a short and intensive programme specifically focused on phoneme-level skills may help ensure all children have the opportunity to develop foundation skills that support further literacy acquisition.

McIntosh, Crosbie, Holm, Dodd, and Thomas (2007) also investigated the benefits of a short, highly intensive and broad PA-focused programme on the reading outcomes of 97 preschool children from low socio-economic localities. Children received 10-weeks of daily PA instruction targeting syllable segmentation, rhyme identification and generation and initial sound identification. Although significant improvements were identified in PA knowledge immediately following instruction, a follow-up investigation by O'Connor, Arnott, McIntosh, and Dodd (2009) indicated that initial gains in PA in preschool did not support accelerated literacy development in the early school years. These studies show that a short 10-week period of high-intensity instruction focused on a broad range of PA skills is less advantageous in generating sustained improvements for reading outcomes. Thus, it can be asked whether a similar 10-week, highly intensive period of instruction, focused on phoneme-level skills, as opposed to on syllables and rhyme, could have a more sustained influence on literacy growth.

Comparison of the studies in Table 1.3 suggests that little is known about the benefits of a short duration, high intensity PA programme focused at the phoneme level when integrated into beginning classroom reading programmes. Short programmes are more time-efficient within the classroom programme and can ensure critical foundation skills are in place to enable children to take advantage of beginning-reading instruction. Classroom programmes of short duration have targeted a broad range of PA skills in both low- and high-intensity formats, but have struggled to demonstrate sustained improvements for literacy beyond five months after the programme's conclusion. Conversely, it is possible that Shapiro and Solity's (2008) high-intensity focus on phoneme-level skills contributed to sustained literacy gains and a significant reduction in the prevalence of reading disorder.

Research suggests that awareness of smaller sound units (e.g., phonemes) may require more explicit and direct instruction (Fletcher, Parkhill, & Gillon, 2010), whereas awareness of larger sound units (e.g., words, syllables and rhyme) may develop from exposure to general classroom instruction. Moreover, raising awareness of smaller sound units (e.g., phonemes) is more likely to generalise to awareness of larger sound units (e.g., syllables and rhyme), whereas the reverse effect is less likely (Brown, 1998; Yeh, 2003). This has implications for time-efficient programmes and provides a strong rationale for including a specific focus on phoneme-level skills. Accordingly, research is needed on whether highly intensive, teacher-delivered PA instruction focused at the phoneme level over a short 10-week period can have a significant effect on reducing the prevalence of reading difficulties in the classroom. This would help bridge the gap between what is shown to be effective under controlled research settings (i.e., 20 hours is sufficient to raise the literacy abilities of the majority of children) and what can be realised within the busy day-to-day classroom setting.

1.4.3.3 Adapting time-efficient phonological awareness programmes for the classroom.

The *Gillon Phonological Awareness Training Programme* (PAT) (Gillon, 2000b) is a time-efficient instructional programme that has been successfully used in a number of individual and small group controlled studies (Gillon, 2000a, 2005; Gillon & McNeill, 2009). Due to its time-efficient framework (e.g., 20 hours of instruction over a 10-week period), the PAT was adapted for this thesis to determine whether time-efficient PA instruction at the classroom-level could achieve salient and sustainable benefits for early reading outcomes.

The PAT was originally designed for an intervention study, to investigate the effect of PA instruction on the PA ability, speech production and literacy development of five- to seven-year-old children with spoken language impairment (Gillon, 2000b). Children who received 20 hours of explicit PA instruction focused at the phoneme level over a 10-week period (two sessions per week) made significant improvements in PA and reading ability, compared to children who received traditional or minimal speech-language therapy. These benefits were maintained 11 months after intervention (Gillon, 2002). The current thesis investigates the effectiveness of a class-adapted version of this programme as a time-efficient supplement to the ‘usual’ classroom literacy curriculum. In this thesis, the ‘usual’ literacy curriculum is defined as reading instruction that has a primary focus on whole language instruction, a secondary focus on phonics instruction and no explicit teaching of PA skills (see Chapter 4 for more details).

1.4.3.4 Quantifying the effect of classroom phonological awareness instruction.

The calculation of effect size can be used to quantify the degree to which an instructional programme influences children’s learning. In education, the most prominent meta-analysis was conducted by Hattie (2009). This analysis involved the synthesis of approximately 800 meta-analyses to determine the most influential factors on learning in primary and secondary education. In Hattie’s study, an effect size of 1.0 (e.g., $d = 1.0$) was associated with a two- to three-year improvement in student performance, or an improvement rate of 50 per cent. Therefore, a new instructional programme that yields an effect size of 1.0 suggests that children receiving the new programme will perform at a higher level than 84 per cent of those children who did not receive the new programme. According to Hattie (2009), effect sizes of 0.4 or greater are desirable in educational research. Achieving effect sizes over 0.4 for the classroom PA programme implemented in this thesis will provide valuable evidence for whether a time-efficient 10-week period of phoneme-focused instruction raises reading achievement and reduces inequality.

1.4.3.5 Section summary.

A gap has been identified in current research on implementing PA instruction as part of a comprehensive classroom reading programme, which highlights the importance of investigating the effect of a short, high-intensity, phoneme-focused period of teacher-led PA instruction for children in the first year of formal schooling. Evaluating the effect of such a programme on literacy outcomes will contribute valuable information for the development of initiatives aimed at reducing inequalities in reading at the level of the classroom.

1.4.4 Reading inequality in the classroom: The case of spoken language impairment.

Inequalities in reading outcomes have been reported for a number of subgroups of children present in the mainstream classroom environment. Subgroups of children over-represented in reading difficulty statistics include: 1) children with spoken language difficulties, 2) children from indigenous or minority populations, 3) children from lower socio-economic backgrounds and 4) boys (Martin et al., 2007; UNESCO, 2009; UNICEF, 2010). Of these subgroups, children with spoken language impairment (SLI) are at particularly high risk for developing reading difficulties because of deficiencies in the underlying spoken language skills necessary to develop written language proficiency (Catts et al., 2001). To evaluate the effect of time-efficient measurement and instruction of PA on reducing reading inequalities in the classroom, children with SLI were selected as an at risk comparison group for experiments reported in this thesis.

1.4.4.1 Defining spoken language impairment.

According to the literature, two types of children with SLI are at high risk for experiencing difficulties learning to read—those with specific language impairment and those with speech impairments that are phonologically based (Gillon, 2004). According to the literature, specific language impairment refers to difficulties with spoken language despite average non-verbal intellectual ability, neurological, sensory, physical and emotional functioning and a positive language-learning environment (Bishop & Norbury, 2008; Conti-Ramsden & Botting, 1999). These children may present with deficits in receptive or expressive language across syntactic, semantic and morphological domains and can have co-occurring speech impairment. Speech impairments that are phonologically based refer to speech errors that are caused by difficulties processing linguistic information (Bird, Bishop, & Freeman, 1995; Gillon, 2004; Lewis, Freebairn, & Taylor, 2000). These can include poor access to phonological representations of words, inaccuracies selecting and sequencing speech sounds and difficulties assembling a phonological plan to produce spoken words. (i.e., this may also be referred to as expressive phonological impairment). Conversely, children may present with speech errors that are due to physical or motor issues. Such difficulties are often referred to as articulation impairments (Dodd, 1995).

Children with specific language impairment and speech impairments that are phonologically based are at greater risk of experiencing reading problems because they have difficulties with the phonological, semantic, syntactic or morphological aspect of spoken language that are necessary for learning to read. Deficits in the phonological aspect of spoken language will affect word-recognition skills, and weaknesses in semantics, syntax and morphology will restrict reading comprehension (Gillon, 2004; Kamhi & Catts, 2012; Nelson, 2010). Children with articulation impairment are not at substantial risk for reading problems because the physical deficits restricting speech sound accuracy are not critical for written language development (Dodd, 1995; Kamhi, Catts, & Mauer, 1990; Stackhouse, 1982). Throughout this thesis, the term 'spoken language impairment' (SLI) is used to refer to children with specific language impairment and children with speech impairments that are phonologically based.

1.4.4.2 Spoken language impairment and learning to read.

Evidence suggests that SLI in the preschool years (i.e., prior to starting school at five years of age) precedes and plays a causal role in the development of reading difficulties in the early and later school years. In a longitudinal study, Catts and colleagues (2002) identified that 70 per cent of children with language impairment in kindergarten (i.e., five years of age) performed below the 25th percentile on a composite measure of reading comprehension by the second and fourth grades. A follow-up study revealed that these children continued to exhibit poor reading ability when assessed in the 10th grade. Consequently, these early difficulties in learning to read appear to be persistent, difficult to remediate and may contribute to growing disparities in reading achievement between children with and without SLI (Morgan et al., 2008; Stanovich, 1986). This indicates that, in the early years of schooling, it is essential to routinely measure the language skills, including PA, of children with a history of spoken language difficulties to identify risk for reading problems.

A deficit in one or multiple areas of spoken language ability may negatively affect learning to read. This was shown by Catts and colleagues (1999) who profiled the spoken language abilities of 183 second grade children with reading comprehension scores at least one standard deviation below the mean. Results showed that the spoken language skills of poor readers were significantly inferior to those of good readers. Specifically, poor readers presented with deficits in the following language areas: in PA (56 per cent), in grammar (56 per cent), in phonological retrieval (45 per cent), in narration (44 per cent) and in vocabulary (39 per cent). Among these language areas, PA plays a critical role in the early stages of reading due to its effect on the development of strong word-recognition skills. As word-recognition skills develop, the language skills of grammar, vocabulary and text-level structures become increasingly important for reading comprehension (Konza, 2006). Children with widespread difficulties that incorporate a number of language areas (e.g., phonology, semantics and/or syntax) are at greater risk for reading disability than children who have trouble in only one language area (Lewis et al., 2000). This is because language skills underlying word recognition and reading comprehension can be simultaneously impaired (Catts & Kamhi, 1999).

SLI may also be exacerbated by the presence of a reading disorder (Catts, Kamhi, & Adolf, 2012). Interaction with printed language enhances the development of complex phonology, advanced vocabulary, stronger grammatical knowledge and text-level language skills. Children with SLI are less likely to read to the same degree as children with typical language development, thereby limiting the amount of exposure to written text and the number of opportunities to develop advanced language skills (Catts, Kamhi, & Adolf, 2012). Consequently, reading problems can contribute to increasing inequalities in language outcomes if they are allowed to manifest through inefficient early identification and instructional procedures in the classroom.

1.4.4.3 Prevalence of spoken language impairment in school-aged children.

In the United States of America, approximately 60 per cent of seven- to eight-year-old children with language specific difficulties and 80 per cent of children with nonspecific language difficulties have reading comprehension scores below the 25th percentile (Catts, 2009; Catts et al., 2002). In New Zealand, there are less rigorous data indicating how many school-aged children present with spoken language difficulties. One report estimates that five per cent of children between five and seven years of age present with an idiopathic speech difficulty (Gillon & Schwarz, 2001). This estimate does not include children with identified or unidentified language impairment or wider communication difficulties. Children with ongoing speech impairment, co-occurring language deficits and resolved speech difficulties tend to have significantly lower PA scores compared to children with typical spoken language (Raitano, Pennington, Tunick, Boada, & Shriberg, 2004). Due to the foundational role PA plays in reading development, it is important for teachers to place a high priority focus on assessing and monitoring PA, particularly for children with SLI. This should be part of a comprehensive process to ensure children have the wide range of language skills necessary for written language development.

1.4.4.4 Spoken language impairment and phonological awareness instruction.

Over the last two decades, research outside the classroom environment has demonstrated that one-to-one or small group instruction in PA can have a positive effect on the reading outcomes of children with SLI (Bus & Van Ijzendoorn, 1999; Ehri et al., 2001; Troia, 1999). More recently, researchers have investigated the benefits of classroom-wide PA instruction on the reading outcomes of all children, including those with SLI. For example, Justice and colleagues (2010) found that children with inferior language skills who received instruction using the RIA programme benefitted less in terms of phoneme awareness, letter-knowledge and print awareness, but showed equal benefit in areas of vocabulary, syntax and rhyme, compared to children with typical language skills. This programme targeted a number of key literacy areas and involved a broad focus on PA (e.g., syllable, rhyme and phonemes). One gap in need of further investigation indicated by this study was whether an explicit time-efficient focus on phoneme-level skills, as demonstrated in one-to-one or small group settings (Gillon, 2005), could reduce the risk of reading disability for children with SLI in the classroom.

A recent meta-analysis evaluating the components of effective classroom instruction for children with reading difficulties, including children with SLI, demonstrated that code-focused instruction (e.g., instruction focused on PA and letter-knowledge to help children decode an alphabetic cipher) exerted the best effect on reading outcomes. Wanzek and Vaughn (2007) reviewed 18 code-focused intervention studies for children with reading difficulties due to language disorder, low phoneme awareness, learning disability or being from a low socio-economic background. Studies were characterised by more than 100 hours of instruction delivered in a small group format. A school staff member administered the instruction with the support of researchers in 14 out of the 18 studies. Results showed that instruction focusing on phonics and blending sounds and words together produced the strongest effect sizes for reading improvement in children already experiencing reading difficulties. Of interest, effect sizes were similar between interventions of long duration (i.e., over one year) and those of shorter duration (i.e., less than one year).

This further supports this thesis's aim, which holds that investigating the benefits of a time-efficient period of classroom-wide PA instruction on the reading abilities of children with SLI will provide information regarding the optimal amount of time teachers should devote to PA instruction prior to accessing supplementary supports or specialist education services. Moreover, comparison of reading outcomes between children with SLI and those showing typical development will provide valuable insight into the effect that teacher-led phoneme-focused instruction has on reducing inequalities in reading outcomes after one year of formal schooling.

1.4.4.5 Section summary.

Children with SLI are one cohort of children in the classroom who are at heightened risk for experiencing early reading disability. Equipping classroom teachers with user-friendly and time-efficient methods of identifying risk for reading disability among this group of children is essential for ensuring these children do not experience unnecessary or exacerbated reading difficulties. Further, providing teachers with time-efficient programmes that ensure foundational skills are in place from the outset of beginning literacy instruction may help to not only reduce inequalities in literacy outcomes, but also to ensure the timely transition to supplementary small group or special education support for children not adequately responding to classroom instruction.

1.5 Thesis Synopsis

1.5.1 Summary and thesis aims.

Ensuring children read efficiently in their own classrooms is a critical issue in reading education. The prevalence of reading difficulties and inequalities between good and poor readers in some of the world's wealthiest countries has raised concerns regarding how far children are falling behind in reading acquisition at an international level (NAEP, 2003; Nicholson, 2009; UNICEF, 2010). Early problems in learning to read are difficult to remediate and come to pervade the social, economic, personal and academic aspects of an individual's life (UNESCO, 2009). Reducing inequalities in reading outcomes is a complex process for educators, researchers and policy makers, and requires collaboration and research initiatives at many levels of the education system.

One part of this process is to ensure key predictors of literacy success can be effectively measured and taught in the daily classroom environment. A convincing body of evidence demonstrates that PA is a critical precursory reading skill and a powerful predictor of early reading success (Catts, Wilcox, Wood-Jackson, Larrivee, & Scott, 1997; Treutlein, Zoller, Roos, & Scholer, 2008; Ehri et al., 2001; Gallagher, Laxon, Armstrong, & Frith, 1996; Gillon, 2004). Controlled research studies show that when children are provided with effective assessment and instruction in PA, reading difficulties can be prevented or minimised (e.g., Gillon, 2000a, 2002, 2005). More recently, researchers have directed their attention towards evaluating how the benefits of PA can be best incorporated into the heterogeneous classroom environment (e.g., Shapiro & Solity, 2008). Many practical issues must be considered when bridging theory and research into the classroom environment. One practical consideration is time-efficiency. Time-efficiency in PA assessment and instruction may not only maximise its inclusion in the classroom programme, but also ensure children at risk for reading difficulties are identified and supported before difficulties become unnecessarily severe. A time-efficient and effective approach to equalising reading outcomes is also more suited to the needs of the busy classroom environment.

One group of children at high risk of experiencing inequality in reading outcomes are those with SLI. This heterogeneous group are reported to be four to five times more likely to experience early literacy failure due to deficiencies in the underlying language skills that support early word recognition and reading comprehension (Catts et al., 2001). Identifying these children and raising their reading performance to an appropriate level through effective classroom PA practices is of critical importance.

New Zealand provides an ideal environment to trial innovative approaches towards the inclusion of PA instruction in the classroom. Over the last decade, New Zealand has consistently demonstrated one of the largest disparities between the scores of good and poor readers in international studies of reading achievement (Martin et al., 2007). These disparities in achievement, in combination with a core reading curriculum that favours whole language strategies, provides a challenging context within which to rigorously test whether time-efficient and effective approaches to the assessment and instruction of PA in the classroom can have a significant influence on raising literacy achievement for the majority of children. Towards this goal, the experiments reported in this thesis addressed the following hypotheses:

1. A computer-based assessment tool, if well developed, will be an efficient and effective method for monitoring PA development and predicting reading outcomes in the first year of formal education.
2. Teacher-implemented classroom-wide phonological awareness instruction that is time-efficient and focused at the phoneme level will significantly raise reading achievement for children with and without spoken language difficulties in the first year of school.

3. Teacher-implemented classroom-wide phonological awareness instruction will have a positive impact on reducing the prevalence of reading difficulties and minimising inequality in reading outcomes when included as part of the beginning literacy curriculum.

The first hypothesis is addressed through research reported in Chapters 2, 3 and 6. Chapter 2 discusses the development of a CBA tool designed to screen and monitor the PA abilities of five-year-old children in the first year of education. In Chapter 3, the time-efficiency of the CBA is compared to a paper-based testing modality and congruency of scores between each assessment medium are examined. Chapter 6 presents validity and reliability information for the computer-based PA assessment when used to track the development of children exposed to current classroom literacy practices in New Zealand. The second and third hypotheses are addressed through research presented in Chapters 4 and 5. Chapter 4 investigates the effect of a short and intensive period of teacher-implemented PA instruction focused at the phoneme-level on reading outcomes for five-year-old children, inclusive of a Māori cohort. In Chapter 5, the effect of teacher-directed classroom PA instruction on the reading outcomes of children with SLI is examined. The influence this has on reducing inequalities in literacy achievement between children with and without an elevated risk for reading difficulty is discussed. Finally, Chapter 7 aggregates the findings from the experiments reported in this thesis in a general discussion and offers future directions for research initiatives that aim to raise achievement and reduce inequalities in reading outcomes for school-aged children.

Chapter 2: Development of a Computer-Based Phonological Awareness Screening and Monitoring Tool

2.1 Introduction

The screening and monitoring of phonological awareness (PA) in the classroom is of great importance to the early identification and prevention of reading disorder (Vellutino, Scanlon, Small, & Fanuele, 2006). This is because PA, particularly at the phoneme level, is a powerful prognostic marker and underlying precursor of early reading success (Adams, 1990). It is possible that equipping teachers with efficient and effective methods of screening and monitoring PA for all children in the classroom may help ensure as many as one in three children underperforming in reading acquisition are identified early before they fall behind their peers in reading development (NAEP, 2003; Nicholson, 2009). A practical consideration when attempting to screen and monitor development of all children in the classroom is that of 'time'. According to McLeod et al. (2003), teaching methods and strategies may go unused by educators if they are time consuming to implement (see Section 1.4.1.3). The purpose of this Chapter is to describe the development of a time-efficient computer-based PA screening and monitoring tool designed to track PA progress and identify risk for reading difficulties among five-year-old children in the first year of school. A key point of difference between existing PA tools and the CBA described in this Chapter is the high-priority focus on measuring PA skills at the phoneme-level.

The aim of developing a computer-based PA screening and monitoring tool as part of this doctoral research is to provide classroom teachers with a time-efficient and user-friendly method of 1) identifying children at risk for reading disorder from the start of beginning literacy instruction, 2) monitoring progress and identifying children not keeping pace with their peers, and 3) profiling patterns of strength and weakness in PA that will contribute to the teacher's understanding of a child's reading ability, inform curriculum design and support referral to special education services for children with significant difficulties. The computer-based PA tool was developed and investigated with a view to measuring the resultant improvements in PA in response to classroom PA instruction (see Chapters 4 and 5) and to establish the validity and reliability of the instrument based on a sample of children receiving the national literacy curriculum during the first year of formal schooling in New Zealand (see Chapter 6).

This tool differs from the majority of commercially available instruments in that it is computer-based as opposed to paper-based (see Section 1.4.2.1). The computer-based PA tool differs from available computer-based assessments (CBA) such as the *Cognitive Profiling System* (CoPS) (Singleton et al., 1996) and the *Performance Indicators in Primary School (*PIPS) (Tymms, 1999) in that it focuses primarily on assessing PA development at the phoneme-level. Computer-based administration allows children to administer the assessment independently because the computer presents all test stimuli and records and scores responses into a separate database (Bjornsson, 2008; Bridgeman, 2009; Martin, 2008; Ripley, 2008; Singleton et al., 1999; Tymms, 2001). This reduces the amount of time teachers need to be directly involved in the administration and scoring of the assessment (Martin, 2008; Olsen, 1990), and may therefore offer a time- and cost-effective alternative for the early identification of reading difficulties in the classroom.

Internationally, exploring time-efficient methods of measuring PA in the classroom may enhance educators' ability to address inequalities in reading outcomes by ensuring children who lack foundation literacy skills are identified and supported in a timely manner. In New Zealand, the construction of a PA tool would provide one of the first instruments available to teachers developed using a New Zealand sample of school-aged children. Further, it would provide New Zealand educators with a method to begin addressing what is reported to be one of the OECD's largest gaps between high- and low-ability students in reading achievement (Martin et al., 2007).

A number of steps are required in the construction of a computer-based PA screening and monitoring tool. Researchers across disciplines suggest that the construction of a new assessment tool requires the alignment of the following three factors: 1) a theoretical framework, conceptualising skills known to be critical to early literacy success (Anderson et al., 2004; Croft, Strafford, & Mapa, 2000), 2) robust test development procedures that ensure tasks and items are suitable for the intended population, discriminate between high- and low-achieving students, and are time efficient to administer (Dewar & Telford, 2003) and 3) appropriate measurement models, to ensure the technical adequacy of a tool for monitoring progress and informing instructional practice (Denton et al., 2006). A theoretical framework for selecting PA as a measurement focus, in particular skills at the phoneme level, is described in Chapter 1 (see Section 1.3.4).

The current chapter describes how the computer-based PA assessment tool was developed, including procedures for the selection of tasks, test items and an administration modality. Subsequent chapters investigate the time-efficiency and effectiveness of the computer-based tool (see Chapter 3), the application of this instrument to the classroom environment for monitoring response to instruction (see Chapters 4 and 5) and the psychometric properties of the tool (see Chapter 6).

2.2 Test Construction

In this section, test development procedures concerning the selection of tasks, items and an administration modality are successively addressed. Underlying the test development process was the need to select tasks and items that, according to the literature, best represent PA ability, differ in difficulty to ensure discrimination between high- and low-ability students and are predictive of later reading success (Vloedgraven & Verhoeven, 2007). This was to ensure that the CBA tool would be a sensitive indicator for children vulnerable to inequalities in reading outcomes.

2.2.1 Selection of phonological awareness tasks.

Researchers have used a wide variety of tasks to measure PA ability in children of different ages (Catts et al., 1997). PA can be measured at the syllable, rhyme and phoneme levels. However, performance at the phoneme level is considered the strongest indicator of later literacy ability (Gillon, 2004). Therefore, screening and monitoring children's ability at this level is an important practice for classroom teachers. Tasks at the phoneme level can vary in the complexity of the cognitive operations being performed (e.g., identifying, blending, segmenting and deleting), the number of phonemes being manipulated (e.g., identifying one phoneme or blending three phonemes) and the demands on working memory (e.g., holding one operation in memory while performing another) (Schreuder & van Bon, 1989; Yopp, 1988).

In general, phoneme identification (i.e., the ability to identify one phoneme in a word) is reported to be one of the easiest tasks at the phoneme level and is a powerful predictor of learning to read (Elbro, Borstrom, & Petersen, 1998; Hoien et al., 1995; Stanovich, Cunningham, & Cramer, 1984). Further, phoneme identification is considered an appropriate task for five-year-old children (Torgesen, 1998). Phoneme blending and segmentation tasks (i.e., blending together or pulling apart the sounds in a word) are thought to provide the most robust relationship with early literacy success (van Bon & van Leeuwe, 2003; Yopp, 1988) and, although difficult at school-entry (Adams, 1990; Chard & Dickson, 1999), become increasingly sensitive indicators as children interact with beginning literacy instruction.

Phoneme deletion requires children to delete a phoneme from a word to create a new word, and is predictive of reading outcomes (Muter & Snowling, 1998). Phoneme manipulation requires the addition or deletion of phonemes to create a new word and is the most difficult of the phoneme-level tasks. The ability to manipulate phonemes generally emerges between five and seven years of age (Johnson & Roseman, 2003; Paul, 2007). The CBA developed for this thesis has a high-priority focus on measuring PA skills at the phoneme level. This is because phoneme-level skills predict later reading outcomes more accurately than measures of syllable or rhyme awareness.

In addition to the ability to predict reading outcomes, it is important to recognise the suitability of various PA tasks based on age and stage of development. To measure the PA skills of high- and low-ability students accurately, a range of tasks graded in level of difficulty must be considered. Low-ability students are more likely to struggle with phoneme-level knowledge (Kamhi & Catts, 2012; Justice et al., 2010; Vloedgraven & Verhoeven, 2007). Thus, the inclusion of an easier task focused at the rhyme level may be necessary to avoid floor effects and to determine a child's exact level of PA knowledge (Adams, 1990; Yopp, 1988). Research has shown that rhyme awareness, although a less powerful predictor than phoneme awareness, provides an independent contribution to reading outcomes (Goswami & Bryant, 1990; Hoien et al., 1995). Rhyme oddity (i.e., identifying which word does not rhyme from a selection of words) is one of the more challenging rhyme awareness tasks (Gillon, 2004) and was selected for inclusion in the CBA for this reason. In summary, the following PA tasks were selected for inclusion in the CBA: rhyme oddity, initial phoneme identity, final phoneme identity, phoneme blending, phoneme deletion and phoneme segmentation. In addition, the National Reading Panel's meta-analysis (Ehri et al., 2001) demonstrated that PA and letter-knowledge are two of the best predictors of how well children will learn to read during the early school years. Therefore, letter-knowledge was also included in the development of the CBA.

2.2.2 Selection of test items.

Following the selection of tasks, the construction of test items within each task needed to be addressed. To achieve this, a review of the literature was undertaken to identify criterion-referenced probes commonly used to measure progress in response to PA instruction. For rhyme oddity, initial phoneme identity and letter-knowledge tasks, probes developed by Bradley and Bryant (1983) and modified by Gillon (2005) were used as test items (for examples see www.education.canterbury.ac.nz/people/gillon/assessment_probes.shtml). These probes were adapted slightly for the CBA by updating the graphics and replacing the voiceless glottal fricative (i.e., ***h***ippo) used in the original initial phoneme identity practice item with a voiced alveolar stop (e.g., ***d***og). This change meant that this practice item was more salient than the original practice item (e.g., ***d*** as opposed to ***h***) and used an animal that was more familiar to five-year-old children. For final phoneme identity, phoneme blending, phoneme deletion and phoneme segmentation, probes developed by Stahl and Murray (1994) were used to guide the construction of test items. The original Stahl and Murray (1994) probes require children to produce a verbal response, whereas use of a computer-based modality in this thesis required a receptive response (e.g., clicking a computer mouse). To guide the selection of new receptive-based test items for the CBA, the increasing difficulty of syllable structures in the original Stahl and Murray (1994) probes (e.g., from simple CVC words to CCVC and CVCC words) were used.

In the construction of items for final phoneme identity, phoneme blending, phoneme deletion and phoneme segmentation tasks, high-frequency words were drawn from *The living word vocabulary: A national vocabulary inventory* (e.g., spoken words) (Dale & O'Rourke, 1981) and the *Ready to Read* series (e.g., written words) (MOE, 2007) where possible. Only words representing tangible objects were selected, to ensure all test items could be presented as images on a computer screen with as little ambiguity as possible. To alter difficulty between test items, syllable structures and manners of articulation were manipulated. As mentioned, the syllable structure of words was in line with that used by Stahl and Murray (1994) to achieve differential difficulty (Anthony & Francis, 2005; Schreuder & van Bon, 1989). Syllable structure can be described by the number and sequence of consonants (C) and vowels (V) that comprise a word. For example, test items containing two- or three-phoneme words (e.g., pie (CV) or dog (CVC)) are easier than test items containing four-phoneme words with initial or final consonant clusters (e.g., space (CCVC)) and hand (CVCC), respectively). All selected words were monosyllabic. Final phoneme identity, phoneme blending, phoneme deletion and phoneme segmentation tasks contained CVC, CCVC and CVCC words. Rhyme oddity and initial phoneme identity test items contained CV and CVC words. Importantly, test items were multiple-choice in format to allow for computer administration. Therefore, each test item contained three options: one correct option and two distractor options. The three response options per test item were similar in syllable structure. However, the syllable structure between test items within a task purposefully varied to ensure a range of difficulty levels.

Manner of articulation was also taken into consideration when selecting test items for final phoneme identity, phoneme blending, phoneme deletion and phoneme segmentation tasks. Manner of articulation refers to how the articulators (e.g., lips, teeth, tongue and soft palate) involved in speech production produce a particular speech sound (Bernthal, Bankson, & Flipsen, 2009). There are five manners of articulation in New Zealand English: stops, fricatives, affricates, nasals and approximants (Hay, Maclagan, & Gordon, 2008). Manner of articulation has been shown to affect the difficulty of test items, and can therefore help discriminate between high- and low-ability students (de Graaff, Hasselman, Bosman, &

Verhoeven, 2007). For example, in phoneme-level tasks, continuant sounds are easier to identify, delete and manipulate (e.g., s, f, v) than non-continuant sounds (e.g., k, g, t, d) (Hubbard & Mahanna-Boden, 2000). Further, identifying the correct option among a choice of three possible responses may be easier when a continuant sound is contrasted with a non-continuant sound. In the case of rhyming tasks, test items can be made easier by having response options end with a consonant as opposed to a vowel (e.g., 'cat, mat' versus 'key, sea') (Schuele & Boudreau, 2008). It has also been reported that rime units are easier to detect when the final sound is visible (e.g., 'pup, cup' versus 'dog, fog'). However with computer-based administration, children do not get additional visual cues from an examiner. To develop test items of varying difficulty, care was taken to use a range of manners of articulation both within test items (i.e., between correct options and distractor options) and between test items within a task. Table 2.1 profiles the linguistic complexity of items within each PA task in terms of syllable structure and manner of articulation. As the table indicates, a variety of syllable structures and manners of articulation were employed to achieve differential difficulty. Few affricate sounds are represented among the test items. However, this is appropriate since proportionality to other manners of articulation is maintained (i.e., there are only two sounds in the affricate sound class in New Zealand English).

Table 2.1

Linguistic complexity of test items within each phonological awareness task

Task	Number of Items	Syllable Structure	Manner of articulation				
			Stops	Fricatives	Affricates	Nasals	Approximants
RO	10	CV, CVC, CVCC	31	10	0	15	7
IPI	10	CV, VC, CVC, CVCC	33	10			
FPI	10	CVC, CCVC, CVCC	28	17	0	11	13
PB	15	CVC, CCVC, CVCC	56	20		22	19
PD	15	CV, CVC, CCVC, CCVC, V, VC, CCVC, CCV	50	15	1	17	15
PS	18	CV, CVC, CCVC, CCV, CVCC	22	9	1	5	4
			e.g., p, b, t, d, k, g	e.g., f, v, θ, ð, s, z, ʃ, ʒ, h	e.g., tʃ, dʒ	e.g., m, n, ŋ	e.g., w, l, r, j

Note. This table outlines the types of syllable structures and manners of articulation present within each PA task. Syllable structures and manners of articulation include correct options and distractor options, in addition to original words in the phoneme deletion task (i.e., the original word from which one phoneme was deleted). RO = rhyme oddity; IPI = initial phoneme identity; FPI = final phoneme identity; PB = phoneme blending; PD = phoneme deletion; PS = phoneme segmentation. This table does not include information on practice items.

2.2.3 Administration using the computer.

Computer-based administration of test content can offer educators several advantages over traditional paper-based methods of testing. These advantages include:

1) Savings in labour time
2) Child self-administration
3) Precision of item presentation
4) Increased availability and delivery through the Internet
5) Minimal training for classroom teachers or other education staff
6) Motivational for children, thereby increasing the reliability of results
7) Uses existing technology in the classroom (e.g., class computer)
8) Low cost once created, without ongoing expense such as the purchasing of test booklets (Bjornsson, 2008; Bridgeman, 2009; Martin, 2008; Ripley, 2008; Singleton et al., 1999; Tymms, 2001).

Adobe Flash CS3 and Action Script were used to produce a multi-media platform for the CBA. The computer presents each task and associated items in a controlled manner using a main character, colourful animated or static images and pre-recorded verbal instructions. Children are encouraged to help the main character, an alien called ‘Poe’, to complete each of the following tasks: 1) rhyme oddity, 2) initial phoneme identity, 3) final phoneme identity, 4) phoneme blending, 5) phoneme deletion, 6) phoneme segmentation, 7) letter-name, and 8) letter-sound knowledge. The character of ‘Poe’ was selected to engage children’s interest and is a figure that is not culturally bound. Each task is presented in a game format, which is reported to be motivating for children, non-threatening and able to elicit more accurate and reliable results (Beech & Singleton, 1997; Singleton et al., 2006).

The teacher is required to set up the computer, open the CBA program, and ensure each child enters their name when prompted, to enable the recording of results into a database. From this point, the teacher is not required, providing the child has adequate computer literacy skills (e.g., can listen to the verbal instructions from the computer and can use a mouse to click their response).

2.2.3.1 Multiple-choice presentation.

Each task begins with two practice items followed by 10 to 18 test items. Each item is presented in a multiple-choice format. Research demonstrates that the optimal number of options per multiple-choice test item is three (Rodrigues, 2005). Therefore, for rhyme oddity, initial phoneme identity, final phoneme identity, phoneme blending and phoneme deletion, each test item consists of one correct option and two distractor options. One distractor item is phonetically similar to the correct option and the other distractor item is phonetically dissimilar. The response format for phoneme segmentation is slightly different in that children click a box (i.e., up to four boxes) for each sound they hear in a spoken word. Due to this difference, phoneme segmentation also includes a demonstration item prior to practice items. In the letter-name and letter-sound tasks, children are required to click the stated letter or sound from a choice of six letters. This is in keeping with probes by Gillon (2005).

The positioning of the correct option (e.g., as the first, second or third option) is varied from item-to-item to minimise the possibility of a correct response due to guessing behaviour (e.g., by a child clicking a favoured position). Each test item begins with simple verbal instructions in line with those used by Bradley & Bryant (1983), Gillon (2005) and Stahl & Murray (1994), followed by the naming and presentation of each multiple-choice response options (i.e., once a response option is presented on the screen, it remains there until the child responds or until a10-second timer runs out). Multiple-choice options are presented as static graphics (e.g., not animated) to ensure test items are engaging but not distracting.

The main character, 'Poe', and the words to be manipulated in the phoneme deletion and phoneme segmentation tasks involve animated images. The multiple-choice format enables items to be administered in a receptive manner, thereby reducing demands on working memory (Reitsma, 2002) and allowing for the use of a computer-based testing modality.

2.2.3.2 Responding with a computer mouse.

Children are required to indicate their response to test items by clicking the computer mouse on a picture or a series of blocks (depending on the task) presented on the computer screen. Prior to commencing the assessment, children participate in a 'mouse practice' activity in which they are given the chance to practice moving and clicking on static images with the computer mouse. For example, one instruction asks children to 'click the red flashing box'. This activity provides children with an opportunity to become familiar with the response format of the assessment. This is particularly important if they have had no experience using a computer mouse. Alternatively, if children prefer to do so, they can point to the computer screen and have the teacher click the mouse on their behalf. Pilot testing (see Chapter 3) revealed no significant difference between participants who independently responded to test items with the computer mouse, and those who preferred to touch the computer screen and have an examiner click the computer mouse on their behalf. However, it should be noted that the latter scenario does increase the time required for the teacher to be part of the administration process. For practice items and test items, participants have 10 seconds to indicate their response. This is to ensure the assessment moves swiftly and is completed in a timely fashion. For practice items, verbal feedback is provided regarding which multiple-choice option is correct and why. This feedback is provided irrespective of whether participants provide a correct or incorrect response. Specific feedback is not provided for test items, however non-specific encouragement such as, 'good listening' and 'you're doing a great job' are offered occasionally to maintain motivation and attention throughout each task.

2.2.3.3 Scoring and time allowance for test items.

The computer records, stores and scores all responses automatically into a single database file, along with the date and time that the assessment was completed. One point is allocated to each test item with a correct response. No points are awarded to test items with an incorrect response. Tasks take an average of four minutes and 30 seconds to five minutes and 30 seconds minutes to complete, and are not adaptive (i.e., the computer does not select test items based on how a child is performing meaning that all items are administered). Figure 2.1 provides an example of an output table of results produced by the CBA.

Database: Rhyme Oddity

Time: Jul 23, 2010 @ 4.25pm

Rows: 1

id	name	date	res1	res2	res3	res4	res5	res6	res7	res8	res9	res10	res11	res12	nbright
10	IALP	2010-07-23	ball	bat	bus	leg	saw	sand	hen	book	kite	lock	shell	sing	4

Figure 2.1. Example of an output table from the computer-based phonological awareness tool

2.3 Phonological Awareness Task and Item Descriptions

The computer-based PA screening and monitoring tool presented here is intended to be administered at school-entry, when children are five years of age, and then re-administered during the middle and end periods of the first year of formal schooling. According to Jenkins (2003), three measurement points per annum in the early stages of education are conducive to helping identify early difficulties in learning to read. The computer-based PA tool is comprised of six PA and two letter-knowledge tasks, equating to 23 practice or demonstration items and 114 test items. A description of each PA task is provided below and a compact disk of examples is included in Appendix F.

2.3.1 Rhyme oddity.

A rhyme oddity task was used to measure rhyme awareness. It consisted of two practice items and 10 test items. Children are required to help an alien character, ‘Poe’, and his friend ‘Clown’ identify which word does not rhyme from a choice of three words, each of which is represented by a static picture. For example, the computer asks, ‘Which word does not rhyme? /cat, mat, bus/’. Children have 10 seconds to select their response by clicking on a picture with the computer mouse. The difficulty between test items and between multiple-choice options within test items was manipulated by using different manners of articulation. This task is approximately three to four minutes in duration.

Example: Practice item 2

Instructions: ‘Which word does not rhyme? /pig, hat, bat/’.

Rhyme Oddity			
Practice Items:			
1	fish	dish	ball
2	pig	hat	bat

Test Items:			
1	cat	mat	bus
2	peg	doll	leg
3	saw	toe	bow
4	sand	hand	cup
5	hen	car	pen
6	dog	book	hook
7	bun	sun	kite
8	tent	lock	sock
9	shell	duck	bell
10	ring	sing	lamb

Note. Grey shading represents the correct multiple-choice option.

2.3.2 Initial phoneme identity.

This task consists of two practice items and 10 test items. Children are introduced to Poe's friends, who each have a favourite sound. Children are required to identify which word from a choice of three starts with a target sound (i.e., the favourite sound of one of Poe's friends). Each word on the computer screen is represented by a colourful static picture. As an example, the second friend Poe introduces is Mouse. The following instructions are provided, 'This is my friend Mouse. Mouse likes words that start with the /m/ sound. What word starts with the /m/ sound? /doll, bear, milk/'. Children have 10 seconds to select their response by clicking one of three images with the computer mouse. Difficulty between test items and within test items (i.e., between correct and distractor options) was adjusted by using different syllable structures and manners of articulation. This task takes approximately five to six minutes to complete.

Example: Practice item 1

Instructions: 'This is my friend Dog. Dog likes words that start with the /d/ sound. What word starts with the /d/ sound? /moon, duck, whale/'.

Initial Phoneme Identity

Practice Items:	Friend:			
1	**D**og	moon	duck	whale
2		horse	rope	door
Test Items:				
1	**M**ouse	doll	bear	milk
2		mat	dog	book
3	**S**eal	bee	sun	tent
4		saw	tie	hook
5	**C**at	bus	kite	arm
6		comb	dish	soap
7	**B**ee	cat	leg	ball
8		car	boat	shoe
9	**F**ish	foot	hat	pig
10		duck	bell	fire

Note. Grey shading represents the correct multiple-choice option; Bolded letters at the start of each friend's name indicates the initial phoneme to be identified.

2.3.3 Final phoneme identity.

As with initial phoneme identity, this task consists of two practice items and 10 test items. Children are required to identify which word from a choice of three words ends with a target sound (each word is represented by a colourful static picture). This task follows a similar structure to the probes of Stahl and Murray (1994), and does not involve Poe's friends. For example, in test item 1 Poe asks, 'Which picture ends with the /p/ sound? /room, seal, soup/'. Children have 10 seconds to select their response by clicking on one of three images with the computer mouse. Difficulty between test items and within test items (i.e., between correct and distractor options) was achieved by altering linguistic complexity (i.e., using CV, CCVC and CVCC syllable structures) and varying manners of articulation. This task takes approximately four to five minutes to complete.

Example: Practice item 1

Instructions: ‘Which picture ends with the /t/ sound? /cat, can, man/’.

Final Phoneme Identity

Practice Items:			
1	ca**t**	can	man
2	shoe	kite	shi**p**
Test Items:			
1	room	seal	sou**p**
2	ha**t**	hole	sun
3	rope	ri**c**e	pan
4	foo**d**	fan	hook
5	green	gra**ss**	fire
6	han**d**	horse	milk
7	ball	goat	ban**k**
8	cake	cam**p**	bed
9	worl**d**	walk	face
10	tent	toast	ma**p**

Note. Grey shading represents the correct multiple-choice option. Bolded letters at the end of each correct multiple-choice option indicate the phoneme to be identified.

2.3.4 Phoneme blending.

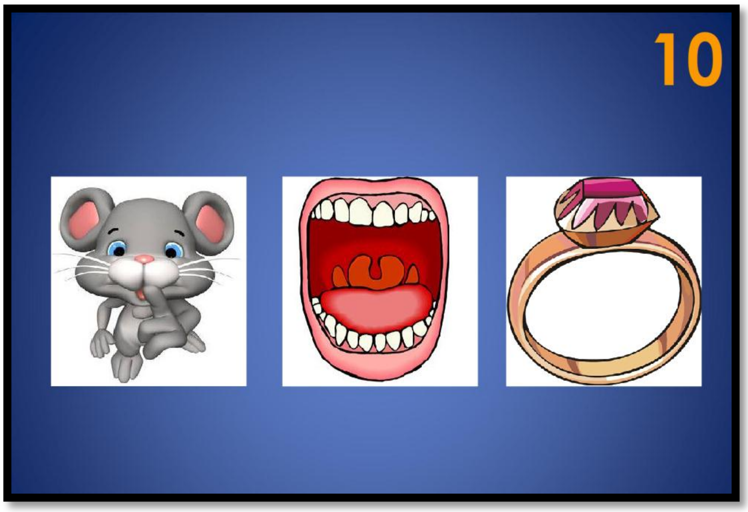
This task consists of four practice items and 15 test items. Children are asked to listen to Poe say a word one phoneme (i.e., sound) at a time, and then select one picture from a choice of three as their answer. For example, in practice item 2, Poe says, 'Click on the picture that you think I am saying: sh–ee–p'. Children have 10 seconds to select their response by clicking on one of three static images with the computer mouse. Items increase in difficulty from simple CVC words to words with initial (i.e., CCVC) and final (i.e., CVCC) blends. There are two practice items before commencing the task, followed by one practice item before test items with initial blends and another practice item before test items containing final blends. Within each test item, one distractor option is phonetically similar to the correct multiple-choice option to increase item difficulty as described below:

1) The first five test items contain simple CVC words with one distractor option containing the same initial phoneme as the correct option. For example, in test item 1 the correct option is 'dot', a phonetically similar distractor option is 'dog' and the final non-similar distractor option is 'man'.
2) The second five test items contain CCVC words with the initial consonant cluster being identical between the correct option and one of the distractor options. For example, in test item 7, the correct option is 'crab', the phonetically similar distractor option is 'crane' and the non-similar distractor option is 'snake'.
3) The final five test items contain CVCC words with the final consonant cluster being similar between the correct option and one of the distractor options. For example, in test item 11, the correct option is 'pond', the phonetically similar distractor option is 'point' and the non-similar distractor option is 'fast'.

This task takes approximately seven to eight minutes to complete.

Example: Test item 2

Instructions: ‘Here is a mouse, a mouth, and a ring. Click on the picture that you think I am saying: m–ou–se’.



Phoneme Blending			
Practice Items:			
1	cake	cape	ring
2	sheet	sheep	mop
Test Items:			
1	dog	dot	man
2	mouse	mouth	ring
3	seal	seat	duck
4	bug	sun	bun
5	cat	cap	lock
Practice Item:			
3	fly	snake	snail
Test Items:			
6	flip	drum	flag
7	crab	crane	snake
8	bread	spade	space
9	tray	clown	train
10	stop	star	plane
Practice Item:			
4	bump	mast	last

Test Items:			
11	point	fast	pond
12	bank	band	toast
13	desk	lamp	lamb
14	wand	mask	world
15	cast	cost	jump

Note. Grey shading represents the correct multiple-choice option.

2.3.5 Phoneme deletion.

This task consists of four practice items and 15 test items. Children are required to delete one sound from a target word to create a new word, and then indicate which picture from a choice of three represents the new word. For example, Poe provides children with the following instructions, ‘This is a cup. Say “cup”. Say “cup” again, but don’t say the /k/ sound at the start. What word do we get? Is it: pup, egg or up?’ Children have 10 seconds to select their response by clicking on one of three static images with the computer mouse. The original word is represented by an animated graphic to maintain interest and attention. Items are graded in difficulty by increasing the complexity of the syllable structure of the target word, from simple CVC to CCVC and CVCC structures. As with phoneme blending, two practice items are introduced prior to commencing the task. One practice item is presented before completing test items with initial consonant blends and another before commencing test items with final consonant blends. Within test items, one distractor option contains similar initial or final phonemes to the correct multiple-choice option as outlined below:

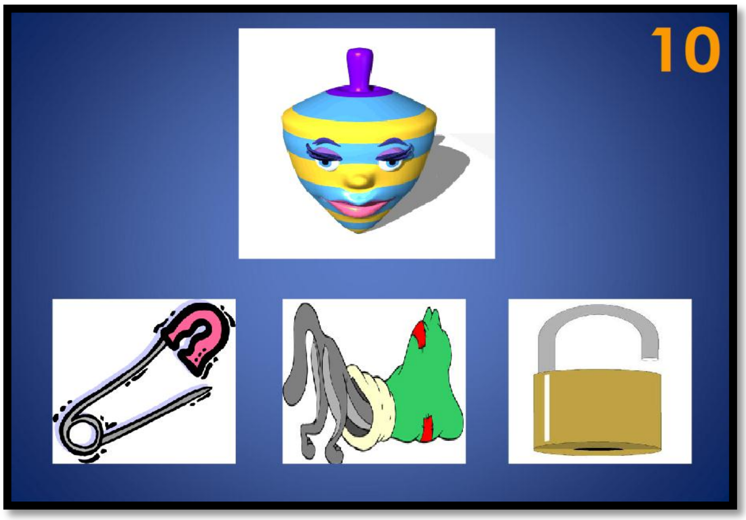
1) The first five test items contain simple CVC or CV words that require the first phoneme (i.e., sound) to be deleted to create a VC or V word. One distractor option contains a similar phonological composition to the correct multiple-choice option. For example, in test item 2 the original word is ‘door’. Hence, the correct option following deletion of the /d/ phoneme is ‘oar’. The phonetically similar distractor option is ‘core’, and the non-similar distractor option is ‘sew’.

2) The second five test items contain CCVC or CCV words that require the first phoneme (i.e., sound) to be deleted to create a CVC or a CV word. Again, one distractor option is phonetically similar to the correct option. For example, in test item 7 the original word is ‘store’. Hence, the correct option following deletion of the /s/ phoneme is ‘tore’. The phonetically similar distractor option is ‘oar’, and the non-similar distractor option is ‘ring’.
3) The final five test items contain CCVC and CVC words that require the final phoneme (i.e., sound) to be deleted to create a CCV or CV word. The correct option and one distractor option contain similar phonology. For example, in test item 11 the original word is ‘beach’. Hence, the correct option following deletion of the /ch/ phoneme is ‘bee’. The similar distractor option is ‘beak’, and the non-similar distractor option is ‘cat’.

This task takes approximately 10 to 11 minutes to complete.

Example: Test item 9

Instructions: ‘This is Spin. Say “Spin” [pause while the child says, “Spin”]. Say “Spin” again, but don’t say the /s/ sound at the start. What word do we get? Is it: pin, in or lock?’



Phoneme Deletion				
Practice Items:				
1	(pie)	eye	moon	hen
2	(peal)	sing	eel	meal
Test Items: *(deleting the first sound)*				
1	(cup)	pup	egg	up
2	(door)	oar	core	sew
3	(farm)	bow	arm	art
4	(bake)	ape	sun	ache
5	(bear)	ear	tear	bye
Practice Item:				
3	(slip)	pip	book	lip
Test Items: *(deleting the first sound)*				
6	(price)	rice	ice	moon
7	(store)	ring	tore	oar
8	(break)	ache	horse	rake
9	(spin)	pin	in	lock
10	(train)	sun	rain	aim
Practice Item:				
4	(fork)	fort	pen	four
Test Items: *(deleting the final sound)*				
11	(beach)	bee	beak	cat
12	(seed)	mouse	sea	seat
13	(tooth)	toot	fan	two
14	(plate)	play	plum	ring
15	(rose)	dog	row	rope

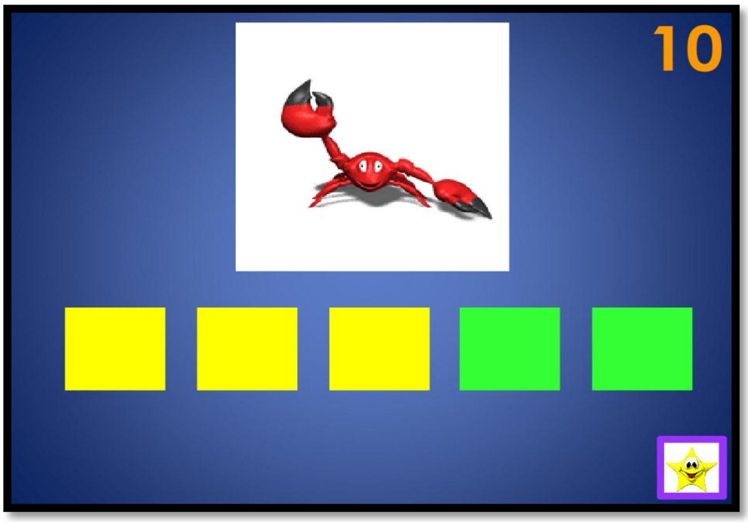
Note. Words in brackets represent the original word from which one phoneme must be deleted to formulate a new word. For example, in test item 11 participants are asked to 'Say "beach". Say "beach" again, but don't say the /ch/ sound at the end'.

2.3.6 Phoneme segmentation.

This task consists of two demonstration items, three practice items and 18 test items. Participants are required to indicate how many sounds they hear in a word by clicking one box for each sound. Up to five boxes can be selected. For example, Poe asks, 'How many sounds do you hear in the word moon?' Children have 10 seconds to click a box for each sound they hear using the computer mouse. Children can click a ‘star’ button to move to the next test item. The first five test items contain a mixture of CVC and CV words (e.g., tooth, cow, soap). The second five test items contain CCVC, CCV and CV words (e.g., crab, star, sew). The final five test items contain CVCC, CVC and CV words (e.g., bank, lock, bear). Overall, items progress in difficulty from CVC, to CCVC, and finally CVCC words. This progression is interspersed with items containing shorter or longer syllable structures to create variability between the numbers of sounds between test items. This task takes approximately four to five minutes to complete.

Example: Test item 8

Instructions: ‘This is Crab. How many sounds do you hear in the word “Crab”?’



Phoneme Segmentation			
Demonstration Items:		Number of Sounds*	Segmented Word Alphabetic (IPA**)
1	sun	3	s-u-n (sʌn)
2	bee	2	b-ee (bi)
Practice Item:			
1	dog	3	d-o-g (dɒg)
Test Items:			
1	moon	3	m-oo-n (mun)
2	tooth	3	t-oo-th (tuθ)
3	cow	2	c-ow (kaʊ)
4	cup	3	c-u-p (kʌp)
5	soap	3	s-oa-p (soʊp)
6	saw	2	s-aw (sɔ)
Practice Item:			
2	fly	3	f-l-y (flai)
Test Items:			
7	flush	4	f-l-u-sh (flʌʃ)
8	crab	4	c-r-a-b (cræb)
9	sew	2	s-ew (soʊ)
10	step	4	s-t-e-p (step)
11	plate	4	p-l-a-t(e) (pleit)
12	star	3	s-t-ar (sta)
Practice Item:			
3	hand		h-a-n-d (hand)
Test Items:			
13	bank	4	b-a-n-k (bæŋk)
14	lock	3	l-o-ck (lɒk)
15	jump	4	j-u-m-p (dʒʌmp)
16	pond	4	p-o-n-d (pɒnd)
17	bear	2	b-ear (beə)
18	cast	4	c-a-s-t (cast)

Note. * indicates the number of phonemes in the target word based on New Zealand English. For example, ‘bear’ contains two phonemes in New Zealand English (i.e., b-ear), whereas in the United States of America, ‘bear’ contains three phonemes (i.e., b-ea-r); ** IPA = International Phonetic Alphabet (International Phonetic Association – IPA, 1999).

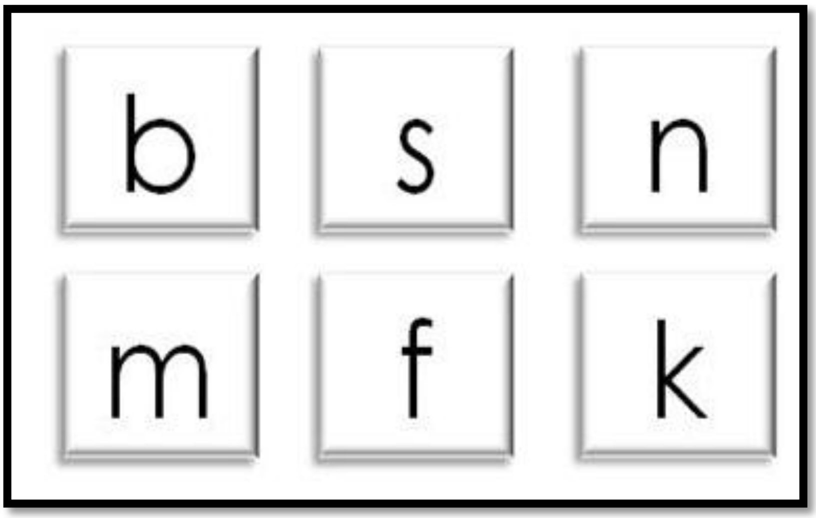
2.3.7 Letter-knowledge.

Letter-knowledge is measured using a letter-name and a letter-sound task. Both tasks are identical, except one measures letter-name and one evaluates letter-sound knowledge.

Letters are presented in lower-case century gothic font in a three-by-two grid on the computer screen. A total of 18 letter-names and letter-sounds are assessed. Children are required to click the letter-name or letter-sound requested by Poe. For example, Poe asks, ‘Show me the letter m’ or ‘Show me the letter that makes the /m/ sound’. Two practice items are presented in both letter-name and letter-sound tasks. These tasks take approximately two minutes each to complete.

Example: Test item 1 (letter-name)

Instructions: ‘Show me the letter m’



Letter-Name and Letter-Sound Knowledge					
Practice Items:					
1) j	2) r				
Test Items:					
1) m	2) s	3) k	4) b	5) n	6) f
7) d	8) h	9) p	10) t	11) w	12) g
13) c	14) z	15) l	16) q	17) v	18) y

2.4 Recognising the Limitations of Computer-Based Assessment

CBA methods are not without limitations. Two key disadvantages include the risk of technology failure and the risk of abuse. There is always the possibility that technology will experience glitches or failures when attempting to administer a CBA (Beech & Singleton, 1997; Tymms, 2001). This could manifest in a number of different ways, from hardware malfunctions, incompatibility of software, to power failures and incompatability of operating systems. Although the reliability of technology is constantly improving, the risk of technology failure provides a strong rationale for having a paper-based counterpart of a CBA readily available. Regarding the risk for abuse, CBA may give the impression of objectivity and professionalism, but may in fact fail to adhere to acceptable levels of test construction and validation standards (BPS, 1999). Therefore, the British Psychology Society (BPS, 1999) suggests that CBA should be endorsed with information on the rationale for the assessment (see Chapter 1), its appropriateness for its intended population (see Section 2.2), its validity and reliability statistics (see Chapter 6) and the administrative and interpretation guidelines. Despite the risks, it is widely acknowledged that the application of CBA to the field of education has a number of advantages for both educator and child that outweigh the potential limitations of using this technology in the classroom (Beech & Singleton, 1997; Tymms, 2001).

2.5 Application of the Computer-Based Tool in this Doctoral Research

A number of considerations must be made when introducing a new assessment instrument into the classroom environment. The aforementioned computer-based screening and monitoring tool examines PA primarily at the phoneme-level and attempts to achieve this in a time-efficient manner through the use of technology. Experiments in this thesis investigate the time-efficiency and effectiveness of the computer-based PA instrument for monitoring response to classroom instruction (see Chapters 3, 4 and 5) and seek to establish the validity and reliability of the test items (see Chapter 6). Practical considerations such as professional development in the use of the computer-based tool for classroom teachers, and the quantity of ongoing professional support will be addressed in future research projects as this is beyond the intended scope of this thesis.

2.6 Discussion

Given the importance of PA to the early identification of reading difficulties, wide scope exists for investigating how computer-based technology can assist teachers to screen and monitor children's PA ability in the classroom. The aim of developing a CBA tool in this thesis is to provide everyday teachers with a time-efficient and user-friendly method of measuring the PA capabilities of beginning readers from the outset of formal schooling. The computer-based PA screening and monitoring tool includes six PA tasks and two letter-knowledge tasks that, according to the literature, best represent PA ability, differ in difficulty and are predictive of later reading outcomes (Vloedgraven & Verhoeven, 2007). Children are able to self-administer the assessment, thereby releasing teachers to focus on instruction in the classroom. It is intended that measurement of children's PA abilities at least three times during the first year of school will help identify risk for reading difficulty (Jenkins, 2003) and shape classroom practice so that equality in reading outcomes can be achieved.

Following development of the computer-based PA assessment, investigation was required to ensure that the computer-based modality was indeed more time-efficient than was a paper-based counterpart and that it could also produce comparable scores to a conventional testing method. Chapter 3 addresses these questions as an important initial step before examining the use of the CBA in an authentic classroom environment.

Chapter 3: Time-Efficiency and Effectiveness of a Computer-Based Phonological Awareness Screening and Monitoring Tool in Comparison to a Paper-Based Counterpart: A Pilot Study

3.1 Introduction

One of the most effective methods of preventing reading difficulties is to ensure that children begin reading instruction with precursory skills that are causally linked with early literacy success (Al Otaiba et al., 2012; Bus & Van Ijzendoorn, 1999; Lonigan et al., 2008). As mentioned, phonological awareness (PA) is one skill that can alert educators to the probability of a child experiencing reading failure (Blachman et al., 2000; Kamhi & Catts, 2012). Therefore, time-efficient methods of measuring PA may help teachers achieve widespread early identification of children at risk of developing reading problems. Time-efficiency in PA assessment is important in the classroom for the following reasons: 1) to reduce demands on teacher time (Dewar & Telford, 2003), 2) to measure PA skills when they are most predictive of reading outcomes (Catts et al., 2002) and 3) to ensure children have the necessary foundational literacy skills to take advantage of classroom reading instruction (Ehri et al., 2001). Implementation of computer-based assessment (CBA) may provide one conduit through which the time-efficient screening and monitoring of PA can be achieved in the classroom. In Chapter 2, the construction of a computer-based PA assessment tool was discussed. Before introducing this instrument into the classroom, research reported in this Chapter seeks to investigate whether the computer-based PA tool is significantly more time-efficient to administer than is a paper-based counterpart, and whether the tool generates equivalent scores to paper-based testing methods.

3.1.1 Time-efficiency.

In general, CBAs are more time-efficient than are equivalent paper-based forms because the computer does the majority of the work typically required of educators, such as preparing and administering test items, recording and scoring responses and interpreting the results to inform instructional practice (Bjornsson, 2008; Martin, 2008; Ripley, 2008; Singleton et al., 1999). It is reported that teachers prefer CBA because they save time in scoring responses and have immediate access to information on child performance (Beech & Singleton, 1997). Moreover, Olsen (1990) demonstrated that a CBA (non-adaptive) was at least 25 per cent faster to administer than an equivalent paper-based form. This is significant when one considers that assessment in the classroom is estimated to consume up to one-third of teacher time (McTighe & Ferrara, 1998). It is possible that CBA will reduce the time demand on educators, while also ensuring valuable prognostic information regarding reading outcomes is collected. Time-efficiency may be further enhanced by focusing on a small number of skills that are highly predictive of early reading success, namely PA and letter-knowledge.

3.1.2 Equivalency of scores.

Establishing equivalency of scores between a CBA and its paper-based counterpart is an important issue to address if teachers are to use CBA in the classroom with confidence (Csapo et al., 2009). According to two recent meta-analyses in reading and mathematics research, computer- and paper-based testing modalities in general do not generate significantly different test scores (Wang et al., 2007; Wang et al., 2008). Further, congruency of scores between computer- and paper-based mediums has been demonstrated in small- and large-scale studies of student achievement (Csapo et al., 2009). In this chapter, establishing equivalency of scores between the computer-based PA assessment reported in Chapter 2 and its corresponding paper-based form will not only increase educator confidence in using the tool, but will also provide initial insight into the validity and reliability of the instrument. For example, comparison of scores between the computer- and paper-based versions of the tool will provide an index of test-retest reliability of assessment content. Further, comparison of computer-based scores with an existing standardised PA assessment (e.g., used to determine study eligibility) will provide information on concurrent validity.

3.1.3 Equivalency of scores for children at risk for reading difficulties.

Establishing equivalency of scores for different cohorts of children in the classroom is critical for ensuring that the computer-based instrument is useful for various subgroups of children in the classroom. For example children with typical development as a group should ideally present with equivalent scores between test modalities, and children at-risk, while presenting with lower scores overall, should also (as a group) present with equivalent between-modality scores. One cohort of children with increased risk for reading problems is those with phonologically based speech impairments (see Section 1.4.4) (Briscoe, Bishop, & Norbury, 2001; Catts et al., 2001; Gillon, 2005). This is because phonologically based speech impairments affect the phonological component of spoken language that underlies successful PA development and skilled word-recognition ability (Kamhi & Catts, 2012). Children with a phonologically based speech impairment show deficits in a number of PA tasks including syllable awareness (Gillon, 2000), rhyme awareness (Bird et al., 1995; Gillon, 2000; Webster & Plante, 1992) and phoneme identity, blending, deletion and segmentation (Bird et al., 1995).

The severity of speech impairment also influences literacy outcomes. For example, children with a severe speech impairment (e.g., percentage of consonants spoken correctly is less than 50 per cent) are at substantial risk for literacy failure (Bowen, 2009). However, even mild speech impairment is associated with reading underachievement. For example, Nathan, Stackhouse, Goulandris, and Snowling (2004) reported that mild speech impairments that persist beyond six years and nine months of age are associated with reading underachievement. Similarly, Bishops, and Adams (1990) suggest that literacy development is likely to be negatively affected if children are still unintelligible by five years and six months of age. To ensure similar time-efficiency and congruency of scores for children with and without elevated risk for reading difficulties, the experiment reported in this chapter included participants with typical spoken language development (TD) and children with moderate-severe speech delay (MSD). Children with MSD needed to present with typically developing receptive language skills. This was to ensure that task performance was reflective of deficits in the phonological domain of language, as opposed to difficulties in processing test instructions.

Therefore, the study described in this chapter reports the results of a pilot investigation to determine the time-efficiency and equivalence of scores between computer-based and paper-based testing modalities. This chapter contributes to existing research by determining whether computer-based methods of administering PA test content can be a time-efficient alternative for teachers in the classroom. This chapter contributes to this thesis by providing a rationale for using CBA as a screening and monitoring tool to evaluate response to classroom reading instruction in Chapters 4 and 5. Specifically, the following research questions and hypotheses are addressed:

1. **Time-Efficiency:** Is computer-based administration of PA test content significantly more time-efficient than a paper-based administration method?
   It was hypothesised that the computer-based administration of PA test content would take less time to administer, score and analyse than the paper-based counterpart.

2. **Effectiveness:** Does administration of PA test content via a computer-based modality generate equivalent test scores to administration using a paper-based modality for children with TD and MSD?
   It was hypothesised that computer-based administration would produce similar scores to a paper-based version of the assessment for children with TD and MSD.

3.2 Method

3.2.1 Participants.

Thirty-three New Zealand children (15 boys; 18 girls) aged between four years and 10 months and five years zero months (*M* age = four years and 11 months, *SD* = 0.57 months) participated in this study. Twenty-one children presented with TD and 12 children presented with MSD, with receptive language ability within normal limits. Eight early-childhood centres from low, middle and high socio-economic areas within the Canterbury region were invited to refer children to the study. Children were required to meet the following inclusion criteria: 1) be aged between four years and 10 months and five years zero months, 2) present with TD or MSD, 3) have no current or previous history of sensory, neurological, physical or intellectual difficulties, 4) speak New Zealand English as their first and only language and 5) have parental permission to participate in the study. All participants were pre-enrolled to begin formal schooling around their fifth birthday.

3.2.1.1 Baseline assessment to satisfy study inclusion criteria.

Thirty-eight children were referred to the study. Children had to present with either TD or MSD to meet the inclusion criteria. For a classification of TD, children needed to score within the average range on standardised measures of language, phonology, intelligence and PA. For a classification of MSD, children needed to produce less than 65 per cent of consonants correctly on a single-word articulation test and perform within the average range on two standardised measures of receptive language ability. Children referred to the study participated in a series of baseline assessments to satisfy inclusion criteria prior to participating in experimental phases. Baseline assessments to meet study inclusion criteria are as follows:

1) **Language:** The *Clinical Evaluations of Language Fundamentals Preschool—Second Edition* (Australian and New Zealand Edition) (CELF-P2) (Wiig, Secord, & Semel, 2006) was administered to obtain a detailed profile of receptive and expressive spoken language skills. This test is suitable for children who are destined for, or have already entered, an academic environment. It provides normative data for children aged three years and zero months to six years and 11 months. ***Content:*** The CELF-P2 is comprised of six core subtests: three assessing receptive language ability and three assessing expressive language ability. The receptive language subtests evaluate children's understanding of sentence structure, concepts and following directions, and basic concepts. The expressive language subtests measure word structure, expressive vocabulary and recall of sentences. The six core subtests are collated into different combinations to produce the following data: a core language score (e.g., sentence structure, word structure and expressive vocabulary), a receptive language index (e.g., sentence structure, concepts and following directions, and basic concepts), an expressive language index (e.g., word structure, expressive vocabulary and recall of sentences), a language content index (e.g., expressive vocabulary, concepts and following directions, and basic concepts) and a language structure index (e.g., sentence structure, word structure and recall of sentences). Supplementary measures including a PA task, pre-literacy scale and pragmatic profile are also provided. ***Reliability and validity:*** Test-retest reliability for the CELF-P2 is satisfactory, with correlation coefficients for the six core subtests ranging from

excellent (0.90) to adequate (0.78) across all ages. Measures of internal consistency are satisfactory, ranging from 0.80 to 0.96 across the subtests. Correlations between the CELF-P2 and its Australian equivalent, the CELF-4, revealed moderate to high coefficients ranging from 0.61 to 0.86 between subtests. ***Determining study eligibility:*** The six core subtests of the CELF-P2 were individually administered to the 38 children referred to this study. Children who performed within the average range in each of the six core subtests were accepted into the study as participants with TD. Children with TD also had to present with typical speech sound development. Children who obtained a receptive language index score (e.g., an aggregation of sentence structure, concepts and following directions, and basic concepts) within the average range (e.g., 85–115, where *M* = 100 and *SD* = +/-15) but produced less than 65 per cent of consonants correctly on a single-word articulation test, as outlined below, were accepted into the study as participants with MSD.

2) **Speech sound development:** The *New Zealand Articulation Test* (NZAT) (MOE, 2004) was administered to measure children's speech sound development. Analysis of the NZAT was supported using Computerised Profiling of Phonology (PROPH) Software (Long, Fey, & Channell, 2002). The NZAT is appropriate for children aged five years and zero months to eight years and 11 months. ***Content:*** The NZAT consists of five subtests: single consonant sounds, initial consonant blends, vowels, multi-syllabic words and a conversational speech sample. The single consonant sounds subtest and the initial consonant blends subtest require children to articulate consonants in a single word by naming a picture presented by an examiner. These are the only subtests that are normed. Entering data into the PROPH software provides additional information including syllable structure

analysis, percentage of consonants correctly articulated, stress patterns, phoneme age of mastery, phonological mean length of utterance and phonological process analysis. ***Reliability and validity of NZAT:*** Reliability of the NZAT single consonant subtest and initial consonant blends subtest is provided using inter-rater reliability, test-retest reliability and internal reliability. Inter-rater reliability between two individuals experienced in phonetic transcription was 98 per cent for single consonants in the initial, medial and final position of single words and 92 per cent for initial consonant blends. Measures of internal consistency produced reliability coefficients ranging from 0.84 to 0.95. ***Determining study eligibility:*** The single consonant sound subtest and the initial consonant blends subtest were individually administered to the 38 children referred to this study. The single consonants sound subtest consisted of 59 picture stimuli to evaluate the production of 23 consonant sounds in the initial, medial and/or final position of words. The initial consonant blends subtest consisted of 23 picture stimuli to evaluate the production of 23 initial consonant blends. Responses were interpreted using the test norms and the results from the PROPH analysis. Percentage of consonants correct (PCC) was the primary measure used to determine study eligibility and subsequent classification as either TD or MSD. PCC is the number of consonants correctly produced divided by the total number of consonants produced. Research suggests that the average PCC for children aged four years and zero months to four years and 11 months is 92.7 per cent (Shriberg, Austin, Lewis, & McSweeny, 1997). Severity classifications by Shriberg and Kwiatkowski (1982) were used for children with PCC scores below 93 per cent, as follows: PCC 85–92 per cent = mild; PCC 65–85 per cent = mild to moderate; PCC 50–65 per cent = moderate to severe; and PCC < 50 per cent =

severe. Thus, children were accepted into the study as participants with TD if they had a PCC score of 93 per cent or greater, or as participants with MSD if they had a PCC score of less than 65 per cent.

3) **PA:** The *Preschool and Primary Inventory of PA* (PIPA) (Dodd et al., 2000) was used to obtain a standardised profile of PA ability. This test is suitable for children aged three years and zero months to six years and 11 months. ***Content:*** The PIPA measures syllable segmentation, rhyme awareness, alliteration awareness, phoneme isolation, phoneme segmentation and letter-knowledge ability. It provides normative data for Australian and British children. ***Reliability and validity:*** Reliability measures of internal consistency, test-retest reliability and inter-rater reliability are provided. Internal consistency reliability coefficients are acceptable and range from 0.70 for phoneme segmentation to 0.98 for letter-knowledge. Test-retest reliability using Pearson product-moment correlation coefficients are reported to be significant for all six subtests ranging from 0.33 for phoneme segmentation to 0.98 for letter-knowledge. Comparison of two examiners' scores on the PIPA did not reveal any significant differences, indicating adequate inter-rater reliability. Significant correlations between the rhyme awareness (0.63) and letter-knowledge (0.92) subtests of the PIPA and the *Phonological Awareness Test* (PAT) (Robertson & Salter, 1997) provide evidence of concurrent validity. Evidence of criterion validity is reported in the form of significant correlations between the performance of 30 children with speech disorders on phoneme isolation (0.39) and alliteration awareness (0.50) tasks in the PIPA and the *Test of Early Reading Ability* (TERA) (Reid, Hresko, & Hammill, 1989). ***Use of the PIPA in this study:*** It was anticipated that a spectrum of strong to weak PA abilities would be evident between children with

TD and MSD. The PIPA was used to profile these differences and to provide an initial step towards establishing concurrent validity with the computer-based PA tool developed in this thesis.

4) **Vocabulary:** The *Peabody Picture Vocabulary Test—Fourth Edition* (PPVT-4) (Dunn & Dunn, 2007) was administered to detail receptive vocabulary ability. This test is suitable for children aged two years and six months through to 90 years of age. ***Content:*** The PPVT-4 comprises of 228 test items divided into 19 subsets, each containing 12 test items. Test items are presented with four pictures on a page, from which children are required to point to the item that represents the word spoken by the examiner. There are two parallel forms of the test: Form A and Form B. ***Reliability and validity:*** Reliability of the PPVT-4 is adequate, with split-half reliability coefficients ranging from 0.94 to 0.95 for Form A and Form B, respectively. Similarly, internal consistency alpha coefficients ranged from 0.97 for Form A to 0.96 for Form B. Further, the average test-retest correlation coefficient was 0.93. A number of measures of validity are provided, including concurrent validity, with satisfactory correlations between the PPVT-4 and the *Expressive Vocabulary Test* (0.82), the *Comprehensive Assessment of Spoken Language* (0.58), the *Clinical Evaluations of Language Fundaments*-4 (0.74) and the PPVT-III (0.84) having been reported. ***Determining study eligibility***: The PPVT-4 Form A was administered to children referred to this study. Children who performed within or above the average range (i.e., above 85, where M = 100 and SD = +/-15) were accepted into the study as participants with TD. Children with TD also had to perform within or above the average range on the Receptive Language Index of the CELF-P2 and above 93 per cent PCC on the NZAT and

PROPH analysis. Children who obtained a score less than 85 (i.e., one standard deviation below the mean) were excluded from the study.

5) **Non-verbal intelligence:** The *Primary Test of Non-verbal Intelligence* (PTONI) (Ehrler & McGhee, 2008) was used to obtain a measure of non-verbal intellectual ability. This test is appropriate for children aged three years and zero months to nine years and 11 months. ***Content:*** The PTONI requires children to examine pictures on a page and then point to one picture that does not belong. The test progresses in difficulty, beginning with lower-order reasoning skills, such as visual and spatial recognition, and moving towards more advanced reasoning skills, such as sequential reasoning and categorical formulation. ***Reliability and validity:*** Internal consistency reliability coefficients using Cronbach's alpha are greater than 0.90 from ages three to nine years. Test-retest reliability for a sample of 94 children was excellent at 0.97. Content validity was established by demonstrating that all of the 75 test items met acceptable standards for item discrimination and item difficulty. ***Determining study eligibility:*** The PTONI was used to determine whether children referred to the study met the criteria of presenting with no current or previous history of intellectual difficulties. Children were required to perform within or above the average range to participate (e.g., 85–115, where M = 100 and SD = +/-15).

6) **History:** A parental questionnaire was distributed to all parents who gave permission for their child to participate in the study. The questionnaire contained questions regarding the physical, neurological, sensory, intellectual and educational (e.g., speech-language therapy) aspects of their child's development and was completed on a voluntary basis. This was to satisfy inclusion criteria that

required participants to present with no current or previous history of physical, neurological, sensory or intellectual difficulties.

Of the 38 children referred to the study, 33 met the selection criteria. From this group, 21 children presented with TD and 12 children presented with MSD. Twenty-six children were of New Zealand European descent, and seven children were of Māori decent. Twenty-one per cent of the children were from early education centres in low socio-economic areas, 36 per cent from middle socio-economic localities and 43 per cent from high socio-economic areas. In New Zealand, the socio-economic status of a school and its local area is defined by a decile ranking. Decile rankings range from 1 (i.e., the lowest ranking, indicating low socio-economic status) to 10 (i.e., the highest ranking, indicating high socio-economic status). Decile rankings are calculated based on socio-economic factors including household income, occupations, household crowding and income support collected during national census periods (MOE, 2011). All schools in New Zealand are assigned a decile ranking. For the purposes of this thesis, a decile ranking from 1–4 is considered low, a ranking from 5–7 is considered medium and a ranking from 8–10 is considered high.

Five children were excluded from the study because their receptive language performance was at least one standard deviation below the mean on the Receptive Language Index of the CELF-P2 and/or the Peabody Picture Vocabulary Test. This meant that they did not meet the criteria for TD development or MSD with receptive language abilities within normal limits.

Table 3.1 profiles the language, speech and PA abilities of the children with TD (n = 21) and the children with MSD (n = 12) who met the study inclusion criteria.

Table 3.1

Baseline verbal and non-verbal measures to address study inclusion criteria

	Age (months)	PTONI	PPVT-4	CELF-P2: RLI	PCC	PIPA		
						RA	IPI	LS
TD (*n* = 21)								
M	58.67	106.14	105.05	107.14	97.34*	9.4*	10.5*	8.1*
SD	0.48	7.99	8.78	9.92	3.30	2.34	3.10	1.65
Range	4;10–4;11	89–120	87–119	89–123	92.7–100	5–12	5–15	6–11
MSD (*n* = 12)								
M	58.83	105.83	104.08	102.5	60.98*	4.25*	4.92*	3.92*
SD	0.72	6.31	7.30	6.30	8.37	1.29	1.83	.90
Range	4;10–5;0	86–109	87–108	86–101	34.6–64.5	3–7	3–9	3–5

Note. PPVT-4 = Peabody Picture Vocabulary Test—Fourth Edition standard scores (*M* = 100, *SD* +/- 15) (Dunn & Dunn, 2007); CELF-P2 RLI = Clinical Evaluations of Language Fundamentals Preschool–2 Receptive Language Index (*M* = 100, *SD* +/-15) (Wiig et al., 2006); PCC = Percentage of Consonants Correct score from the NZAT (MOE, 2004) analysed using PROPH (Long et al., 2002); PTONI = Primary Test of Non-verbal Intelligence standard scores (*M* = 100, *SD* +/-15) (Ehrler & McGhee, 2008); PIPA = Preschool and Primary Inventory of PA standard scores (*M* =10, *SD* +/-3) (Dodd et al, 2000); RA = Rhyme Awareness, IPI = Initial Phoneme Identity, LS = Letter-Sound Knowledge; *significant difference between children with TD and children with MSD, $p < .001$.

Baseline testing did not reveal any significant differences between children with TD and children with MSD on measures of chronological age ($t(31) = 0.79$, $p = .43$) and receptive language ability on the PPVT-4 ($t(31) = 0.32$, $p = .75$) and CELF-P2 ($t(31) = 1.45$, $p = .16$). A significant group difference was identified for PCC articulated on a single-word articulation test ($t(31) = 17.81$, $p < .0001$). The children with TD produced 97.34 per cent ($SD = 3.30$) of consonants correctly, while children with MSD produced 60.98 per cent ($SD = 8.37$) of consonants correctly. The PCC and common error patterns made by children with MSD are profiled in Table 3.2. These error patterns primarily reflect delayed development. However, one participant showed signs of disordered phonological development.

All speech samples were collected using an Olympus Digital Voice Recorder (VN-2100PC) and transcribed using broad phonetic transcription techniques by the author and an independent examiner. The independent examiner was experienced in phonetic transcription and was blinded to the children's group status. Inter-rater reliability of the speech samples was 98 per cent. Discrepancies were resolved through collaborative listening of samples leading to 100 per cent agreement prior to data being entered into Computerised Profiling Software (Long et al., 2002). As anticipated, children with TD performed significantly higher on measures of rhyme awareness ($t(31) = 7.01$, $p < .0001$), initial phoneme identity ($t(31) = 5.67$, $p < .0001$) and letter-sound knowledge ($t(31) = 8.08$, $p < .0001$).

Table 3.2

Speech error patterns of children with moderate-severe speech delay

	Age	PCC	Gender	Examples of speech error patterns	Delay/Disorder
Group A					
1	4;11	64.2	M	Substitutions 97%; Omissions (on cluster element) 3% (e.g., velar fronting)	Delay
2	4;11	64.5	M	Substitutions 85%; Omissions 15% (e.g., velar fronting, stopping, lateralisation of /s/, /z/, /sh/ and /ch/)	Delay & Disorder
3	4;11	63.1	F	Substitutions 95%; Omissions (on cluster element) 5% (e.g., stopping, cluster reduction)	Delay
4	4;11	62.7	F	Substitutions 92%; Omissions 8% (e.g., velar fronting)	
5	5;00	64.2	F	Substitutions 88%; Omissions (on cluster element) 12% (e.g., velar fronting)	Delay
6	4;10	61.3	F	Substitutions 79%; Omissions 21% (e.g., palatal fronting, velar fronting)	Delay
Group B					
1	4;11	64.6	M	Substitutions 96.2%; Omissions 3.8% (e.g., stopping)	Delay
2	5;00	62.3	M	Substitutions 91.3%; Omissions 8.7% (e.g., final consonant deletion)	Delay
3	4;11	63	M	Substitutions 92%; Omissions 8% (e.g., velar fronting, cluster reduction, final consonant deletion)	Delay
4	4;11	34.6	F	Substitutions 74.2%; Omissions 25.8 (e.g., final consonant deletion)	Delay
5	5;00	63.3	F	Substitutions 69.7%; Omissions 30.3% (e.g., palatal fronting, velar fronting)	Delay
6	4;10	64	F	Substitutions 89.5%; Omissions 10.5% (e.g., stopping, velar fronting)	Delay

Note. Each numeral under 'Group A' and 'Group B' represents a participant with MSD; Age is represented as years; months; PCC = percentage of consonants correctly articulated in the NZAT (MOE, 2004) analysed using PROPH (Long et al., 2002); Gender is represented by M = male and F = female; Substitutions = the percentage of speech errors that involves substituting one sound for another; Omissions = the percentage of speech errors that involved deleting a sound from a word.

3.2.2 Procedures.

A crossover research design was used to compare time-efficiency and congruency of scores between the computer-based PA screening and monitoring tool and an identical paper-based counterpart. Crossover research designs allow for control of order effects by systematically varying exposure to each experimental condition (Portney & Watkins, 2009). Children with TD and children with MSD were randomly allocated to either Group A or Group B experimental assessment conditions. Group A consisted of 10 children with TD and six children with MSD, while Group B consisted of 11 children with TD and six children with MSD. Table 3.3 illustrates that there were no significant differences between Groups A and B on measures of age ($t(31) = 0.05$, $p = .96$), gender ($t(31) = 0.19$, $p = .85$), speech sound development ($t(31) = 0.10$, $p = .92$) or receptive language ability on the PPVT-4 ($t(31) = 0.29$, $p = .78$) and CELF-P2 Receptive Language Index ($t(31) = 17.96$, $p = .86$). Further, no significant between-group differences were identified on measures of rhyme awareness ($t(31) = 0.13$, $p = .89$), initial phoneme identity ($t(31) = 0.11$, $p = .91$) or letter-sound knowledge ($t(31) = 0.02$, $p = .98$).

Table 3.3

Group equivalence on speech, language and phonological awareness measures

	Age (months)	PPVT-4	CELF-P2: RLI	PCC	PIPA		
					RA	IPI	LS
Group A (*n* = 16)							
M	58.81	105.13	105.75	84.47	7.5	8.44	6.63
SD	0.54	7.20	8.64	17.06	3.01	3.78	2.58
Range	58–60	87–110	86–123	61–100	3–12	3–15	3–11
Group B (*n* = 17)							
M	58.82	104.29	105.18	83.79	7.35	8.59	6.65
SD	0.62	9.20	9.39	20.48	3.43	3.97	2.55
Range	58–60	86–112	87–121	35–100	3–14	3–15	3–11

Note. PPVT-4 = Peabody Picture Vocabulary Test—Fourth Edition standard scores (*M* = 100, *SD* = +/- 15) (Dunn & Dunn, 2007); CELF-P2 = Clinical Evaluations of Language Fundamentals Preschool–2 Receptive Language Index (*M* = 100, *SD* +/- 15) (Wiig et al., 2006); PCC = Percentage of Consonants Correct score from the NZAT (MOE, 2004) analysed using PROPH (Long et al., 2002); PIPA = Preschool and Primary Inventory of PA standard scores (*M* = 10, *SD* +/-3) (Dodd et al., 2000), RA = Rhyme Awareness, IPI = Initial Phoneme Identity; LS = Letter-Sound Knowledge.

Experimental phases: There were two experimental assessment phases. During the first assessment phase (P1), Group A received a paper-based version of the PA assessment, administered by an examiner, while Group B received the same assessment content, but delivered by a computer program. The second assessment phase (P2) took place two weeks later, at which time Group A received the CBA and Group B received the paper-based version of the assessment. Figure 3.1 illustrates the crossover research design encompassing the two experimental assessment phases.

	Phase 1 (P1)		**Phase 2 (P2)**
Group A **(*n* = 16)**	**PAPER**  Is paper-based testing less time-efficient?	Are computer- and paper-based modalities strongly correlated? ↔	**COMPUTER**  Is CBA more time-efficient?
	↕ Do computer- and paper-based modalities produce significantly different results at P1?		↕ Do computer- and paper-based modalities produce significantly different results at P2?
Group B **(*n* = 17)**	**COMPUTER**  Is CBA more time-efficient?	Are computer- and paper-based modalities strongly correlated? ↔	**PAPER**  Is paper-based testing less time-efficient?

Figure 3.1. Crossover research design investigating time-efficiency and equivalence of scores between computer- and paper-based modalities

Computer-based and paper-based assessment modalities: There were two assessment modalities under investigation in this experiment: a computer-based modality and a paper-based modality. Tasks from the computer-based PA screening and monitoring tool described in Chapter 2 were selected based on their appropriateness for children four years and 10 months to five years and zero months of age. These tasks included rhyme oddity, initial phoneme identity, letter-name recognition and letter-sound knowledge. In this modality, the computer presented all test items visually (e.g., pictures) on a laptop screen, along with pre-recorded verbal instructions. Participants were required to respond independently by clicking a computer mouse or, if they chose to do so, point to the screen and have a researcher click the mouse on their behalf.

In experimental assessment P1, three children in Group B selected to point to the computer screen rather than use the computer mouse. In P2, two children in Group A chose to point to the computer screen. Although numbers were small, the average performance of children who selected to point to the screen was compared to children who used the computer mouse. Results showed no significant difference in performance on measures of rhyme oddity ($t(31) = 0.11$, $p = .91$), initial phoneme identity ($t(31) = 0.16$, $p = .88$), letter-name recognition ($t(31) = 0.21$, $p = .84$) or letter-sound knowledge ($t(31) = 0.18$, $p = .86$). The computer recorded, stored and scored all responses. The CBA was administered on the same SONY VAIO (VGN-NS15G) laptop for all participants. The CBA used in this study was non-adaptive. Chapter 2 provides a detailed description of the rhyme oddity, initial phoneme identity and letter- knowledge tasks chosen for this study (see Section 2.3) and Chapter 6 provides information on the validity and reliability of these tasks.

The paper-based method of assessing PA ability involved presenting children with picture cards for rhyme oddity and initial phoneme identity tasks, as well as groups of six letters on A4-sized sheets for letter-knowledge tasks. Pictures and letter-fonts were identical between the computer-based and paper-based conditions. The examiner had to present the test stimuli (e.g., picture or letter cards) in a prescribed order, which was identical to the CBA, while simultaneously scoring responses. The same examiner completed all paper-based administrations with each participant and had their voice recorded as the computer-based verbal instructions. This was to help minimise external influences between computer-based and paper-based assessment modalities.

Computer-based assessment and paper-based assessment content: As mentioned, the computer- and paper-based modalities included tasks of rhyme awareness, phoneme awareness and letter-knowledge. Test stimuli were identical between modalities, except for the use of animated graphics in the computer-based modality for the main character, 'Poe', and her friends. All test images were identical between the computer- and paper-based versions of the assessment. All test items were administered using a multiple-choice format with an array of three response options (i.e., one correct and two distractors) for each of the rhyme oddity and initial phoneme identity tasks, and six response options (e.g., one of six letters) for letter-name and letter-sound knowledge tasks. A brief description of the tasks used in this experiment is provided below:

1) **Initial phoneme identity:** Participants were asked to select a word (represented by a picture) from a choice of three words that started with the same initial phoneme as the name of one of Poe's friends. For example, the computer or an examiner showed the participant a character picture of a Bee and said, 'This is my friend Bee. Bee starts with a /b/ sound. What word starts with a /b/ sound? /car, boat, shoe/'. Pictures of these words then appeared on the computer screen, or picture cards were placed on the table in front of the participant by the examiner. The participants were required to click or point to the picture that started with the same initial sound as Bee. This task consisted of two practice items and 10 test items and took five to six minutes to administer.

2) **Rhyme oddity:** Participants were required to indicate which word from a choice of three did not rhyme with the remaining two words (each word represented by a picture). For example, the computer or examiner showed the participant three

pictures and said, ‘Which word does not rhyme: book, hook, sail?’ The participants were asked to click or point to the picture that did not rhyme with the other two pictures. This task consisted of two practice items and 10 test items and took approximately three to four minutes to administer.

3) **Letter-knowledge:** Letter-knowledge was assessed using a letter-name and a letter-sound task. In both tasks, either the computer or an examiner presented six letters in a three-by-two grid on the computer screen or tabletop. The participants were required to click or point to the letter-name or letter-sound requested by the computer or examiner. For example, in the letter-name subtest, the computer or examiner asked the participant to ‘Show me the letter m’. Eighteen letter-names and letter-sounds were assessed in lower-case century gothic font. Two practice items were presented in both tasks and it took approximately two minutes to administer each task, for a total of four minutes.

3.2.3 Reliability.

In the CBA modality, reliability data was collected in three ways: 1) the computer recorded all responses into a database, 2) an examiner scored responses in real-time using pen and paper and 3) video recordings captured each assessment from start to finish. In the paper-based assessment modality, reliability data was collected using: 1) an examiner who scored responses in real-time and 2) video recordings of all assessments. Twenty-five per cent of data was selected from each modality and was reviewed by an independent examiner who was blind to participants’ group status. First, the reliability of scores between different methods of recoding data was reviewed. The percentage of agreement between the three methods of data recording in the computer-based modality was 100 per cent. Similarly, the percentage of agreement between real-time examiner scores and video recordings in the paper-based modality was 100 per cent. Next, the time taken to administer the computer-based and paper-based modalities was compared, with 100 per cent inter-rater agreement being achieved for the duration of tasks across both modalities using video observations.

3.3 Results

Independent group t-tests were used to identify significant differences in time-efficiency and to establish equivalency of scores between computer- and paper-based assessment modalities. A significant difference between the paper- and computer-based modalities, in which the computer-based modality was shown to be significantly faster than the paper-based counterpart, was required to prove the time-efficiency of the CBA. In contrast, no significant difference between scores on the computer- and paper-based modalities was required to establish equivalency of scores between these two testing modalities.

Pearson product-moment correlation coefficients were calculated to examine the consistency of responses between experimental assessment P1 and P2. Statistical Package for the Social Sciences (SPSS) software (Version 19.0) was used to analyse study data. Effect size indices using Cohen's *d* were calculated for statistically significant results. Cohen's *d* is calculated as the difference between the groups' means divided by the root mean square of the groups' standard deviations (Portney & Watkins, 2009). Conventional values for Cohen's *d* are small effect size $d = 0.20$, medium effect size $d = 0.50$ and large effect size $d = 0.80$ (Cohen, 1988).

3.3.1 Time-efficiency of the computer-based phonological awareness assessment.

Each task took significantly less time to administer using the computer-based administration modality. Specifically, the following t-test results were obtained: rhyme oddity ($t(64) = 71.61$, $p < .0001$, $d = 1.91$), initial phoneme identity ($t(64) = 39.37$, $p < .0001$, $d = 2.84$), letter-name ($t(64) = 116.49$, $p < .0001$, $d = 2.12$) and letter-sound knowledge ($t(64)$

= 123.54, *p* < .0001, *d* = 3.89). The average total testing time using the computer was 13 minutes and 20 seconds compared to 19 minutes and 20 seconds when using the paper-based modality. In total, the computer-based version of the assessment took 31 per cent less time to administer when compared to the paper-based version. This represents a statistically significant difference in total administration time (*t*(64) = 238.38, *p* < .0001, *d* = 5.60). It is important to note that these results represent the amount of child time required to complete each assessment. These results do not include the time a teacher could save while a child is independently administering the CBA or the time saved by not having to organise paper materials prior to assessment, score results during the assessment or analyse the results post-assessment. Table 3.4 profiles the average administration time per task and the average total administration time for the child. An approximation of the amount of teacher administration time required for testing in a classroom context is also provided.

Table 3.4

Administration time for computer-based and paper-based modalities

	Average administration time minutes and seconds (time in seconds)				Total (average) administration time for child (seconds)	Class example: Total (average) administration time for teachers
	RO	IPI	LN	LS		
Computer	3;30 (210)	5;35 (335)	2;05 (125)	2;10 (130)	13 minutes; 20 seconds (800)	0 minutes; 0 seconds
Paper-based	5;00 (300)	6;10 (370)	4;20 (260)	3;50 (230)	19 minutes; 20 seconds (1, 160)	19 minutes; 20 seconds
Significance (*p*)	<.0001	<.0001	<.0001	<.0001	<.0001	

Note. RO = rhyme oddity; IPI = initial phoneme identity; LN = letter-name knowledge; LS = letter-sound knowledge. The hypothetical class example does not include time required for the teacher to turn on the computer and open the assessment software.

Comparison of these results to the selection of PA tools available to classroom educators profiled in Table 1.2 (see Chapter 1), illustrates that the computer-based PA screening and monitoring tool offers a time-efficient alternative when paucity of time is an issue in the classroom. An adapted version of Table 1.2 is provided as part of a general discussion in Chapter 7 (i.e., Table 7.1) to demonstrate how these results, in addition to other thesis results, contribute to gaps in the existing research base. It is important to note that only four tasks from the computer-based tool were included in the experiment reported in this chapter, as opposed to the entire assessment of eight tasks. Moreover, given that this study is pilot in nature, it is important to interpret these results in a conservative manner.

3.3.2 Equivalency of scores between computer-based and paper-based modalities.

Independent group t-tests were used to compare performance on computer-based and paper-based modalities during P1 and P2.

P1: In P1, participants in Group A received the paper-based version of the assessment and participants in Group B received the computer-based counterpart. Results revealed no significant differences between the scores obtained by the two groups, irrespective of modality of assessment. Specifically, the following t-test results were obtained: rhyme oddity ($t(31) = 0.05$, $p = .96$), initial phoneme identity ($t(31) = 0.64$, $p = .52$), letter-name knowledge ($t(31) = -0.72$, $p = .48$) and letter-sound knowledge ($t(31) = -0.31$, $p = .76$).

P2: In P2, participants in Group A received the CBA modality and participants in Group B received the paper-based counterpart. Consistent with P1, congruent scores between the two assessment modalities were identified. The following t-test results were obtained: rhyme oddity ($t(31) = -0.12$, $p = .91$), initial phoneme identity ($t(31) = 0.44$, $p = .67$), letter-name knowledge ($t(31) = 0.62$, $p = .54$) and letter-sound knowledge ($t(31) = -0.44$, $p = .67$).

These results provide evidence of equivalency of scores between computer- and paper-based administration modalities. Figures 3.2 and 3.3 illustrate the comparable scores obtained between the two assessment modalities during P1 and P2.

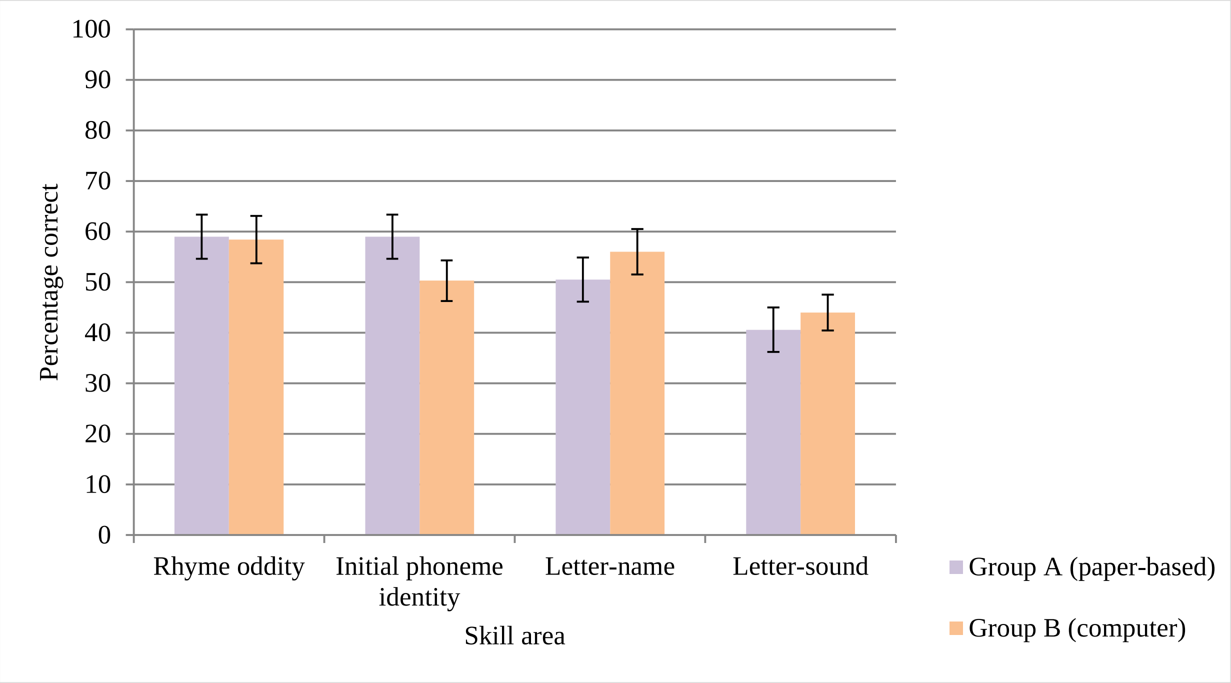


Figure 3.2. Equivalency between computer-based and paper-based scores in P1

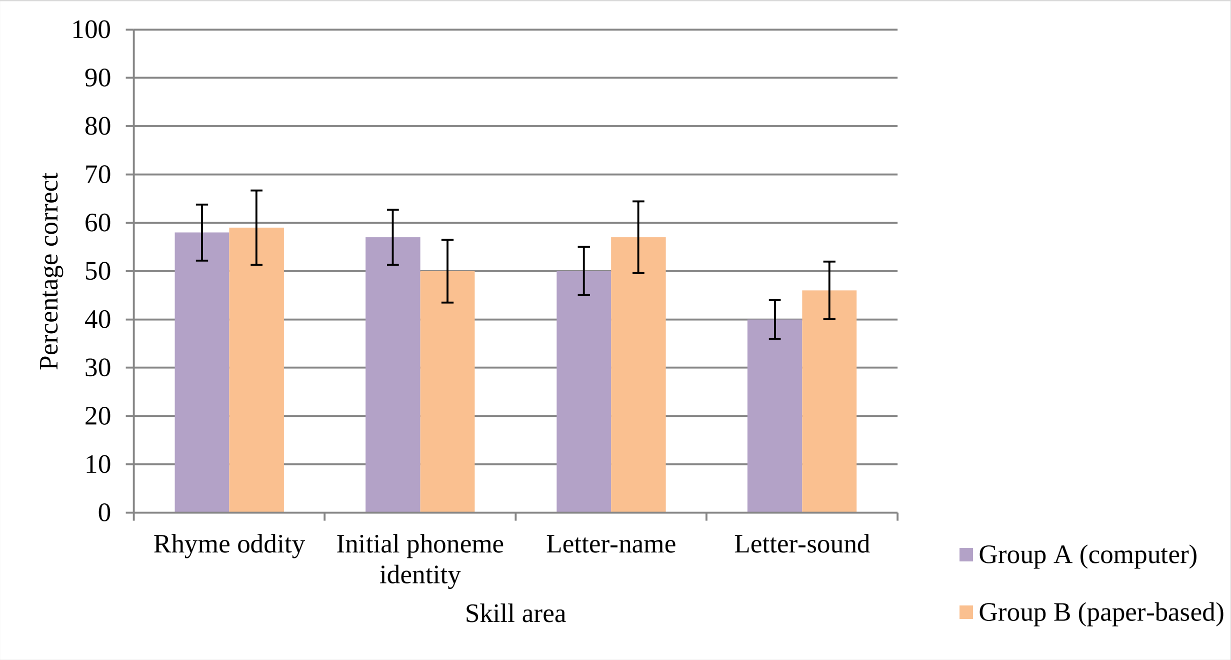


Figure 3.3. Equivalency between computer-based and paper-based scores in P2

3.3.3 Children with moderate-severe speech delay.

Children with MSD are at risk of having trouble in PA development due to deficiencies in processing the phonological component of language (Gillon, 2005; Nelson, 2010). This can negatively affect the acquisition of good word-recognition skills (Al Otaiba et al., 2012; Catts et al., 1999). Therefore, it is important that consistent scores are obtained between paper-based and computer-based assessment methods for this cohort, particularly if CBA is to be reliability used within the classroom environment. Independent group t-tests between participants with MSD in Group A (n = 6) and participants with MSD in Group B (n = 6) did not show a significant difference during P1. T-test results for each task were as follows: rhyme oddity ($t(10) = 0.24$, $p = .81$), initial phoneme identity ($t(10) = 0.48$, $p = .64$), letter-name knowledge ($t(10) = 0.19$, $p = .86$) and letter-sound knowledge ($t(10) = 0.06$, $p = .96$). Similarly, no significant between-group differences for children with MSD were identified during P2, as demonstrated by the following t-test results: rhyme oddity ($t(10) = 0.26$, $p = .80$), initial phoneme identity ($t(10) = 0.54$, $p = .60$), letter-name knowledge ($t(10) = 0.09$, $p = .93$) and letter-sound knowledge ($t(10) = 0.37$, $p = .72$). Thus, children with MSD obtained similar scores irrespective of the assessment modality used. Figure 3.4 illustrates congruency of scores between computer-based and paper-based administration modalities for children with MSD in Groups A and B during P1.

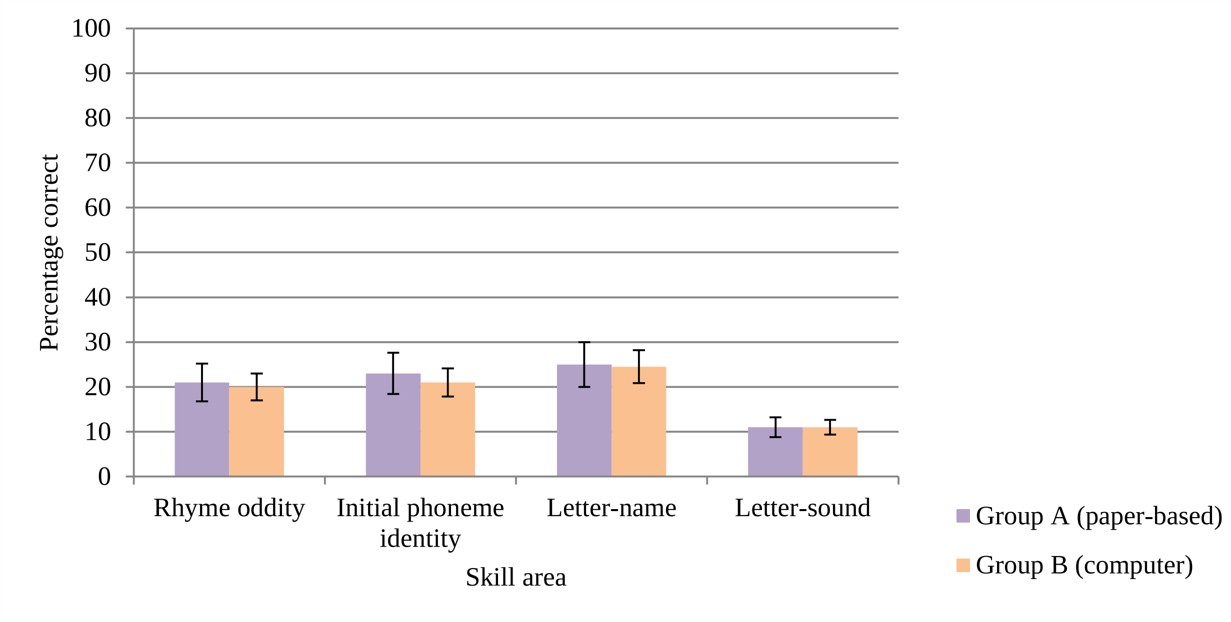


Figure 3.4 Equivalency of scores for children with MSD in Groups A and B in P1

3.3.4 Initial evidence of reliability and validity for the computer-based phonological awareness tool.

Reliability of PA assessment content: Pearson product-moment correlation coefficients revealed strong correlations between each participant's performances from P1 and P2, regardless of a change in the modality of administration. Specifically, correlation coefficients for each task were as follows: rhyme oddity ($r(31) = 0.95$, $p < .001$), phoneme identity ($r(31) = 0.94$, $p < .001$), letter-name knowledge ($r(31) = 0.99$, $p < .001$) and letter-sound ability ($r(31) = 0.99$, $p < .001$). These results not only indicate that the computer- and paper-based modalities are highly correlated, but also that there is high test-retest reliability of assessment content.

Concurrent validity of the CBA: Positive correlations between the PIPA, which was administered at baseline, and CBA results were identified on measures of rhyme oddity ($r(31) = 0.83$, $p < .001$), initial phoneme identity ($r(31) = 0.88$, $p < .001$) and letter-sound knowledge ($r(31) = 0.89$, $p < .001$). These tasks were compared because they are common to both the PIPA and the computer-based PA tool developed for this thesis. The PIPA is a standardised and paper-based PA assessment. Therefore, achieving significant and positive correlations between this assessment and the CBA provides initial evidence of concurrent validity for the CBA. Detailed evidence of the validity and reliability of the CBA following implementation with a larger sample is provided in Chapter 6. Figure 3.5 illustrates the correlation between the initial phoneme identity scores on the PIPA and initial phoneme identity scores on the CBA task.

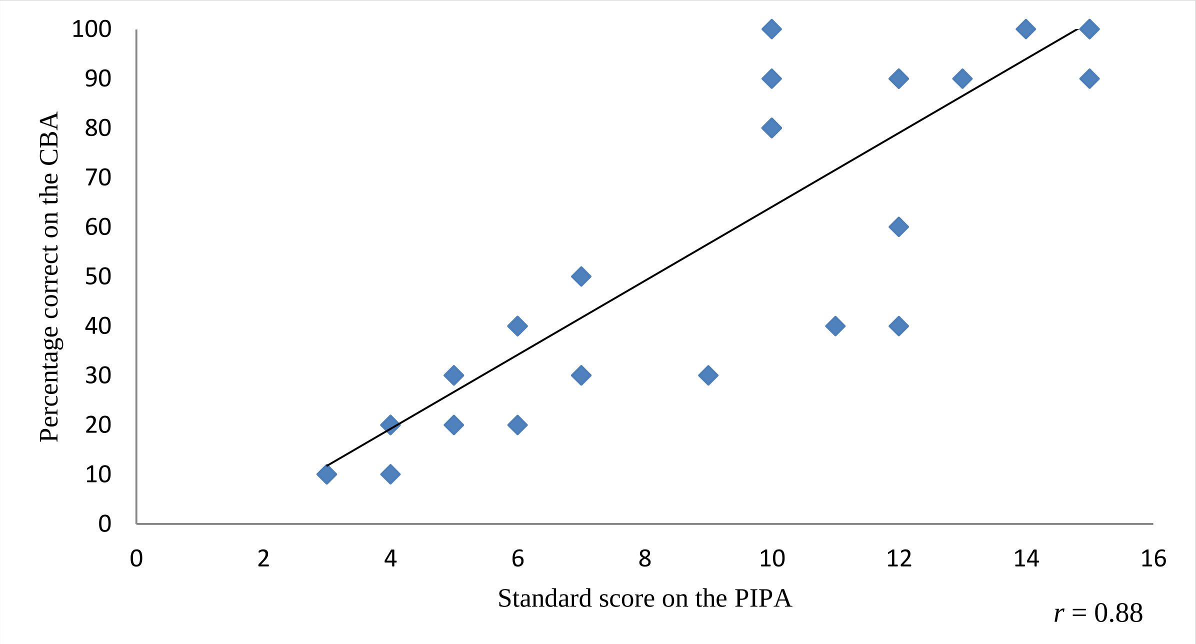


Figure 3.5. Correlation between scores on the PIPA and CBA initial phoneme identity task

Overall, the results of this experiment confirm that CBA can be more time-efficient for the measurement of PA ability and can effectively generate equivalent scores to more conventional paper-based methods of assessment for young children. Figure 3.6 provides a summary of these results.

	Phase 1 (P1)		Phase 2 (P2)
Group A **(*n* = 16)**	**PAPER**  19 minutes; 20 seconds	strong correlations ($r > 0.80$) ←→	**COMPUTER**  13 minutes; 20 seconds 31% faster to administer
	↕ No significant difference between modalities at P1		↕ No significant difference between modalities at P2
Group B **(*n* = 17)**	**COMPUTER**  13 minutes; 20 seconds 31% faster to administer	strong correlations ($r > 0.80$) ←→	**PAPER**  19 minutes; 20 seconds

Figure 3.6. Summary of time-efficiency and equivalency of scores between computer- and paper-based assessment modalities

3.4 Discussion

The experiment reported in this chapter investigated the time-efficiency and equivalency of scores between computer-based and paper-based administration of PA assessment content for children with TD and children with MSD, just prior to school-entry. Thirty-three participants were randomly allocated to one of two experimental assessment groups (Group A and Group B) that, according to baseline assessment, were not significantly different in terms of age, gender and spoken language abilities. Group A received the paper-based version of the assessment first, followed two weeks later by the computer-based version. Group B received the same assessments but in the reverse order of delivery. Assessment content was identical between computer- and paper-based modalities, with the only difference being administration via a computer or an examiner.

3.4.1 Time-efficiency.

The first hypothesis stated that computer-based administration would take significantly less time than would its paper-based counterpart. This hypothesis was supported by data analysis. All four tasks took significantly less time to administer using the computer-based method in comparison to paper-based testing. In total, CBA contributed to a 31 per cent reduction in administration time. These results are consistent with previous research demonstrating that educational assessments can be administered more time-efficiently through the use of computers (Olsen, 1990; Tymms, 2001). Comparison of total administration times for the four tasks from the computer-based PA tool and a selection of PA measures currently available to classroom teachers reveals that CBA can provide a time-efficient alternative for classroom teachers when paucity of time is an issue (see Table 7.1). A key point of difference between the computer-based tool in this experiment and other computer tools currently available to measure cognitive, reading or mathematics skills (e.g., CoPS and PIPS) is the high-priority focus on skills at the phoneme-level in the CBA instrument used here. It is possible that this focus on a small range of skills, as opposed to a broad range of PA tasks, aids in the time-efficiency of the instrument.

Increasing the time-efficiency of PA assessment holds important implications for ensuring teachers nationally and internationally can easily obtain information on children who may or may not be at risk for reading problems. Assessment in the classroom can consume up to one-third of teaching time (McTighe & Ferrara, 1998) and strategies that are time-consuming may be minimised or omitted from the classroom programme (McLeod et al., 2003). In New Zealand, almost 70 per cent of new entrant teachers do not use a nationally available SEA because it is too time consuming (Dewar & Telford, 2003). Internationally, researchers emphasise the need for improved screening instruments within a RTI framework that are efficient to administer and sensitive in their prediction of reading outcomes (Al Otaiba et al., 2012). The results of this experiment contribute to national and international needs by demonstrating that CBA of PA tasks can significantly reduce the quantity of assessment time required, while obtaining the same results as paper-based testing. Moreover, it is important to note that the results of the experiment reported in this chapter represent significant savings in the time that the child is engaged in the assessment. They do not include the time the teacher saves while the child is independently completing the assessment, or the time saved due to not having to organise paper materials prior to the assessment or to score and interpret the results. This provides a strong rationale for further investigation into the use of computer-based methods for the screening and monitoring of PA development in children at the threshold of school-entry.

3.4.2 Equivalency of scores.

The second hypothesis stated that computer- and paper-based assessment modalities would generate comparable scores for children with TD and MSD. Statistical analyses supported this hypothesis in three primary ways. First, children in Groups A and B obtained similar scores during P1 and P2, regardless of assessment modality. Further, results showed that equivalent scores between computer- and paper-based modalities were not only realised at a group level, but were also achieved for children with MSD. That is, children with MSD obtained similar scores despite the modality used to present test stimuli. Ensuring that a computer-based modality generates congruent results across modalities for children who are likely to experience inequalities in PA development is an important step in using this technology to identify risk for reading failure in the classroom. Second, participants' scores from P1 and P2 were highly correlated, despite the change in mode of assessment delivery. Third, performance on the CBA was highly correlated with performance on similar tasks (e.g., rhyme oddity, initial phoneme identity and letter-sound knowledge) from the PIPA. These results provide strong evidence of equivalency of scores between the computer-based PA tool and paper-based testing procedures in the context of this thesis.

Previous research suggests that the modality of assessment delivery generally does not have a significant effect on test scores (Wang et al., 2007; Wang et al., 2008). The current experiment supports these findings on measures that are specific to PA and letter-knowledge for two groups of children among whom inequalities in reading outcomes may exist in the early school years. Establishing equivalency of scores for both children with TD and MSD when exposed to either computer- or paper-based administration means that teachers can more confidently use a particular instrument knowing that the modality they choose in their assessment practice will not advantage or disadvantage children in their classroom.

3.4.3 Limitations and future directions.

It is important to acknowledge the limitations of the experiment reported in this chapter. First, generalisation of results is limited by the use of a small sample size and the fact that this experiment is a pilot investigation. Future research can overcome these limitations through use of larger sample sizes, as well as through replication studies using a similar sample size. Further, investigations that account for time-savings beyond child administration time (i.e., that include the time the teacher saves while the child independently administers the assessment and savings in time spent scoring and interpreting the results) will contribute to the ecological validity of the current experiment and general research area. A number of additional variables affect the time-efficiency of an assessment. These include the educator's own knowledge and confidence in using PA assessment tasks, and how he or she applies his or her knowledge to classroom instruction. Variables such as socio-economic status and gender must also be considered in future research.

This study contributes to educational research in New Zealand and abroad by providing a rationale for further exploration into the use of computers as time-efficient options for teachers to achieve widespread screening and monitoring of PA in the classroom. Helping teachers to collect information on PA development easily will provide valuable information on risk for reading difficulties, and will inform curriculum decision making, while reducing demands on teacher time. It is intended that the computer-based screening and monitoring tool developed in this thesis could serve as a time-efficient adjunct or alternative to the tools and resources already in use in the classroom environment.

Given that the time-efficiency and congruency of scores of this tool have been established in this chapter, the next step was to integrate the use of the computer-based PA tool into the classroom environment to monitor response to reading instruction in the first year of school (see Chapters 4 and 5). This also includes evaluating the validity and reliability of the instrument and identifying areas that require modification before it can be used beyond a research capacity (see Chapter 6).

Chapter 4: Classroom Phonological Awareness Instruction and Literacy Outcomes in the First Year of School

4.1 Introduction

There is strong agreement among reading researchers that instruction in phonological awareness (PA) and letter-sound correspondences and how these skills apply to word recognition are critical components in teaching children to read (Al Otaiba, Puraniki, Zilkowksi, & Curran, 2009; Justice, 2006; Wagner & Torgesen, 1987). However, less is known about the optimal duration and intensity of such instruction when integrated as part of the classroom reading curriculum. The intervention study described in this chapter, therefore, investigates the benefits of time-efficient PA instruction delivered by classroom teachers on raising reading achievement in the first year of school. Investigating the optimal duration and intensity of classroom PA instruction, as the focus of this chapter, will help inform educators' on whether a short period of PA teaching is effective in raising reading outcomes for the majority of children, or if longer-term instruction is necessary.

4.1.1 Benefits of a time-efficient period of classroom phonological awareness instruction.

Results from previous classroom-based studies demonstrate that two years of intensive instruction in phoneme awareness and phoneme–grapheme correspondences can produce significant and long-term benefits for reading outcomes (Shapiro & Solity, 2008). In one study, this translated to a 20 per cent reduction in the prevalence of reading difficulties (Shapiro & Solity, 2008). Studies of shorter duration (e.g., less than one academic year) show immediate benefits for reading post-instruction, but are less advantageous in achieving sustained reading improvements beyond five months of a programme's conclusion (see Table 1.3, in Chapter 1). Examining whether instructional reading programmes of shorter duration can be modified to achieve sustained literacy growth is important from several perspectives. From a classroom management perspective, time-efficient programmes may enable teachers to maximise learning time and increases the ease with which a programme can be integrated into the existing classroom schedule. From a learning perspective, time-efficient programmes are important for ensuring children have the PA and alphabetic knowledge to take advantage of beginning-reading instruction (Kamhi & Catts, 2012). Further, this may help ensure that early identification of children at risk for reading disability takes place, to avoid unnecessary failure in literacy acquisition (Singleton et al., 1996). It is plausible that time-efficient instruction may be supported by focusing PA instruction at the phoneme-level.

4.1.2 Increasing time-efficiency of instruction by focusing on phonological awareness at the phoneme-level.

A commonality among recent classroom investigations is a broad focus on a range of PA skills (e.g., syllable, rhyme and phoneme levels) either as the primary target, or as part of a comprehensive programme involving a number of reading-related skills (see Table 1.3, in Chapter 1). Research demonstrates that a focus on developing skills at the phoneme-level is most successful in generating long-term improvements in reading ability (Ehri et al., 2001; Shapiro & Solity, 2008). Evidence suggests that instruction in smaller phonological units (e.g., phonemes) improves awareness of larger phonological units (e.g., syllables and rhymes) but that the reverse is less likely to occur (Brown, 1998; Yeh, 2003). It is plausible that instruction in phoneme-level skills may be more time-efficient than teaching skills at each level of awareness (e.g., syllable, rhyme and phoneme levels). This may be one way in which short-duration programmes can be modified to achieve sustained literacy improvements. Moreover, pairing PA instruction with phonics (e.g., alphabetic knowledge or phoneme–grapheme correspondences), and ensuring that teaching is direct and explicit, further enhances the quality of PA training (Hiebert & Fisher, 2002; Torgesen, Rashotte, Alexander, Alexander, & MacPhee, 2003). These are important aspects to consider when implementing PA in the classroom.

4.1.3 Adapting the Gillon Phonological Awareness Training Programme (PAT) for the classroom.

A short and highly intensive PA programme implemented in one-to-one or small group sessions under controlled research conditions was adapted for this thesis; namely, the PAT (Gillon, 2000b). In the original PAT, children receive 20 hours of instruction over a 10-week period in one-to-one or small group formats. This programme primarily focuses on developing awareness at the phoneme-level using explicit and systematic teaching strategies. For the current investigation, the PAT was used at the classroom-level with adaptations including 1) the enlargement of materials and 2) the provision of guidelines to teachers on how to adjust activities to match a wide range of ability levels in the classroom. The children in the study described in this chapter received 20 hours of intensive, explicit and systematic instruction focused at the phoneme level over 10 weeks from their classroom teacher in their first year of schooling. This instruction was paired with phonics instruction (i.e., letter-sound knowledge), which was already part of the 'usual' classroom curriculum. The consequent reading outcomes were then compared.

The 'usual' classroom literacy curriculum in this thesis refers to a whole language approach to literacy instruction, in combination with phonics instruction to teach letter-sound knowledge. Whole language instruction focuses on meaning and encourages children to read whole words and sentences in the context of real literacy experiences (Pressley, 2006). Phonics refers to a method of teaching children to read by drawing attention to letters or letter patterns and the sounds they represent (Tunmer, Chapman, & Prochnow, 2006). PA is different from phonics in that it deals specifically with the sound structure of words, however its instruction is best paired with letter knowledge to ensure the link between speech and print is highlighted (Gillon, 2004). The 'usual' literacy curriculum referred to in this thesis—and that which is typically implemented in New Zealand classrooms—does not include a specific focus on teaching PA skills.

In this chapter, the following research questions are addressed:

1) Can teacher-implemented PA instruction in the classroom, focused at the phoneme level for 20 hours over a 10-week period, significantly improve the PA, reading and spelling abilities of five-year-old children, in comparison to the 'usual' literacy curriculum?

   It was hypothesised that children exposed to a short and intensive period of classroom PA instruction focused at the phoneme level would show significantly higher scores on end-of-year phoneme awareness and early literacy measures in comparison to children who received the 'usual' literacy curriculum only.

2) Can a short and intensive period of teacher-implemented PA instruction have a positive effect on reducing the prevalence of reading difficulties after one year of formal schooling?

   It was hypothesised that children exposed to a short and intensive period of classroom PA instruction would demonstrate fewer reading difficulties after one year of schooling in comparison to children who received the 'usual' reading program.

4.2 Method

4.2.1 Participants.

One-hundred and twenty-nine New Zealand children (54 boys and 75 girls) aged between five years and zero months and five years and two months (*M* = 60.41 months, *SD* = 0.59 months) from 12 classrooms and their respective teachers participated in this study. All 12 classrooms were located in the suburbs of the same metropolitan city and were located in government-funded schools. Two classrooms from different schools, but with similar socio-economic, student academic achievement and teacher profiles, were selected to participate as experimental classes. One experimental classroom, consisting of 18 children and their teacher, was randomly assigned to experimental Group A, and the second experimental classroom, consisting of 16 pupils and their teacher, participated as experimental Group B. The remaining 10 teachers and a subset of children from their respective classrooms were automatically assigned to comparison Group C. This group comprised of 95 children. The subset of children in these 10 classrooms ranged from seven to 14 participants. Groups A and B were located in middle socio-economic areas, whereas comparison Group C represented an equal spectrum of socio-economic backgrounds.

4.2.2 School selection and teacher participants.

A stratification process was used to select and invite schools to participate in the study. One-hundred and ten government-funded primary schools in the Christchurch region were stratified into groups of high, middle and low socio-economic groupings based on deciling ranking (see Chapter 3, Section 3.2.1.1). In this thesis, a decile ranking from 1 to 4 was considered low, 5 to 7 was considered middle and 8 to 10 was considered high. Ten schools from each socio-economic grouping were selected at random and comprised the 30 schools invited to participate in the study. From these 30 schools, two of the 12 teachers were asked to participate as part of an experimental group (randomly assigned to either

Group A or B) that would implement classroom PA instruction. These two teachers and their classrooms were selected because the children presented with similar spoken and written language profiles, socio-economic rankings, teacher characteristics and instructional reading programmes. The remaining 10 teachers were asked to continue with their ‘usual’ literacy curriculum for the duration of the study. This primarily involved a whole language approach to teaching reading. All teachers were appropriately qualified and registered to teach in New Zealand Table 4.1 illustrates the characteristics of the teachers in Groups A, B and C.

Table 4.1

Teacher characteristics for Groups A, B and C

Teacher	Age (Years)	Experience (Years)	Decile*	Gender	Class Size
Group A	39	13	7	F	18
Group B	40	14	8	F	16
Group C-1	37	15	10	F	7
Group C-2	38	13	10	F	8
Group C-3	39	16	9	F	8
Group C-4	37	12	6	F	7
Group C-5	37	11	7	F	14
Group C-6	37	14	6	F	13
Group C-7	39	18	6	F	10
Group C-8	38	16	4	F	10
Group C-9	39	13	3	F	8
Group C-10	38	6	2	F	10

Note. Each numeral beside a label ‘Group C’ represents one teacher in the comparison group; * In New Zealand, each school is assigned a decile ranking, which provides an indication of the socio-economic community within which the school is located. This ranking is based on household income, occupation, household crowding, educational qualifications and income support obtained from national census data. The decile ranking is placed on a scale from 1 to 10, with 1 representing a lower socio-economic community and 10 representing an affluent socio-economic community (MOE, 2011).

4.2.3 Child participants.

Teachers distributed consent forms to parents asking for permission for their child to participate in the study. The inclusion criteria were broad to ensure representation of a range of skill levels present in the classroom. Participants were required to: 1) be enrolled to commence their first year of formal education at the start of 2010; 2) have written parental permission to participate in the study; 3) present with sensory, neurological and physical abilities not requiring specialised equipment and/or additional professional support (e.g., the use of sign language or a language interpreter) to achieve accurate testing; and 4) be present at school during prescribed assessment periods.

Parental consent was obtained for all children in the classrooms that were assigned as experimental Groups A (n = 18) and B (n = 16). A subset of seven to 14 children from each of the 10 classrooms that comprised comparison Group C received parental consent to participate. In New Zealand, children typically start Year 1 (i.e., the first year of school) on the day of their fifth birthday or as close to this day as practically possible. All participants spoke standard New Zealand English as their first language. Table 4.2 illustrates the demographic profile of participants in comparison to the Canterbury region from which the study sample was derived. It is important to note that Groups A, B and C were inclusive of a small Māori cohort of children.

Table 4.2

Demographic composition of Groups A, B and C in comparison to primary school-aged students in the Canterbury region

	Gender		Ethnic Grouping				
	Girls	Boys	New Zealand European	Māori	Pacific Islander	Asian	Other
Group A (*n* = 18)	10 (55.55%)	8 (44.44%)	13 (72.22%)	3 (16.67%)	0 (0%)	1 (5.56%)	1 (5.56%)
Group B (*n* = 16)	9 (56.25%)	7 (43.75%)	12 (75.10%)	2 (12.50%)	1 (6.25%)	1 (6.25%)	0 (0%)
Group C (*n* = 95)	56 (58.95%)	39 (41.05%)	64 (67.37%)	16 (16.84%)	8 (8.44%)	4 (4.21%)	2 (2.11%)
*Canterbury Demographic Data	44,798 (49.43%)	45,833 (50.57%)	68,702 (75.80%)	10,909 (12.04%)	3,238 (3.57%)	5,735 (6.33%)	2,044 (2.26%)

Note. Percentages represent a proportion of the group as opposed to the entire sample (e.g., in Group A, 10 girls represent 55.55% of the total number of participants in Group A because 10 divided by 18 equals 55.55%); * Canterbury demographic data taken from Education Counts (2010).

4.3 Procedure

A quasi-experimental design was used to investigate the PA, reading and spelling development of five-year-old children who received teacher-implemented PA instruction or the ‘usual’ literacy curriculum during their first year of school. This design was selected because research was conducted in everyday classroom environments as opposed to highly controlled clinical settings. This quasi-experimental design included a delayed-treatment component. The study took place over a full school year, which in New Zealand runs from February to December and is divided into four terms; each approximately 10 weeks in duration and separated by a two-week holiday break. The study was designed around the four school terms throughout the year 2010 as follows:

Term One:	‘Usual’ literacy curriculum for all groups
Term Two:	Classroom PA instruction for Group A
	The ‘usual’ curriculum for Groups B and C
Term Three:	Classroom PA instruction for Group B
	The ‘usual’ curriculum for Groups A and C
Term Four:	‘Usual’ literacy curriculum for all groups

4.3.1 Professional development for classroom teachers instructing in phonological awareness.

Three levels of professional development were provided to experimental teachers. The first level involved two one-hour meetings with the lead researcher to discuss the programme’s theory and structure. The second level involved providing teachers with an instruction manual outlining the programme’s goals, content, suggested activity dialogue, pre-made resources and activity adaptation charts (see Appendix A). The third level involved the lead researcher co-teaching the first three to four weeks of the programme alongside the classroom teacher to ensure modelling and support was provided. The teachers independently administered the programme from weeks six to 10, with phone and email support being made available if necessary. A total of approximately eight hours of professional development was provided to each of the experimental teachers. The 10 teachers in Group C did not receive any professional development during this period. They were asked to continue with their 'usual' literacy curriculum for the duration of the study.

4.3.2 Assessment phases and measures.

All participants received a comprehensive baseline assessment of their language, PA and early literacy skills at school-entry, in addition to follow-up assessments of PA, reading and spelling development at the middle and end of the school year. As participants in Groups A and B received classroom PA instruction during term two and term three respectively, additional assessment periods were warranted to measure pre- to post-instructional change. This meant that Groups A and B were assessed on a per term basis. In practice, assessments for Groups A and B were conducted at:

- the start of the school year
- the start of term two (i.e., just prior to Group A PA instruction)
- the end of term two (i.e., after Group A PA instruction, just prior to Group B PA instruction and coinciding with the middle-of-year assessment for all groups)
- the end of term three (i.e., after Group B PA instruction)
- the end of term four (i.e., coinciding with the end-of-year assessment for all groups).

Formal assessment measures of language and non-verbal abilities: The following formal measures were administered at school-entry to profile the language, speech, PA and non-verbal intellectual abilities of all participants in experimental and comparison conditions. All administration guidelines were strictly adhered to throughout delivery of the following assessments (see Chapter 3, Section 3.2.11 for specific details of these assessments):

- CELF P-2 (Wiig et al., 2006)
- PIPA (Dodd et al., 2000)
    - The rhyme oddity, initial phoneme identity and letter-sound knowledge subtests were selected from this assessment
- NZAT (MOE, 2004)
    - Analysed using PROPH Software (Long et al., 2002)
- PPVT-4 (Dunn & Dunn, 2007)
- PTONI (Ehrler & McGhee, 2008).

The *Neale Analysis of Reading Ability—Third Edition* (NARA) (Neale, 1999) was administered when participants turned six years of age, coinciding with the end of the school year. ***Content:*** This test provides a measure of reading accuracy (decoding) and reading comprehension of connected text and is standardised on Australian children from six years of age. Children are presented with a series of passages of increasing difficulty, which they are required to read aloud. A reading accuracy score is calculated based on how fluently children are able to read the passages to an examiner (e.g., omitted words, substitutions and phonemic errors are recorded). After each passage, children are required to answer a series of comprehension questions, which provides a reading comprehension score. There are two parallel forms in the NARA: Form 1 and Form 2. Raw scores can be converted into a percentile rank, stanine, performance descriptor, national reading level and a reading age level. ***Reliability and validity:*** Test-retest reliability for this assessment is satisfactory with a coefficient of 0.95 for reading accuracy and 0.93 for reading comprehension. Internal consistency using Kuder-Richardson reliability coefficients is adequate, ranging from 0.71 to 0.95 for Forms 1 and 2 across both reading accuracy and comprehension in the first year of schooling. Evidence of predictive validity is provided, whereby results after one and two years of schooling are highly correlated for reading accuracy (0.83) and reading comprehension (0.78). Concurrent validity shows that the NARA is highly correlated on measures of reading accuracy and comprehension on both Forms 1 and 2 (i.e., above 0.88) with the *Schonell Graded Word Reading Test* (Schonell & Goodacre, 1974). ***Application to the current study:*** Form 1 was used at the end of the first year of school, when children were six years of age, to obtain a standardised measure of reading accuracy and comprehension. Performance was interpreted by comparing raw scores to a 'Reading Age Level'. Children needed to obtain a raw score of 10 on reading accuracy and a raw score of three on reading comprehension to achieve a 'Reading Age Level' of six years and zero months. Children who obtained a reading accuracy raw score below 10 or a reading comprehension score below three were considered as performing at below an age-expected reading level after one year of formal schooling.

Informal assessment measures of PA, reading and spelling development: The following assessments were administered to all participants at the start, middle and end of the school year. They were also used as pre- and post-instruction measures for Groups A and B:

1) **Computer-Based PA Assessment** (Carson, Gillon, & Boustead, 2011): This computer-based assessment (CBA) measures rhyme oddity, initial phoneme identity, final phoneme identity, phoneme blending, phoneme deletion and phoneme segmentation ability. Letter-name and letter-sound knowledge are also assessed. The rhyme oddity, initial phoneme identity and letter-knowledge tasks are modelled on paper-based probes developed and reported by Gillon (2000a, 2002), which in turn were based on earlier work by Bradley and Bryant (1983).

The final phoneme identity, phoneme blending, phoneme deletion and phoneme segmentation tasks are modelled from work developed and reported by Stahl and Murray (1994). The computer administers all test items, records and scores all responses, and allows the child to administer the test independently, using a computer mouse to record his or her response (see Chapter 2 for a description of the CBA). Across the six PA tasks and two letter-knowledge tasks, there are 114 test items. One point is allocated to each test item with a correct response. Points were added within each task to create a total task score. These total scores were then used in data analysis in this chapter. Detailed information regarding the validity and reliability of the computer-based PA tool is provided in Chapter 6.

2) **Real and Non-Word Reading:** Real word reading was measured using the *Burt Word Reading Test—New Zealand Revision* (BURT) (Gilmore, Croft, & Reid, 1981). This test provides a measure of word-recognition ability and requires children to read single words across a test sheet until 10 consecutive errors are made. The words are represented in a graded order of difficulty. Internal reliability is excellent (0.97). Although this test does not provide normative data for children under the age of six years, it was used to provide information on early decoding and sight word abilities. This test was scored by providing one point to each word read correctly. Thus, the total number of words correctly read was the score entered into and used in data analysis. Non-word reading was measured using the 10 non-words (i.e., 10 simple CVC words using short vowels) from the *Non-Word Reading Task in the Reading Freedom Diagnostic Reading Test* (Calder, 1992). This task was scored by providing one point for each correct phoneme–grapheme conversion (as opposed to reading the entire word correctly). Each word contained three phonemes and thus a total of 30 points could be

obtained. The total number of correct phoneme–grapheme conversions was used in the data analysis.

3) **Real and Non-Word Spelling:** Real word spelling was measured using the *Schonell Essential Spelling Test* (Schonell, 1932). This test provides a measure of single word spelling ability and requires children to spell single words spoken by an examiner until 10 consecutive words are spelled incorrectly. The words are graded in order of difficulty. Satisfactory correlations between the *Schonell Essential Spelling Test* and the *Phonic Inventories* (0.60) have been reported (Potter, 2009). Scoring conventions require a whole word to be spelled correctly to obtain a point. Thus, the total number of words spelled correctly was used in the data analysis. Non-word spelling was measured using 10 non-words from the Pseduoword Spelling Subtest of the TOPA-2+ (Torgesen & Bryant, 2004). Internal reliability, inter-rater reliability and test-retest reliability estimates are greater than 0.80 for all age groups. Each correct phoneme–grapheme conversion was awarded a point, with the total number of correct phoneme–grapheme conversions being used to analyse progress.

Tables 4.3 and 4.4 illustrate the language and literacy abilities of participants in Groups A, B and C at the start of the school year.

Table 4.3

School-entry performance on formal measures of verbal and non-verbal skills

	CELF-P2		PTONI	PIPA			PCC
	RLI	ELI		RO	IPI	LS	
Group A (*n* = 18)							
M	102.7	100.3	106.2	5.9	7.9	6.3	94.1
SD	8.3	2.2	6.1	2.2	3.7	2.0	11.7
Range	85–115	83–110	92–115	0–9	0–11	4–9	63–100
Group B (*n* = 16)							
M	102.9	100.6	106.9	6.4	7.9	6.9	94.2
SD	7.0	2.3	5.1	2.1	2.7	1.5	11.1
Range	85–110	80–111	95–113	3–9	0–10	0–8	64–100
Group C (*n* = 95)							
M	98.3	96.8	103.2	6.3	9.1	7.2	93.2
SD	10.8	1.1	7.3	2.5	2.6	2.1	11.0
Range	79–119	80–116	85–116	2–13	3–15	4–11	61–100

Note. CELF-P2 RLI = Clinical Evaluation of Language Fundamentals—Preschool 2, Receptive Language Index (*M* = 100, *SD* +/-15) (Wiig et al., 2006); CELF-P2 ELI = Clinical Evaluation of Language Fundamentals—Preschool 2, Expressive Language Index (*M* = 100, *SD* +/-15) (Wiig et al., 2006); PTONI = Primary Test of Non-Verbal Intelligence standard score (*M* = 100, *SD* +/-15) (Ehrler & McGhee, 2008); PIPA = Preschool and Primary Inventory of PA standard scores (*M* = 10, *SD* +/-3), where RO = rhyme oddity, IPI = initial phoneme identity and LS = letter-sound knowledge (Dodd et al., 2000); PCC = Percentage of Consonants Correct from analysis utilising PROPH software where 93 per cent is considered typical (Long et al., 2002).

Table 4.4

Group performance on informal measures of literacy ability at school-entry

	Phonological Awareness								Reading		Spelling	
	RO*	IPI*	FPI*	PB*	PD*	PS*	LN*	LS*	Real Word	Non-Word	Real Word	Non-Word
Group A (*n* = 18)												
M	5.4	5.4	0.3	1.3	0.3	0.3	13.7	10.9	4.1	2.3	1.1	2.7
SD	2.4	2.2	0.5	1.1	0.5	0.5	4.3	4.7	3.2	3.9	0.9	1.9
Range	0–8	1–8	0–1	0–3	0–1	0–1	4–18	1–16	0–10	0–4	0–3	0–6
Group B (*n* = 16)												
M	5.7	4.8	0.5	1.6	0.5	0.4	13.9	11.5	3.9	2.6	1.1	2.8
SD	2.4	1.8	0.7	1.2	0.5	0.5	3.9	4.7	2.7	3.6	0.7	1.8
Range	1–9	1–6	0–2	0–3	0–1	0–1	3–18	1–16	0–10	0–4	0–2	0–6
Group C (*n* = 95)												
M	5.3	5.3	0.6	1.3	0.5	0.6	13.4	11.1	3.7	2.2	1.2	2.5
SD	2.5	3.0	0.7	1.2	0.8	0.9	4.6	5.4	4.5	3.4	1.7	0.3
Range	0–10	0–10	0–2	0–5	0–4	0–4	2–18	0–18	0–10	0–7	0–9	0–9

Note. *Part of a computer-based PA assessment that is under development (Carson et al., 2011) in which RO = rhyme oddity, IPI = initial phoneme identity, FPI = final phoneme identity, PB = phoneme blending, PD = phoneme deletion, PS = phoneme segmentation, LN = letter-name and LS = letter-sound; Real Word Reading = The Burt Word Reading Test—New Zealand Revision (Gilmore et al., 1981); Non-Word Reading = Calder Non-Word Reading Probes (Calder, 1992); Real Word Spelling = Schonell Spelling Test (Schonell, 1932); Non-Word Spelling = TOPA-2+ (Torgesen & Bryant, 2004).

A one-way analysis of variance did not show a significant difference between the three groups on formal baseline measures of receptive language (F(2, 126) = 2.50, p = .09), expressive language (F(2, 126) = 1.66, p = .19), receptive vocabulary (F(2, 126) = 2.37, p = .10), non-verbal intelligence (F(2, 126) = 3.00, p = .05), speech sound development (PCC F(2, 126) = 0.11, p = .90), PIPA-Rhyme Oddity (F(2, 126) = 0.22, p = .81), PIPA-Initial Phoneme Identity (F(2, 126) = 2.55, p = .08) or PIPA-Letter-Sound Knowledge (F(2, 126) = 1.45, p = .24). Nor were there any significant differences between the three groups on informal measures of rhyme oddity (F(2, 126) = 0.22, p = .80), initial phoneme identity (F(2,126) = 0.25, p = .78), final phoneme identity (F(2, 126) = 0.98, p = .38), phoneme blending (F(2,126) = 0.59, p = .56), phoneme deletion (F(2,126) = 0.65, p = .53) or phoneme segmentation (F(2, 126) = 1.25, p = .29), letter-name knowledge (F(2, 126) = 0.11, p = .90), letter-sound knowledge (F(2, 126) = 0.06, p = .94), real word reading (F(2, 126) = 0.12, p = .89), non-word reading (F(2, 126) = 0.09, p = .91), real word spelling (F(2,126) = 0.08, p = .93) or non-word spelling (F(2,126) = 0.13, p = .88).

4.3.3 Assessment administration and scoring reliability.

Assessments were administered individually to each child by the primary researcher or a qualified speech-language therapist trained in test administration procedures for this study. Children were tested in a quiet area near their classroom across two sessions for initial school-entry testing and then across one session for middle- and end-of-year assessments. Data was scored in real-time with 50 per cent of measures being scored twice, using digital recordings (stored on a DVD). Inter-rater reliability for measures of PA, language and non-verbal intelligence was 100 per cent. Inter-rater reliability for speech sound performance on the NZAT was 98.2 per cent.

4.3.4 Classroom phonological awareness programme.

The PAT was adapted for the classroom environment and used as the instructional programme for this study. The PAT has been used successfully in a number of randomised clinical trials with a variety of children at risk for reading difficulties (e.g., Gillon, 2000a, 2005; Gillon & McNeill, 2009). The content of the PAT covers rhyme knowledge, phoneme analysis, phoneme identity, phoneme segmentation, phoneme blending and linking speech to print. The programme is based on the following principles: 1) instruction should focus on developing skills at the phoneme level, 2) PA activities should be linked with letter-sound knowledge, 3) teaching should be explicit and direct, as opposed to implicit and 4) more intensive individual or small group instruction may be warranted for children with severe PA difficulties. The PAT was adapted for the classroom by enlarging the available resources and expanding the programme adaptations in the original manual, to show teachers how each activity could be modified to match different student abilities. This adapted classroom version of the programme required approximately 20 per cent of the classroom literacy teaching time.

The original PAT requires two one-hour sessions per week until 20-hours of instruction has been completed. Following collaboration with the teachers of Group A and Group B, it was agreed that four 30-minute sessions per week for 10-weeks during the morning literacy block was most suited to the needs of the classroom timetable. PA instruction targeted rhyme oddity for one week, before quickly progressing to explicit teaching of phoneme-level skills for nine weeks. Phoneme-level skills, in particular phoneme segmentation and blending, have been identified as key intervention targets for maximising literacy outcomes (Ehri et al., 2001). Therefore, these skills received the most attention during the 10-week period. Outside the specified instructional periods, Groups A and B continued with the 'usual' literacy curriculum. Table 4.5 illustrates the weekly schedule of PA skills targeted. Specific activity examples can be found in Appendix A.

Table 4.5

Content of the classroom phonological awareness programme

Week	PA skill	Activity description
1	Rhyme	Rhyme bingo and odd-one-out activities were used by the teachers to encourage children to listen for the onset and riming components of spoken words.
2	Initial Phoneme Identity	Initial sound bingo, initial sound matching and odd-one-out activities were used to draw children's attention to the first sound in spoken words. Medial or final sounds were included for children with more advanced skills.
3	Final Phoneme Identity	Final sound bingo, final sound matching and odd-one-out activities were used to draw children's attention to the final sound in spoken words. Medial or final sounds were introduced for children with advanced skills.
4 & 5	Phoneme Blending	Drawing, singing and bingo games were used to teach children to blend words together. Two and three phoneme words were predominantly used. However, words with four phonemes and initial and final blends were used for children with more advanced abilities.
6 & 7	Phoneme Segmentation	Drawing, singing and bingo games were used to teach children how to segment sounds in words. Two and three phoneme words were used. However, four phoneme words and initial and final blends were used to extend students.
8 & 9	Manipulation	Large letter cards or a white board was used to teach children to manipulate letter-sounds in words to create new words.
10	Review	Activities from each of the nine weeks of instruction were reviewed. Focus was directed towards phoneme segmentation and blending activities.

Linking speech to print: All activities required a demonstration of how the PA task related to print. For example, during or after initial sound bingo in week two, children were asked to select three pictures from the bingo board, articulate the first sound they heard, and then write the letter for that sound on a laminated piece of card in front of them.

A 30-minute classroom session involved a 5-minute review of activities from the previous session and discussion about how listening for sounds in words helps with reading and spelling. The next 20 minutes were devoted to two activities of approximately 10 minutes each in duration. Each 10-minute activity targeted the PA skill for that week and ensured that an explicit link to print was demonstrated. For example, when listening for initial sounds in words, children were encouraged to write the letters that represented those sounds on a laminated piece of card in front of them. Each session finished with 5 minutes of shared reading using a book from the classroom. As the teacher read the book with the children, she would emphasise the target PA skill for that week. For example, during the initial phoneme identity week, the teacher would draw children's attention to the first sound in words with prompts such as, 'Who can tell me what the first sound is in cat?' and 'What letters make the /k/ sound?'

Children in Groups A and B were not required to reach a pre-determined performance criterion before moving on to the next PA skill in the classroom programme. Instead, children were exposed to a range of PA activities known to support literacy development and teachers were encouraged to modify and adapt activities to match the different ability levels of children in their classroom. For example, the teacher would modify a phoneme blending task by initially asking children in the class to blend the sounds of a word together and would then adapt the task for a low-ability student by asking that child to identify the first or final sound in that word (see Appendix A for examples of activity adaption charts).

4.3.5 The 'usual' literacy curriculum.

The 'usual' literacy curriculum is underpinned by a whole language approach to the teaching of reading skills. This approach focuses on teaching children to extract meaning from words, sentences and paragraphs as a whole, rather than focusing on letters and sounds. Each classroom in this study was also implementing a phonics programme. The teachers of

Groups A and B and eight teachers from Group C used the Jolly Phonics Program. This programme involves teaching children letter-sound correspondences and includes a section that instructs children on how to blend sounds together to form simple words (e.g., CVC). Teachers of Groups A and B reported using this programme to teach letter-sound skills at the start of their morning literacy block and using the concept of blending sounds together during shared booked reading. Seven of the teachers in Group C used the Jolly Phonics Program to teach letter-sound knowledge, but did not report using the blending section of this programme. One teacher from Group C reported using the blending section of this programme regularly. Two teachers from Group C used school-developed programmes to teach letter-sound knowledge. None of the classrooms included a programme that specifically targeted PA knowledge in an explicit and systematic manner.

The ‘usual’ literacy curriculum across Groups A, B and C involved 15 minutes of guided reading with the teacher in small groups, during which meaning-based strategies such as using knowledge of sight words, looking at the pictures, and attempting to read to the end of the sentence were utilised. Shared book reading as a whole class for approximately 10 to 15 minutes also involved the use of meaning-based strategies. Each day, children were given up to 15 minutes for silent reading for which they selected a book from the class or school library. The teaching of letter-sound knowledge using the Jolly Phonics Program or using a school-developed programme usually involved 20 to 25 minutes of instruction and often occurred as one of the first literacy activities of the day. Additional literacy-based activities included practice writing the ‘letter of the week’ and story writing using a sight-word prompt sheet that included high-frequency words such as ‘a’, ‘the’ and ‘my’.

4.3.6 Independent review of post-treatment data.

An independent examiner conducted all post-instructional testing for both Groups A and B to ensure data was collected by an individual who was ‘blinded’ to the experimental versus comparison conditions (Troia, 1999). Further, 30 per cent of post-instructional assessment measures were randomly selected from the DVD recordings and reviewed by an independent examiner with a qualification in speech and language therapy. A 100 per cent agreement rate was achieved between the real-time examiner results and the independent review of the post-assessment measure DVD recordings.

4.3.7 Treatment fidelity.

Teachers in Group A and B were required to complete a PA teaching log for each week of instruction. In this log, teachers had to name the PA skill that was targeted (e.g., phoneme blending), the activities that were used from the programme to address that target (e.g., phoneme blending bingo) and the duration spent on each activity (e.g., 10 minutes). The teachers were also required to write a short paragraph outlining the response of children to this instruction; in particular, whether any of the children demonstrated difficulty relative to their peers. Comparison teachers were also required to complete a weekly teaching log for the same periods over which Groups A and B were receiving PA instruction. In this log, teachers were asked to document the types of literacy activities that were implemented in the classroom (e.g., guided reading), the types of teaching methods and strategies employed (e.g., context-based cues or letter-sound knowledge) and the duration spent on each activity. In addition, the lead researcher visited each teacher in the study twice during each school term and twice during Group A and Group B's 10-week period of instruction to act as an observer in the classroom and to record details of the classroom literacy programme for treatment validity purposes.

All PA sessions were recorded using a Sony DCR-DVD201 camcorder. Twenty per cent of DVD footage was randomly selected and reviewed by an independent researcher to ensure that each PA skill area was targeted during the programme and that each skill area included teaching that linked speech to print. One hundred per cent of the reviewed data was validated as accurately illustrating the instructional content reported in Table 4.5. Ten teachers in the comparison classrooms participated in four recording sessions to enable data gathering on what constituted activities and strategies within the ‘usual’ literacy curriculum (i.e., approximately once per term). Of this data, 20 per cent was reviewed by an independent examiner who validated that the instructional strategies in comparison classrooms consisted of guided, shared and silent reading with a focus on meaning-based cues. The examiner also validated the use of letter-sound knowledge instruction and the absence of explicit and systematic teaching in phoneme identification, phoneme blending, phoneme segmentation, phoneme deletion and phoneme manipulation skills.

It was possible that exposure and practice at implementing PA activities would affect the reversal back to the ‘usual’ literacy curriculum for experimental teachers. To ensure teachers in Groups A and B returned to the ‘usual’ literacy curriculum following the 10-week PAT programme, classroom reading instruction was recorded twice for Group A (i.e., in terms three and four following instruction) and once for Group B (i.e., in term four). Review of each recording session showed that instructional strategies and resources were predominantly focused on whole language instruction (e.g., there were no time slots allocated specifically to PA), but teachers were more likely to spontaneously draw children’s attention to the initial sounds in words and how to blend and segment sounds in words during classroom reading and spelling activities.

4.4 Results

Repeated measures analysis of variance (ANOVA) were used to examine group differences in PA, reading and spelling performance between children who did (i.e., Groups A and B) and did not receive (i.e., Group C) classroom PA instruction in the first year of school. Independent group t-tests were also conducted to examine differences between experimental Groups A and B. SPSS (Version 19.0) was used to analyse the study data and conventionally reports effect size using partial eta-squared. For the purposes of this thesis, partial eta-squared was converted to eta-squared using the following calculation (Levine & Hullett, 2002):

$$\eta^2 = SS_{between} / SS_{total}$$

Eta-squared is defined as the proportion of total variance attributed to a treatment condition (Field, 2009). To allow for a more direct comparison of Hattie's effect sizes for educational research (see Section 1.4.3.4) and to present effect sizes as standard units, eta-squared for all significant results were converted to Cohen's *d*, using the following calculation (Cohen, 1988):

$$d = \frac{2\sqrt{\eta^2}}{\sqrt{1-\eta^2}}$$

Conventional values for Cohen's *d* are: 0.2 = small effect; 0.5 = medium effect; 0.8 = large effect. These can then be applied to Hattie's effect size ranges for educational research, in which $d \geq 0.4$ is within the zone of desired effect and $d \geq 1.0$ indicates that an instructional programme helped children increase their performance by at least one standard deviation (see Section 1.4.3.4).

4.4.1 Literacy outcomes following classroom phonological awareness instruction.

Group performances on measures of PA, reading and spelling abilities were compared at the start, middle and end of the school year. A multivariate approach to repeated measures, Wilk's Lambda (Assessment T1, T2 and T3 x Group), was used to explore between-group differences on measures of PA, reading and spelling development over time. A significant group by time effect when adjusted for sphericity using the Greenhouse-Geisser Correction method was identified for measures of initial phoneme identity ($F(3.403, .851) = 9.09$, $p <$

.001), final phoneme identity ($F(2.820, .705) = 22.31$, $p < .001$), phoneme blending ($F(3.554, .889) = 9.17$, $p < .001$), phoneme deletion ($F(3.650, .912) = 16.72$, $p < .001$), phoneme segmentation ($F(3.580, .895) = 23.99$, $p < .001$), real word reading ($F(3.078, .769) = 18.54$, $p < .001$), non-word reading ($F(3.091, .773) = 16.82$, $p < .001$), real word spelling ($F(2.961, .745) = 31.45$, $p < .001$) and non-word spelling ($F(3.698, .925) = 13.68$, $p < .001$). No significant group by time effect was identified for rhyme oddity ($F(3.758, .940) = 0.97$, $p = .42$). Linear and quadratic group by time results from repeated measures analyses validated significantly different growth trajectories for phoneme-level skills and literacy measures, but not for rhyme oddity. Figures 4.1 and 4.2 compare the group by time interaction for non-word reading and phoneme segmentation respectively during the first year of school. This pattern of group by time interaction was similar for initial phoneme identity, final phoneme identity, phoneme blending, phoneme deletion, real word reading, real word spelling and non-word spelling. Figure 4.3 illustrates the group by time interaction for rhyme oddity, which did not reveal any significant between-group differences.

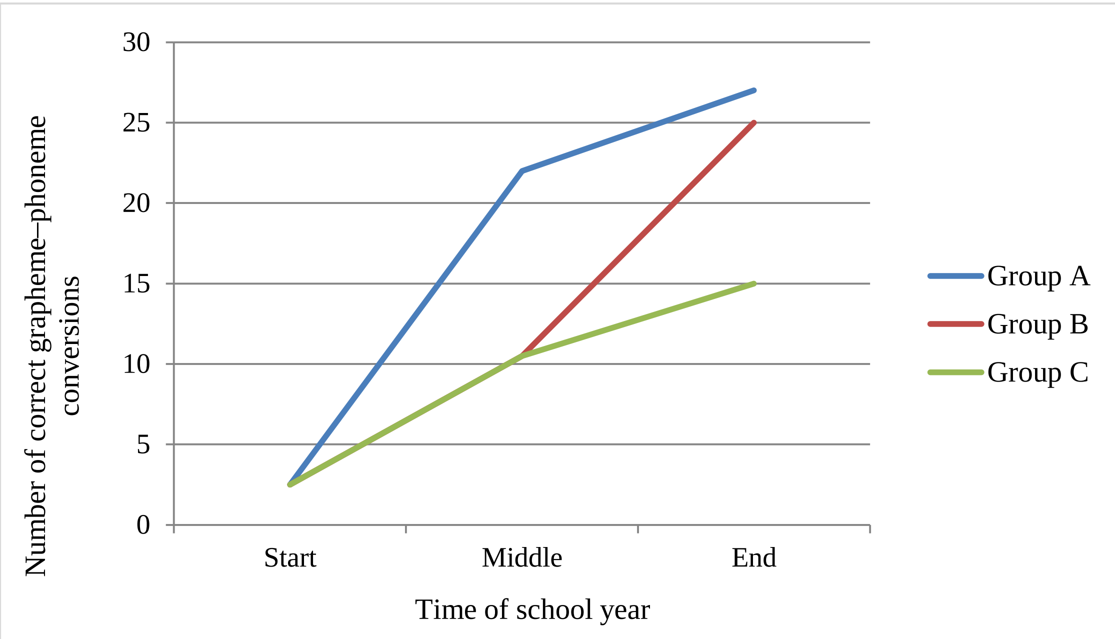


Figure 4.1. Non-word reading performance by groups in the first year of school

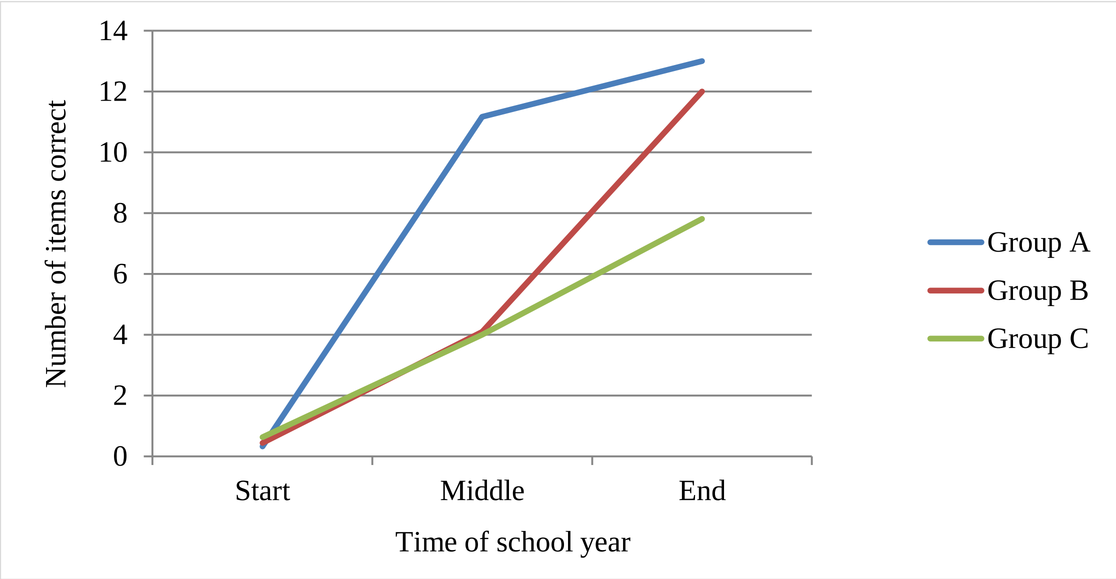


Figure 4.2. Phoneme segmentation performance by groups in the first year of school

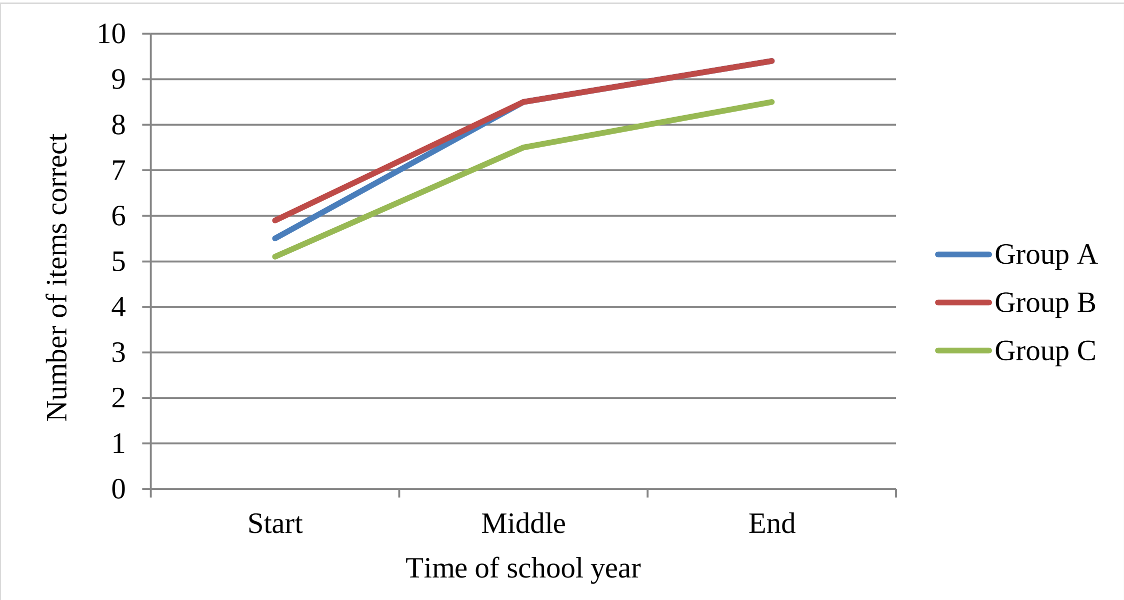


Figure 4.3. Rhyme oddity performance by groups in the first year of school

Tamhane's T2 post-hoc tests revealed that Groups A and B did not perform significantly different to each other on measures of phoneme awareness, early reading and spelling development. However, they did perform significantly different to Group C on all measures, except for rhyme oddity. As mentioned, research suggests that instruction in smaller phonological units (i.e., phonemes) can increase awareness of larger phonological units (i.e., rhyme and syllables), but that the reverse is less likely to occur (Brown, 1998; Yeh, 2003). This may provide one explanation as to why the development of rhyme awareness was consistent between groups irrespective of instructional reading approach. Table 4.6 profiles the between-group significant values obtained for each dependent measure following post-hoc analysis.

Table 4.6

Between-group significant values on assessment measures following Tamhane's T2 post-hoc analysis

	Group Combination		
	A & B#	A & C#	B & C#
Rhyme Oddity	.99	.32	.20
Initial Phoneme Identity	.92	.04	.00
Final Phoneme Identity	.77	.00	.00
Phoneme Blending	.55	.00	.00
Phoneme Deletion	.29	.00	.00
Phoneme Segmentation	.05	.00	.00
Real Word Reading	.06	.00	.00
Non-Word Reading	.16	.00	.04
Real Word Spelling	.23	.00	.00
Non-Word Spelling	.44	.00	.01

Note. # = between group significant values where $p < .05$ is considered significant; A = Group A; B = Group B; C = Group C

4.4.1.1 Gain score analysis.

Gain scores were calculated to measure growth in response to instructional literacy approach. Gain scores were calculated from the start (i.e., T1) to the middle (i.e., T2), and from the middle (i.e., T2) to the end (i.e., T3) of the school year. For example, gain score 1 = T2 score minus T1 score; gain score 2 = T3 score minus T2 score (Portney & Watkins, 2009). A one-way ANOVA from the start to the middle of the school year demonstrated that Group A made significantly greater grains on all study measures except for rhyme oddity when compared to Groups B and C. A one-way ANOVA from the middle to the end of the school year showed that Group B significantly improved on all measures except for rhyme oddity. Effect sizes were calculated for semester one and two gain scores using eta-squared (Levine & Hullett, 2002) and converted to Cohen's *d*. Table 4.7 illustrates the mean gain scores and effect sizes from the start to the middle and the middle to the end of the school year.

Table 4.7

Mean gain scores from the start to the middle and from the middle to end of the school year for Groups A, B and C

	Start to Middle of Year				Middle to End of Year			
	A	B	C	Effect Size (*d*)	A	B	C	Effect Size (*d*)
RO	2.9	2.6	2.2	0.3	1.2	1.2	1.1	0.0
IPI	4.2*	2.1	2.2	0.7	0.3	$2.8^{\times}$	0.9	0.9
FPI	8.3*	4.3	4.2	1.3	0.3	$3.9^{\times}$	1.6	1.5
PB	11.8*	8.1	7.2	0.8	1.4	4.7	3.3	0.6
PD	10.8*	6.7	5.7	1.0	1.9	$5.4^{\times}$	2.6	0.7
PS	11.2*	5.8	5.2	1.3	1.8	$5.9^{\times}$	2.6	1.0
RW Reading	17.5*	9.5	8.6	0.1	10.8	$18.6^{\times}$	8.9	1.2
NW Reading	20.2*	9.4	7.8	1.1	3.8	$13.7^{\times}$	6.4	1.1
RW Spelling	8.3*	4.4	3.1	1.5	5.5	$9.4^{\times}$	4.6	1.2
NW Spelling	14.3*	9.3	6.8	1.0	3.7	$8.8^{\times}$	5.2	0.6

Note. * Represents a skill area in which Group A showed a significantly higher gain score than Groups B and C on Tamhane's T2 Post-Hoc Tests ($p < .05$); $^{\times}$ Represents a skill area in which Group B showed a significantly higher gain score than Groups A and C on Tamhane's T2 Post-Hoc Tests ($p < .05$); RO = rhyme oddity; IPI = initial phoneme identity; FPI = final phoneme identity; PB = phoneme blending; PD = phoneme deletion; PS = phoneme segmentation; RW Reading = real word reading; NW Reading = non-word reading; RW Spelling = real word spelling; NW Spelling = non-word spelling.

During the second half of the school year, detailed analysis revealed that Group B demonstrated significantly more growth in phoneme blending when compared to Group A ($p = .03$) (i.e., Group A approached ceiling scores post-instruction), but not in comparison to Group C ($p = .09$). It is plausible that by the third term of school, interaction with the 'usual' literacy curriculum and increased exposure to written text was sufficient to promote the development of phoneme blending skills. Further, Groups A and C demonstrated gain scores that were not significantly different on measures of phoneme deletion ($p = .11$), phoneme segmentation ($p = .21$), real word reading ($p = .07$) and real ($p = .09$) and non-word spelling ($p = .08$) during this period. This suggests that while Group A continued to acquire phonological knowledge after receiving PA instruction, their development was not at the same accelerated rate when instruction was removed. Group A showed superior growth to Group C in real word reading ($p = .04$). This could suggest that exposure to PA instruction earlier in the year supported enhanced decoding skills of non-words.

The NARA was administered after one year of schooling when participants were six years of age. This was six months post-instruction for Group A and three months post-instruction for Group B. A one-way ANOVA followed by post-hoc tests showed that Groups A and B performed significantly better than participants in Group C in reading fluency ($F(2, 126) = 39.94$, $p < .001$, $\eta^2 = 0.39$, $d = 1.60$) and comprehension tasks ($F(2, 126) = 38.43$, $p < .001$, $\eta^2 = 0.38$, $d = 1.57$). The resulting effect sizes were considered large. Importantly, only 5.88 per cent of children who received PA instruction performed below age-expected levels in reading fluency after one year, compared to 26.32 per cent of children who received the ‘usual’ literacy curriculum. Similarly, 5.88 per cent of children who received PA instruction performed below the age-expected range in reading comprehension at six years of age, compared to 31.58 per cent of children who received the ‘usual’ curriculum. These results demonstrate that sustained benefits for literacy growth were achieved beyond the immediate conclusion of the programme and that this represented a similar decrease in the prevalence of reading difficulties to that reported by Shapiro and Solity (2008). Figure 4.4 illustrates the percentage of children who did not meet an age-expected level in reading accuracy and comprehension after one year of schooling and who were therefore considered at risk for potential reading underachievement.

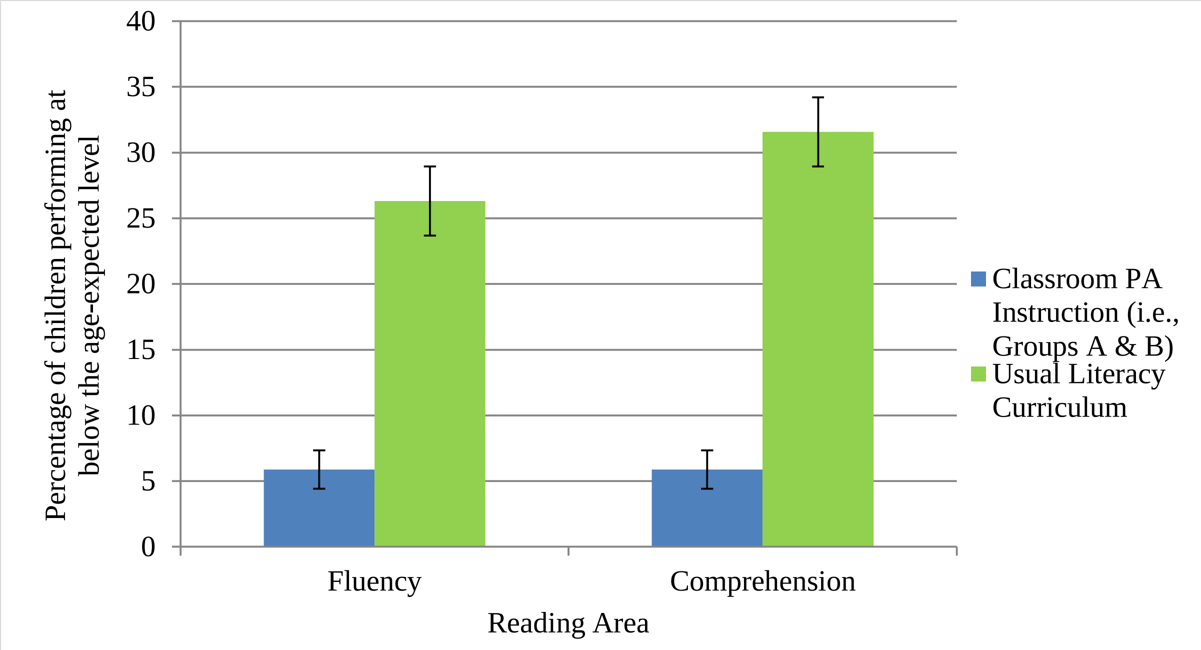


Figure 4.4. Percentage of children performing below the age-expected level after one year of school, according to the Neale Analysis of Reading Ability

4.4.2 Isolating periods of significant improvement in classroom phonological awareness for Groups A and B.

Further analysis was necessary to isolate whether improvements in PA, reading and spelling occurred during the 10-week block of PA instruction, as opposed to during another period in the first or second half of the school year. Pre- and post-instructional data was used to isolate the school terms within which significant improvements in PA, reading and spelling occurred for Groups A and B. At the start of term two, just prior to Group A receiving PA instruction, independent sample t-tests revealed no significant differences between Groups A and B on measures of rhyme oddity ($t(32) = -1.93$, $p = .06$), initial phoneme identity ($t(32) = -0.24$, $p = .81$), final phoneme identity ($t(32) = -1.28$, $p = .21$), phoneme blending ($t(32) = -1.06$, $p = .30$), phoneme deletion ($t(32) = -1.86$, $p = .07$), phoneme segmentation ($t(32) = -1.63$, $p = .11$), real word reading ($t(32) = -1.88$, $p = .07$), non-word reading ($t(32) = -1.51$, $p = .14$), real word spelling ($t(32) = -1.94$, $p = .06$) or non-word spelling ($t(32) = -0.90$, $p = .37$).

Group A then received 10-weeks of PA instruction in term two. At the end of term two, Group A performed significantly higher than Group B on measures of initial phoneme identity ($t(32) = 4.85$, $p < .001$, $d = 1.7$), final phoneme identity ($t(32) = 5.93$, $p < .001$, $d = 2.1$), phoneme blending ($t(32) = 4.11$, $p < .001$, $d = 1.45$), phoneme deletion ($t(32) = 4.18$, $p < .001$, $d = 1.48$), phoneme segmentation ($t(32) = 6.76$, $p < .001$, $d = 2.39$), real word reading ($t(32) = 7.93$, $p < .001$, $d = 2.80$), non-word reading ($t(32) = 4.13$, $p < .001$, $d = 1.46$), real word spelling ($t(32) = 4.81$, $p < .001$, $d = 1.70$) and non-word spelling ($t(32) = 3.24$, $p < .001$, $d = 1.15$). Rhyme oddity did not reveal any significant between-group differences ($t(32) = -1.28$, $p = .21$, $d = 0.45$). Figure 4.5 illustrates the significant difference between Group A and Group B's performances on the non-word reading task following Group A's PA instructional period.

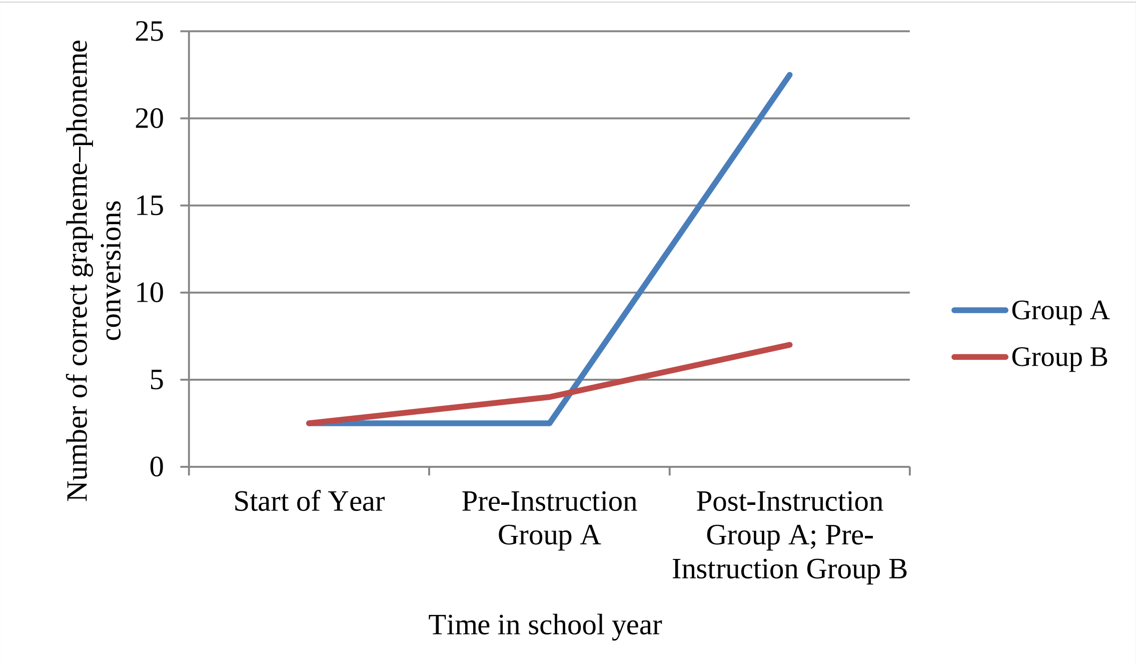


Figure 4.5. Non-word reading performance during the first year of school for experimental classrooms

In term three, Group B received 10 weeks of PA instruction and Group A resumed their ‘usual’ literacy curriculum. At the end of term three, Group B no longer performed significantly lower than Group A on measures of initial phoneme identity ($t(32) = 0.50$, $p = .62$, $d = 0.18$), final phoneme identity ($t(32) = 1.93$, $p = .06$, $d = 0.68$), phoneme blending ($t(32) = 0.56$, $p = .58$, $d = 0.20$), non-word reading ($t(32) = 1.99$, $p = .06$, $d = 0.70$), real word spelling ($t(32) = 1.04$, $p = .31$, $d = 0.37$) and non-word spelling ($t(32) = 0.52$, $p = .61$, $d = 0.18$). Rhyme oddity also continued to demonstrate no significant differences between groups ($t(32) = -0.52$, $p = .61$, $d = 0.18$). Significant between-group differences were identified for measures of phoneme deletion ($t(32) = 2.34$, $p = .03$, $d = 0.83$), phoneme segmentation ($t(32) = 2.29$, $p = .01$, $d = 0.81$) and real word reading ($t(32) = 2.26$, $p = .03$, $d = 0.80$). However, by the end of the school year, these significant differences were no longer present: phoneme deletion ($t(32) = 1.15$, $p = .26$, $d = 0.40$), phoneme segmentation ($t(32) = 2.03$, $p = .06$, $d = 0.70$) and real word reading ($t(32) = 0.16$, $p = .88$, $d = 0.05$). It is plausible that these immediate differences were related to Group A receiving instruction one term earlier and their having had extra time to consolidate and practice skills. Figure 4.6 illustrates the changes in non-word reading ability over the first year of schooling as a function of timing of classroom PA instruction.

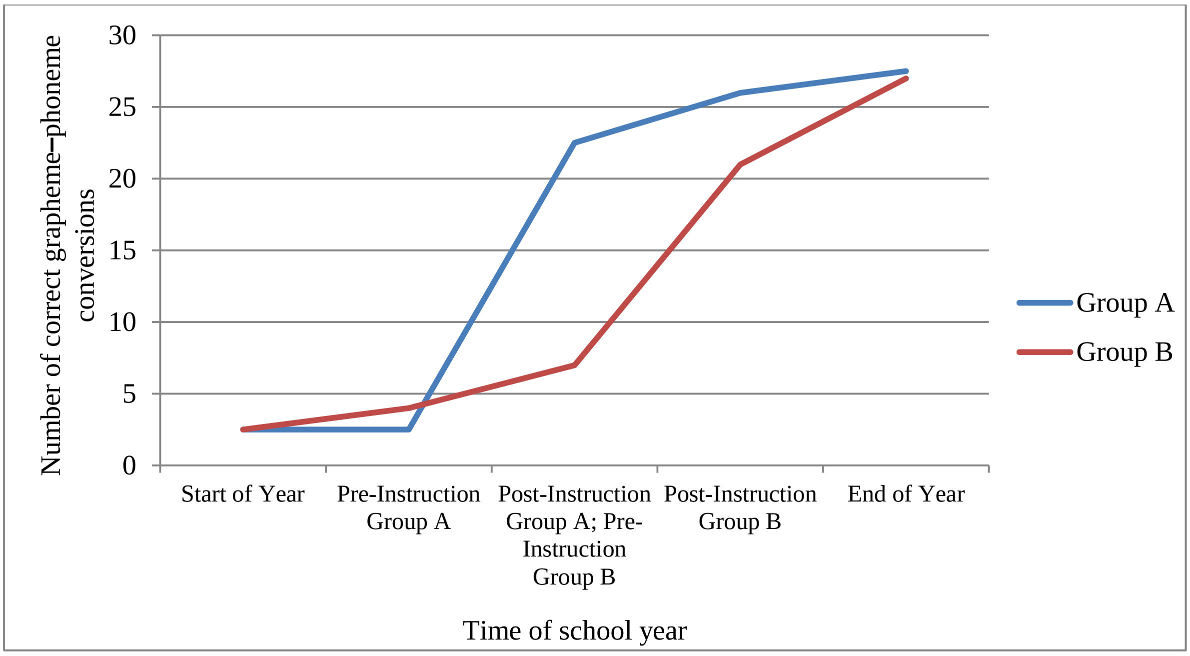


Figure 4.6. Non-word reading performance during the first year of school for experimental classrooms

Collectively, these results contribute to the existing literature by demonstrating that a short and intensive period of teacher-directed PA instruction focused at the phoneme-level can have a significant effect on raising reading achievement immediately after and up to five months post-instruction for children in the classroom environment.

4.5 Discussion

This study contributes to existing literature by demonstrating that a short and intensive period of PA instruction focused at the phoneme-level and delivered by classroom teachers can have a significant effect on accelerating reading outcomes. One-hundred and twenty-nine five-year-old children participated in the experiment reported in this chapter. These children were divided into three groups: experimental Groups A and B and comparison Group C. Children in the experimental groups received 10 weeks of teacher-led classroom-wide PA instruction supplementary to the 'usual' classroom reading programme. Children in the comparison group received the 'usual' curriculum only, which did not include a focus on PA instruction. Data was analysed to determine the effect of PA instruction on end-of-year reading outcomes, in addition to the effect on reducing the prevalence of reading difficulties before the second year of schooling.

4.5.1 Raising reading achievement through classroom phonological awareness instruction.

The first hypothesis stated that children exposed to a short and intensive period of classroom PA instruction focused at the phoneme-level would show significantly higher scores on end-of-year phoneme awareness and early literacy measures in comparison to children who received the 'usual' literacy curriculum. This hypothesis was supported by statistical analyses. Children in Groups A and B performed significantly higher on phoneme awareness, reading and spelling tasks after one year of school in comparison to children in Group C. This indicates that instructional effects were maintained up to six months post-instruction for Group A and three months post-instruction for Group B. A specific focus on phoneme-level skills allowed for maximal teaching time to be spent on critical precursors to literacy success. All three groups performed similarly on rhyme oddity throughout the school year, regardless of the classroom reading programme they received.

Previous classroom investigations of short duration (see Table 1.3, in Chapter 1) that included a focus on PA instruction have generally struggled to demonstrate sustained improvement for reading outcomes beyond five months of the programme's conclusion (Fuchs et al., 2001; Justice et al., 2010; McIntosh et al., 2007). In the current study, a high-priority focus on developing skills at the phoneme-level was employed, whereby nine weeks of the 10 week programme directly targeted initial phoneme identity, final phoneme identity, phoneme blending, phoneme segmentation and phoneme manipulation. Research demonstrates that phoneme-level skills are critical to early literacy success, are often deficient in children at risk for reading disorder and can be stimulated in children as young as four years of age with spoken language difficulties (Bradley & Bryant, 1983; Catts et al., 1999; Gillon, 2005). A specific focus on these skills may have contributed to sustained improvements in phoneme awareness, reading and spelling for children in the current study.

Further, research posits that increased awareness of smaller sound units, such as phonemes, is likely to support increased awareness of larger phonological structures such as onset-rime and syllable units (Brown, 1998; Yeh, 2003). Given that children in Groups A and B received instruction in rhyme and phoneme awareness, it could be anticipated that these children would show superior performance on rhyme awareness measures. However, comparison of gain scores demonstrated that all groups performed equally well on rhyme awareness. This suggests that 'usual' classroom instruction may be sufficient in raising awareness of larger sound units (e.g., rhymes, syllables), but insufficient to raise awareness at the critical phoneme-level. This highlights the importance of including PA instruction focused at the phoneme-level as part of classroom reading programmes.

4.5.2 Reducing the prevalence of reading difficulties through phonological awareness instruction.

The second hypothesis stated that children exposed to a short and intensive period of PA instruction as a group would demonstrate fewer reading difficulties after one year of schooling in comparison to children who received the 'usual' reading programme. Data analyses supported this hypothesis. By six years of age, only 5.88 per cent of the children who received PA instruction were performing below the age-expected level in word-decoding skills, compared to 26.32 per cent of children who did not receive PA instruction. Similarly, only 5.88 per cent of children who had received PA instruction, as compared to 31.58 per cent of children who had continued with the 'usual' curriculum, performed below the age-expected level in reading comprehension after one year of school. Therefore, teacher-delivered PA instruction in this study coincided with a more than 20 per cent reduction in the prevalence of children considered at risk for reading difficulties by six years of age. This provides initial evidence that classroom PA instruction can have a positive effect on reducing the number of school-aged children struggling with reading development.

Previous investigations in the classroom have varied widely in duration, intensity and content of PA instruction. Shapiro and Solity (2008) demonstrated one of the most salient and sustained benefits for reading outcomes following teacher implementation of classroom phoneme awareness instruction. This instruction was delivered three times a day for two years and resulted in a 20 per cent reduction in the prevalence of reading difficulties. In comparison, the current investigation involved 10 weeks (totalling 20 hours) of direct and intensive phoneme-focused instruction led by classroom teachers. Significant improvements in phoneme awareness, reading and spelling were observed up to six months post-instruction and resulted in a similar reduction in the prevalence of reading difficulties to that reported by Shapiro and Solity's (2008) longer-term study. Moreover, the PA programme employed in this study only required 20 per cent of classroom literacy teaching time, whereas Shapiro and Solity's (2008) investigation required adaptations to the entire school day (i.e., to accommodate three sessions a day).

This result demonstrates that short-duration, highly intensive, phoneme-focused and teacher-led instruction can be highly effective in accelerating literacy outcomes and reducing the prevalence of reading difficulties in the classroom environment. Shorter programmes may be easier and more manageable for teachers to integrate into existing reading curriculums, and may help ensure children possess the necessary skills to take advantage of beginning-reading instruction. Moreover, time-efficient programmes allow educators to allocate valuable teaching time to other critical areas of literacy development (i.e., comprehension) in the first year of schooling.

4.5.3 National and international classroom implications.

In New Zealand, a whole language approach to reading instruction continues to be the method of choice for many educators (Tunmer et al., 2006). New Zealand presents with one of the largest spread of scores between high- and low-ability readers among OECD countries (Martin et al., 2007). In this thesis, the 'usual' literacy curriculum refers to instruction that is predominantly whole language in focus, with the teaching of phonics as an adjunct. The 'usual' literacy curriculum does not include PA instruction. A whole language approach assumes that children will learn how speech relates to print and the connections between phonemes and graphemes naturally, with the teaching of phonological and word-level skills being implicit (Pressley, 2006). In contrast, the classroom PA programme employed in this study used explicit and systematic instructional strategies. Children who were exposed to explicit and systematic PA instruction outperformed those children who received implicit and less systematic whole language instruction. Explicit and systematic instructional strategies may have enhanced children's awareness of the crucial links between the spoken and written language paradigms that support proficient reading development. This in turn may have increased the rate of successful decoding experiences, providing richer access to positive literacy experiences.

PA instruction exerts a greater influence on reading and spelling development when paired with the teaching of letter-sound knowledge (Ehri et al., 2001). Consistent with this notion, children exposed to the teacher-implemented PA programme, in addition to an established classroom phonics programme, performed significantly higher on end-of-year reading and spelling measures, compared to children who received phonics only instruction in the context of a whole language curriculum. Moreover, these gains were achieved before the stage at which children are typically recommended for Reading Recovery. This represents the potential for a shift from the concept of recovery to that of prevention. Further, this can be done in a time-efficient way, as a short period of highly intensive phoneme-level instruction can operate in unison with the predominant approach to reading instruction being implemented in New Zealand schools. These findings provide support for the New Zealand Ministry of Education strategy to raise Māori achievement - *Ka Hikitia - Managing for Success: The Māori Education Strategy 2008 – 2012* (MOE, 2012a). This is because a small cohort of Māori children were included in Groups A and B who showed significant improvements in reading outcomes following classroom PA instruction. Overall, these results suggest that nationally the integration of whole language, phonics and PA instruction may support the elevation of reading achievement and the reduction of the reading inequalities prevalent in the New Zealand education system.

Internationally, researchers posit a need to investigate the optimum duration of classroom PA instruction (Al Otaiba et al., 2012). Such information would provide direction for when to engage further support for struggling readers. The results reported in this chapter suggest that children who show little improvement after 10 weeks of explicit and highly intensive PA instruction may require more in-depth small group or one-to-one support inside or outside the classroom. For classrooms globally, a significant advantage of employing a short period of instruction is that educators can quickly identify children who are not responding to instruction, in turn allowing for the early identification and remediation of reading difficulties before children enter the second year of formal schooling. This may help other OECD countries presenting with large discrepancies between good and poor readers to achieve greater equality in reading outcomes for school-aged children.

4.5.4 Limitations and future directions.

The positive outcomes from this study must be interpreted in the context of study limitations. First, the use of a quasi-experimental design in which participants were not randomly assigned at an individual level to each instructional condition may limit the causal relationships postulated in this study. A quasi-experimental design was employed because participants were already found as part of intact groups (i.e., classrooms). Attempts to counteract the lack of random assignment included the use of comparison Group C and ensuring that Groups A, B and C were not significantly different at the start of the study. It is plausible that the generalisability of the findings may be confounded by variability in teacher, child and location factors that exist between educational settings. However, all participants resided in the same metropolitan city, and those participants who received intervention came from average socio-economic backgrounds.

It must also be acknowledged that gains in reading and spelling may in part be related to the quantity of professional development provided to teachers in the experimental condition. Teachers of Groups A and B received eight hours of professional development and in-class support, whereas teachers of children in Group C did not receive any formal professional development. Further, it cannot be firmly ruled-out that Group A and Group B teachers did not inadvertently continue to implement the newly learned PA strategies when they reverted back to their 'usual' reading programme. These limitations warrant further investigation through replication studies involving a range of education contexts and multiple instructional sites. The limited number of children with SLI also necessitates future investigation through the use of larger sample sizes. Further investigation to help children with SLI transfer their phonological knowledge to written language tasks and benefit equally from instruction is also necessary.

In summary, teaching children to become efficient readers in their own classrooms is paramount to future academic learning and life-long success. The findings reported in this chapter contribute to existing literature by demonstrating that a short and intensive period of teacher-led instruction in PA focused at the phoneme-level has the potential to exert a significant influence on elevating reading achievement and reducing the prevalence of reading problems. Evaluating the effect of such a programme on the reduction of inequalities in reading outcomes between children with and without heightened risk for reading disorder would provide further support for utilising the modified PAT as a time-efficient instructional tool in the classroom. To aid in this endeavour, Chapter 5 examines the response of children with SLI to classroom PA instruction and compares this to the performance of children with typical language skills. The effect on inequalities in reading outcomes after one year of formal education is then discussed.

Chapter 5: Classroom Phonological Awareness Instruction and Literacy Outcomes in the First Year of School for Children with Spoken Language Impairment

5.1 Introduction

Children with spoken language impairment (SLI) are often slower to develop phonological awareness (PA) knowledge, particularly at the phoneme level, in comparison to their peers with typical development (TD) in the classroom (Catts et al., 1999; Gillon, 2005; Raitano et al., 2004; Rvachew, Ohberg, Grawberg, & Heyding, 2003). Difficulties developing PA increase the risk of experiencing early word-recognition difficulties, which in turn has a negative effect on the development of skilled reading comprehension (Rvachew et al., 2003). It is not surprising, therefore, that children with SLI are four to five times more likely to experience inequalities in reading outcomes (Catts et al., 2001). In a longitudinal study, Catts and colleagues (2002) reported that up to 50 per cent of children with SLI in kindergarten demonstrate poor reading performance in the early school years. From this group, nearly 60 per cent will present with deficits in critical PA skills (Catts et al., 1999). Moreover, it is estimated that 60 to 70 per cent of children with some form of language impairment perform below the 25th percentile in reading comprehension by seven to eight years of age (Catts et al., 2002). In this thesis, children with SLI were selected as an at-risk comparison group. This chapter investigates the response of children with SLI to classroom PA instruction, and examines the effect this has on equalising reading equality in comparison to children with TD during the first year of school.

5.1.1 Learning to read and spoken language impairment.

It is often reported that children with SLI rely more heavily on context, as opposed to phonemic decoding skills, to recognise words when learning to read (e.g., Al Otaiba et al., 2012). Relying on this strategy may prove successful when beginning-reading books use high-frequency words, are predictive in structure and contain picture cues, but is a less reliable strategy as text difficulty increases and word choice is no longer constrained to the same degree (Al Otaiba et al., 2012). Instructional methods that encourage the use of context, such as the whole language approach that is typically used in New Zealand classrooms, may inadvertently reinforce this compensatory strategy rather than strengthening critical phoneme awareness skills. Research suggests that without specific instruction in PA, children with SLI are more likely to experience persistent difficulties in acquiring PA knowledge (Bird et al., 1995; Snowling, Bishop, & Stothard, 2000; Webster, Plante, & Couvillian, 1997). Teaching children to become aware of the individual sounds in words and how to apply this awareness to learning to read may help vulnerable readers to begin self-teaching in word recognition (Share, 1995), avoiding the negative 'Matthew Effect' (Stanovich, 1986) and thereby developing a positive self-perception as a reader (Chapman & Tunmer, 1997). If this is the case, including a focus on PA instruction at the phoneme-level as part of classroom reading instruction is an important consideration when attempting to reduce inequalities in reading outcomes.

5.1.2 Recent reports of inequalities in response to beginning classroom reading instruction.

Explicit instruction in PA and the application of this knowledge to the decipherment of the meaning of an alphabetic orthography has proven successful for children with SLI in one-to-one or small group settings under controlled research conditions (Bradley & Bryant, 1983; Ehri et al., 2001; Gillon, 2004). More recently, researchers have evaluated the benefits of PA instruction as part of a comprehensive literacy curriculum in the wider classroom environment. As discussed in Chapter 4, Justice and colleagues (2010) examined the benefits of *Read It Again* (RIA), a language and literacy supplement, for preschool children considered to be at risk for literacy difficulties. When analysing outcome measures based on child language abilities, results showed that children with lower language skills benefitted less from RIA instruction in phoneme awareness, letter-knowledge and print awareness compared to children with more typical language skills. A number of important language and pre-reading skills were targeted including PA, vocabulary and narrative ability. PA sessions focused on a broad range of skills (e.g., syllable, rhymes and phonemes) and thus differed in content to the classroom programme reported in Chapter 4, which primarily focused on teaching skills at the phoneme-level.

Similarly, Fuchs, Fuchs, Thompson, Al Otaiba, Yen, Yang, Braun, & O'Connor (2002) reported that the number of school-aged children with speech and/or language impairment who demonstrated improved literacy outcomes following classroom PA and decoding instruction was equal to the number of children with speech and/or language impairment who demonstrated no improvement. Again, a broad range of PA skills was targeted in their study (e.g., syllables, rhyme and phonemes). In line with previous research, the work of Justice et al. (2010) and Fuchs et al. (2002) shows that children with poor spoken language skills demonstrate a slowed and more variable response to language and literacy instruction, in comparison to children with average to above-average language abilities (Boudreau & Hedberg, 1999; O'Connor, Notare-Syverson, & Vadasy, 1996; Penno, Wilkinson, & Moore, 2002).

Research also shows that there is usually a small group of children who show very little improvement in response to instruction. These children are often referred to as 'non-responders' or 'treatment resisters' (Al Otaiba & Fuchs, 2002; Al Otaiba & Torgesen, 2007). Children who respond poorly to instruction often present with poor phoneme awareness, deficits in phonological memory, difficulties in processing orthographic information, paucity of attention, behavioural problems or lower intellectual abilities (Al Otaiba & Fuchs, 2002; Nelson, Benner, & Gonzalez, 2003). It is estimated that approximately three to five per cent of children will respond poorly to supplementary instruction (Wanzek & Vaughn, 2009). In some cases, it may be difficult for educators and researchers to identify why a child does not respond favourably to instruction. Attempting to uncover factors contributing to reading failure and a poor response to instruction is therefore necessary to avoid further underachievement and reading inequality.

5.1.3 Attempting to reduce inequality by focusing on phonological awareness at the phoneme level.

Given that children with poor spoken language skills appear to benefit less (e.g., Justice et al., 2010) and show a more variable response (e.g., Fuchs et al., 2002) to classroom instruction focused on a broad range of PA skills, it appears worthwhile to investigate whether inequalities in instructional outcomes can be reduced when a specific focus on phoneme-level skills is employed. Research under controlled clinical settings demonstrates that a specific focus on PA at the phoneme-level can have a significant effect on advancing the reading skills of children with expressive phonological impairment, speech and language impairment, developmental verbal dyspraxia and Down Syndrome (Gillon, 2000, 2002, 2005; McNeill, Gillon, & Dodd, 2009; van Bysterveld, Gillon, & Foster-Cohen, 2010). Exporting these benefits to the heterogeneous classroom environment to accelerate the reading developing of children at-risk for literacy difficulties may provide one useful pathway towards ensuring every child in the classroom reaches their potential in reading development.

A large number of programmes have been developed to support instruction in phoneme awareness and early word-recognition ability (Fuchs et al., 2001; Gillon, 2000a; Justice et al., 2010; McIntosh et al., 2007; Solity & Shapiro, 2008). As stated in Section 1.4.3, these programmes vary in intensity and duration of instruction. Evaluating whether the phoneme awareness and early reading skills of children with SLI can be successfully raised in a time-efficient manner (i.e., 20 hours over 10 weeks) as part of the classroom reading programme will contribute towards initiatives aimed at raising literacy achievement and thereby reducing inequalities in reading outcomes for this group of children. To address this area of investigation, children presenting with SLI in the study reported in Chapter 4 were identified, and their results were extracted for detailed analysis. As mentioned in Section 1.4.4.1, SLI in this thesis refers to children with 'specific language impairment' and children with 'phonologically based' speech impairment. The following research questions are addressed:

1) Can teacher-implemented PA instruction focused at the phoneme-level for 20 hours over a 10-week period in the classroom environment significantly improve the reading and spelling abilities of five-year-old children with SLI?

   It was hypothesised that children with SLI would show significantly higher scores on phoneme awareness and early literacy measures following teacher-directed classroom PA instruction.

2) Do children with SLI equally benefit from classroom-wide teacher-implemented PA instruction at the phoneme-level in comparison to children with TD during the first year of schooling?

   It was hypothesised that although children with SLI would show significant improvement in phoneme awareness and early literacy following teacher PA *instruction, they would also demonstrate variability in the extent to which they benefitted from instruction when compared to peers with TD.*

3) What is the effect on reducing reading inequalities between children with SLI and TD when exposed to 10 weeks of teacher-directed classroom PA instruction in comparison to the 'usual' classroom literacy curriculum?

It was hypothesised that the majority of children with SLI who received classroom PA instruction would approach an age-appropriate level in reading accuracy after one year of schooling, making their performance comparable to that of the children with TD who received the 'usual' classroom curriculum. However, it was anticipated that they would remain deficient when compared to children with TD who also received class PA instruction.

5.2 Method

5.2.1 Participants.

In the study reported in Chapter 4, 34 children (i.e., experimental Groups A & B) received classroom PA instruction from their teacher, while 95 children (i.e., comparison Group C) continued with the 'usual' literacy curriculum. In each study condition, a subset of children presented with SLI. The data of this subset of children was extracted to evaluate 1) how children with SLI respond to classroom PA instruction focused on phoneme-level skills, 2) whether instruction coincides with improved reading outcomes after one year of schooling and 3) the effect on reducing inequalities in reading achievement. Figure 5.1 profiles the number of participants who presented with SLI and TD in the experimental and comparison conditions in the study reported in Chapter 4.

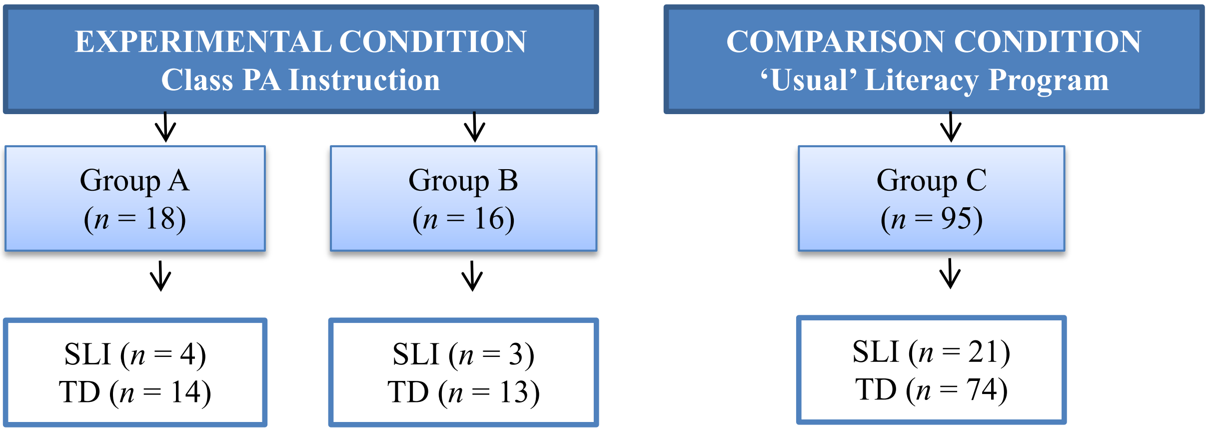


Figure 5.1. Participants with SLI and TD in the experimental and comparison conditions from the main study reported in Chapter 4

For the purposes of this chapter, the children from Groups A and B were aggregated to form one experimental group. This was to achieve a larger sample size of children with SLI who had received classroom PA instruction. It is important to acknowledge that aggregation of children in Groups A and B may introduce an error margin into study results because Group A received instruction 12 weeks earlier than Group B. Nonetheless, the goal of this Chapter is to evaluate reading outcomes after one year of school, as opposed to terms within the school year. In total, the experimental condition consisted of seven children with SLI and 27 children with TD. In the comparison condition, 21 children presented with SLI and 74 children presented with TD. All criteria and procedures reported in Chapter 4 (e.g., inclusion criteria, formal and informal assessments, professional development for classroom PA teachers, description of the classroom PA program, the 'usual' literacy curriculum and treatment fidelity) were identical for the information extracted and reported in this chapter.

Classifying spoken language impairment and typical spoken language development: Baseline data collected from the study in Chapter 4 was used to classify children as presenting with either TD or SLI. To be classified as having TD children needed to score within or above the average range on the following assessments:

- CELF-P2 (Wiig et al., 2006): Receptive language index and expressive language index; a composite score of 85–115 is considered the average range
- PIPA (Dodd et al., 2000): Rhyme oddity, initial phoneme identity and letter-sound knowledge subtests; a standard score of 7–13 is considered the average range
- PTONI (Ehrler & McGhee, 2008): A score of 85–115 is considered the average range
- NZAT (MOE, 2004): Single consonants and consonant blends subtests; a PCC above 93 per cent for five-year-old children is considered satisfactory (Shriberg et al., 1997).

To be classified as having SLI, children needed to perform at least one standard deviation below the mean on one of the baseline language measures (e.g., CELF-P2 or PIPA) or present with phonologically based speech errors with a PCC below 93 per cent. In practice, children who perform at least one to two standard deviations below the mean are considered to have impaired language skills, and thus a cut-off of at least one standard deviation below the mean was used in this study (Tomblin, Freese, & Records, 1992). A PCC of 93.4 per cent to 100 per cent is considered typical for children aged five years and zero months to five years and 11 months (Shriberg et al., 1997), and the following descriptors are often applied to describe PCC scores: mild = 85–100 per cent; moderate = 65–85 per cent; moderate-severe = 50–65 per cent; severe = < 50 per cent (Shriberg & Kwiatkowski, 1982). Children in this study who had a PCC below 93 per cent were considered to present with phonologically based speech impairment, ranging in severity from mild to severe.

All children with SLI in the experimental condition (100 per cent) and the majority of children with SLI in the comparison condition (80.95 per cent) performed at least one standard deviation below the mean on a language measure (e.g., receptive language, expressive language, PA) and presented with a PCC below 93 per cent in speech sound development. That is, the majority of children presented with deficits in both language and speech (i.e., phonological) development. All children with SLI in this study performed within the average range on a measure of non-verbal intelligence. Table 5.1 profiles the language, PA, speech and non-verbal intellectual abilities of children with SLI and TD in the experimental and comparison conditions at the start of the study.

Table 5.1

Spoken language profiles of children with typical development and spoken language impairment in the experimental and comparison conditions

	Experimental Groups A and B Combined		Comparison Group C	
	TD (*n* = 27)	SLI (*n* = 7)	TD (*n* = 74)	SLI (*n* = 21)
CELF-P2				
Receptive				
M	105.96	90.57*	102.13	85.29**
SD	4.37	4.24	8.95	4.00
Range	93–115	85–95	86–119	79–95
Expressive				
M	104.67	84*	100.32	84.71**
SD	3.66	2.83	8.88	3.48
Range	98–111	83–89	86–116	82–93
PIPA				
Rhyme oddity				
M	6.93	3.14*	6.96	4.10**
SD	1.36	1.77	2.33	1.58
Range	5–9	0–6	3–13	2–7
Initial phoneme				
M	9.22	2.57*	9.84	6.76**
SD	1.31	3.21	2.44	1.30
Range	4–11	0–6	3–15	6–10

Letter-sound				
M	7.30	3.86*	7.8	5.05**
SD	0.78	1.86	1.86	1.12
Range	6–8	0–6	4–11	4–7
PTONI				
M	108.93	97.43*	105.59	94.43**
SD	2.81	3.82	5.69	6.30
Range	103–115	92–102	87–116	85–103
PCC				
M	99.06	75.33*	97.59	77.86%**
SD	1.44	12.77	5.56	11.83
Range	96.2–100	63.3–89	94–100	65–98

Note. CELF-P2 Receptive = Receptive Language Index of the Clinical Evaluations of Language Fundamentals—Preschool, Second Edition (*M* = 100, *SD* = +/-15) (Wiig et al., 2006); CELF-P2 Expressive = Expressive Language Index of the Clinical Evaluations of Language Fundamentals—Preschool, Second Edition (*M* = 100, *SD* = +/-15) (Wiig et al., 2006); PIPA = Preschool and Primary Inventory of PA (*M* = 10, *SD* = +/-3) (Dodd et al., 2000); PTONI = Primary Test of Non-verbal Intelligence (*M* = 100, *SD* = +/-15) (Ehrler & McGhee, 2008); TD = typical development; SLI = spoken language impairment; * indicates that children with SLI in the experimental condition (i.e., Groups A & B) performed significantly lower than children with TD in the experimental condition (p <.001); **indicates children with SLI in comparison Group C performed significantly lower than children with TD in comparison Group C (p <.001).

According to baseline assessment results, children with TD in the experimental condition performed significantly higher than experimental children with SLI on measures of receptive language ($t(32) = 8.35$, $p < .0001$), expressive language ($t(32) = 13.85$, $p < .0001$), rhyme oddity ($t(32) = 6.18$, $p < .0001$), initial phoneme identity ($t(32) = 8.60$, $p < .0001$), letter-sound knowledge ($t(32) = 7.59$, $p < .0001$), non-verbal intelligence ($t(32) = 8.96$, $p < .0001$) and PCC ($t(32) = 9.85$, $p < .0001$). Similarly, children with TD in comparison Group C performed significantly higher than comparison children with SLI on measures of receptive language ($t(93) = 8.36$, $p < .0001$), expressive language ($t(93) = 7.86$, $p < .0001$), rhyme oddity (*t*(93) = 3.05, $p < .0001$), initial phoneme identity (*t*(93) = 5.55, $p < .0001$), letter-sound knowledge (*t*(93) = 6.44, $p < .0001$), non-verbal intelligence (*t*(93) = 7.75, $p < .0001$) and PCC (*t*(93) = 10.82, $p < .0001$).

Children with TD in the experimental and comparison conditions performed within the average range on all baseline measures, except for on the rhyme oddity task in the PIPA (Dodd et al., 2000). The typical range for standard scores on this task is 7–13. Both experimental and comparison children with TD obtained a mean score just below seven (see Table 5.1).

Comparison of children with SLI in the experimental and comparison conditions revealed that both groups on average performed within normal limits in receptive language development and non-verbal intelligence, according to the ranges reported in the standardised measures. Although performing outside the average range, children with SLI in the experimental and comparison conditions were not significantly different to each other on measures of rhyme oddity (*t*(26) = 1.35, $p = .19$), letter-sound knowledge (*t*(26) = 2.05, $p = .06$) or speech sound production (*t*(26) = 0.48, $p = .64$). Children with SLI in the comparison condition, although performing lower than the average range, did perform significantly higher than children with SLI in the experimental condition on initial phoneme identity (*t*(26) = 5.00, $p < .0001$). Given this advantage in phoneme-level knowledge, end-of-year results demonstrating significantly higher phoneme-level knowledge for experimental children with SLI over comparison children with SLI, would add strength to experimental findings. This is because experimental children with SLI would need to ‘catch-up’ and then ‘over-take’ comparison children with SLI. Moreover, these differences in incoming language skills reflect the heterogeneous nature of SLI and the types of issues teachers encounter in the classroom environment.

The study reported in Chapter 4 employed broad inclusion criteria to ensure the selected sample represented the spectrum of abilities found in a typical classroom environment. Thus, it was anticipated that the language profiles of children with SLI would be heterogeneous. That is, some children with SLI may present with deficits in expressive language, receptive language, and phonology, PA or a mixture of deficits in more than one language domain. The spoken language profiles of children with SLI in the experimental and comparison conditions are illustrated in Table 5.2. This table shows that children with SLI in the experimental condition had deficits primarily in the domains of expressive language (71.43 per cent), speech sound production (100 per cent) and PA (rhyme oddity = 100 per cent; initial phoneme identity = 100 per cent and letter-sound = 100 per cent). Children with SLI in the comparison condition presented with a more varied profile of weaknesses in their receptive language (52.3 per cent), expressive language (57.14 per cent), speech production (76.19 per cent) and PA (rhyme oddity = 85.71 per cent; initial phoneme identity = 71.43 per cent and letter-sound = 80.95 per cent).

Table 5.2

Spoken language profiles of children with spoken language impairment in experimental and comparison conditions

SLI Experimental Groups A & B (*n* = 7)	Gender	CELF-P2 RLI	CELF-P2 ELI	PCC	PIPA		
					RO	IPI	LS
1A*	F	89	86	63	6	0	6
2A	M	85	83	64	0	0	4
3A	M	95	83	89	4	6	4
4A	F	93	84	88	3	0	0
1B**	F	93	89	79	3	0	5
2B	M	85	80	64	3	6	4
3B	M	94	83	70	3	6	4
% of participants **not** WNL		0%	71.43%	100%	100%	100%	100%
SLI Group C (*n* = 21)							
% of participants **not** WNL		52.3%	57.14%	76.19%	85.71%	71.43%	80.95%

Note. *1A = ‘A’ indicates that this child with SLI in the experimental group belonged to Group A in the main study (see Chapter 4); **1B = ‘B’ indicates that this child with SLI in the experimental group belonged to Group B in the main study; CELF-P2 RLI = Receptive Language Index of the Clinical Evaluations of Language Fundamentals—Preschool Second Edition (*M* = 100, *SD* = +/-15) (Wiig et al., 2006); CELF-P2 ELI = Expressive Language Index of the Clinical Evaluations of Language Fundamentals—Preschool Second Edition (*M* = 100, *SD* = +/- 15) (Wiig et al., 2006); PCC = percentage of consonants correct as measured by the NZAT (MOE, 2004), where a PCC above 93 per cent is considered typical development (Shriberg et al., 1997); PIPA = Preschool and Primary Inventory of PA, where RO = rhyme oddity, IPI = initial phoneme identity, LS = letter-sound knowledge (*M* = 10, *SD* = +/- 3) (Dodd et al., 2000); ‘Red’ font indicates performance below the average range.

An important and potentially confounding variable for children with SLI is whether they are receiving specialised support outside the classroom, and what that support entails. In New Zealand, the Ministry of Education's Group Special Education service provides speech-language therapy support to children struggling with spoken language development. Support may be provided: 1) directly, from a speech-language therapist; 2) in a consultative model, in which the speech-language therapist supports the classroom teacher, or 3) in a framework within which a communication support worker (i.e., a paraprofessional) implements activities provided by a therapist that target spoken language goals. Experimental and comparison children with SLI were receiving the following levels of support in addition to the classroom programme:

- **Experimental children with SLI:** In this condition, two children were receiving speech-language therapy for vocabulary development and were seen once by their regular speech-language therapist during the 10-week block of classroom PA instruction. Two children were receiving speech-language therapy for speech sound development but were on a break from therapy during classroom PA instruction. Three children had been referred for an assessment, or were on the waiting list, for specialised support for language development.
- **Comparison children with SLI:** In this condition, five children were receiving 30-minutes of speech-language therapy on a fortnightly to monthly basis for semantic or syntactic language development. Two children were receiving support for speech sound development from a communication support worker using a programme provided by the school's speech-language therapist. Five children had been referred for speech-language therapy support. Nine children had not been referred for speech-language therapy support.

5.2.2 Procedure.

The procedure was described in the main study reported in Chapter 4 (see Section 4.3). A quasi-experimental design with a delayed-treatment component was employed to compare PA, reading and spelling development between children who received classroom-wide PA instruction and children who continued with the ‘usual’ literacy curriculum. The responses of children with SLI in the experimental and comparison conditions were extracted for analysis.

All children received a baseline assessment of language, PA, speech and non-verbal intellectual abilities. The computer-based PA screening and monitoring tool was used to measure response to classroom PA instruction and response to the ‘usual’ literacy curriculum. Tasks from the CBA reported in Chapters 2 and 3 were used to measure progress in PA development and included: rhyme oddity, initial phoneme identity, final phoneme identity, phoneme blending, phoneme deletion and phoneme segmentation. According to data presented in Chapter 6, these tasks were, for the most part, age-appropriate for children during the timeframe when classroom PA instruction was being implemented (i.e., five years and six months of age). Informal measures of reading and spelling (i.e., real and non-word) were also used to track development in literacy skills (see Chapter 4, Section 4.3.2). After one year of schooling, when participants were six years of age, the NARA was administered to obtain a standardised measure of reading accuracy and reading comprehension ability. This would help profile the nature of inequalities in reading outcomes between children with SLI who did and did not receive classroom PA instruction.

5.2.3 Reliability.

Reliability procedures are reported in Chapter 4 (see Section 4.3.3) and are identical for data reported in this chapter. Twenty per cent of baseline data, pre- and post-instructional data and end-of-year data for experimental and comparison children with SLI and TD were reviewed using DVD recordings by an independent examiner who was blind to participant group status. This was to ensure accurate classification of SLI and TD and to ensure precise tracking of changes in PA, reading and spelling abilities over time. Inter-rater agreement between real-time data collection and reviews of DVD recordings by the independent examiner was 100 per cent for baseline data, 98.7 per cent for pre- and post-instructional data and 99.2 per cent for end-of-year data. Discrepancies were discussed until 100 per cent agreement was obtained for all data reviewed. Treatment fidelity data for children exposed to classroom PA instruction and the 'usual' literacy curriculum is provided in Chapter 4, Section 4.3.7.

5.3 Results

To evaluate the effect of classroom PA instruction on reading outcomes for children with SLI, data was analysed to determine: 1) pre- to post-instructional improvements, 2) development in PA, reading and spelling, compared to experimental children with TD and 3) end-of-year reading performance relative to comparison children with SLI, comparison children with TD and experimental children with TD. Paired t-tests and independent group t-tests were used to identify significant differences in mean performance using SPSS (Version 19.0). As mentioned in previous chapters, to determine the magnitude of differences between groups, the effect size index Cohen's *d* was calculated (Portney & Watkins, 2009). Conventional values of interpreting Cohen's *d* were used as follows: small effect size d = 0.20, medium d = 0.50 and large d = 0.80 (Cohen, 1988). According to Hattie (2009) effect sizes greater than 0.4 are desirable in education research and effect sizes greater than 1.0 are considered indicative of a one standard deviation increase or 50 per cent improvement rate in student achievement.

5.3.1 Pre- to post-instructional improvement for experimental children with spoken language impairment.

Paired t-tests were used to investigate pre-to-post instructional improvement for children with SLI who received classroom PA instruction either in the second term (i.e., children with SLI in Group A) or the third term (i.e., children with SLI in Group B) of the first year at school. This data was collected immediately before (i.e., within one week of instruction commencing) and after (i.e., within one week of instruction ceasing) instruction. Table 5.3 presents the pre- and post-instructional mean scores for experimental children with SLI on PA, reading and spelling measures.

Table 5.3

Pre- to post-instructional performance for children with spoken language impairment who received classroom phonological awareness instruction

	RO*		IPI*		FPI*		PB*		PD*		PS*		LN*		LS*	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
M	4.43	7.14	2.29	7.86	0.57	5.71	4.14	10.14	1.86	5.29	1.86	8.71	7.86	14	3.57	11.86
SD	3.51	1.57	0.76	1.86	0.79	1.11	3.18	2.54	1.86	1.60	1.77	1.38	4.10	1.53	3.15	1.35
Range	1–10	5–10	1–3	5–10	0–2	4–7	1–9	6–14	0–5	3–8	0–4	7–11	3–13	13–16	1–8	10–13
Significance	$p = .09$		$p < .0001$		$p < .0001$		$p = .0021$		$p = .003$		$p < .0001$		$p < .003$		$p < .0001$	

	Reading				Spelling			
	Real-word		Non-word		Real-word		Non-word	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
M	3.86	17*	1	4*	1	5.43*	1.29	9.86*
SD	3.48	2.16	1.41	1.53	0.82	0.98	1.25	2.67
Range	1–9	14–20	0–3	2–6	0–2	4–7	0–3	6–13
Significance	$p < .0001$		$p = .0025$		$p < .0001$		$p < .0001$	

Note. *measures from the computer-based PA under investigation in this thesis project; RO = rhyme oddity; IPI = initial phoneme identity; FPI = final phoneme identity; PB=phoneme blending; PD = phoneme deletion; PS = phoneme segmentation; LN = letter-sound; LS = letter-sound; Real-word reading = The Burt Word Reading Test (Gilmore et al., 1981); Non-word reading = Calder non-word reading probes (Calder, 1992); Real-word spelling = Schonell (Schonell, 1932); Non-word spelling = Pseudoword Spelling Subtest of the TOPA-2+ (Torgesen & Bryant, 2004).

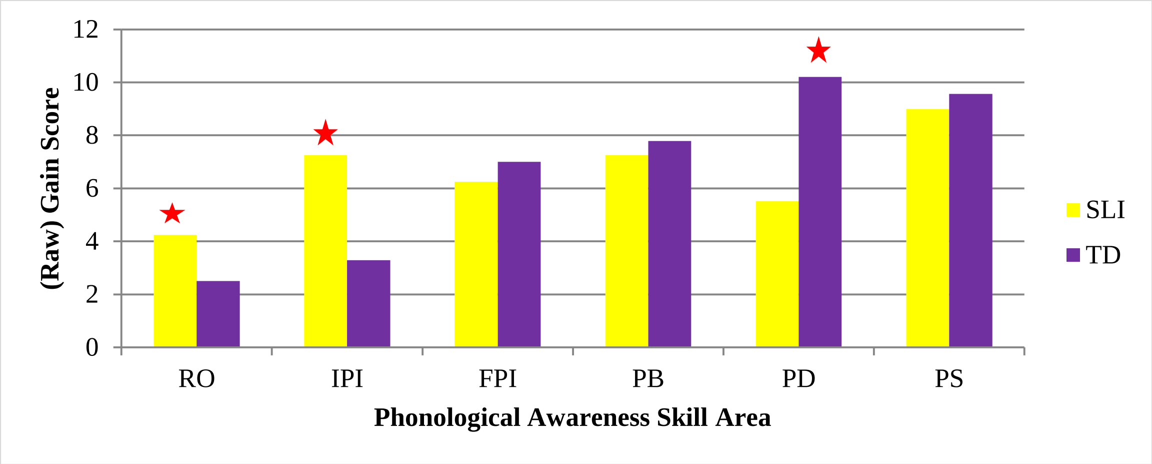
Table 5.3 demonstrates that children with SLI who received 10-weeks of classroom PA instruction showed significant improvements on measures of: initial phoneme identity ($t(6) = 7.33$, $p < .0001$, $d = 3.92$), final phoneme identity ($t(6) = 9.98$, $p < .0001$, $d = 5.34$), phoneme blending ($t(6) = 3.90$, $p = .002$, $d = 2.08$), phoneme deletion ($t(6) = 3.70$, $p = .003$, $d = 1.98$), phoneme segmentation ($t(6) = 8.08$, $p < .0001$, $d = 4.32$), letter-name recognition ($t(6) = 3.71$, $p = .003$, $d = 3.03$), letter-sound knowledge ($t(6) = 6.40$, $p = < .0001$, $d = 5.23$), real word reading ($t(6) = 8.49$, $p < .0001$, $d = 4.54$), non-word reading ($t(6) = 3.81$, $p = .0025$, $d = 2.04$), real word spelling ($t(6) = 9.17$, $p < .0001$, $d = 4.90$) and non-word spelling ($t(6) = 7.69$, $p = .0001$, $d = 4.11$). According to the conventional values for Cohen's *d* and Hattie's (2009) educational research, effect sizes were large. Paired t-tests did not reveal any significant pre-to post-instructional differences for rhyme oddity ($t(6) = 1.86$, $p = .09$, $d = 1.0$).

5.3.2 Growth in phoneme awareness, reading and spelling.

Gain scores were calculated to measure growth in response to classroom PA instruction and to determine whether children with SLI benefited equally from instruction in comparison to children with TD. Gain scores were calculated from pre- to post-instruction for experimental children with SLI and TD. For example, gain score = pre-instructional score minus post-instructional score (Portney & Watkins, 2009). Independent sample t-tests on gain scores showed that children with SLI and children with TD differed in how they benefitted from classroom PA instruction. Children with SLI and TD appeared to gain equally in the development of deeper-level phoneme awareness skills, including phoneme blending ($t(32) = 0.69$, $p = .50$, $d = 0.24$) and phoneme segmentation ($t(32) = 1.22$, $p = .23$, $d = 0.43$). Children with TD showed significantly more growth in phoneme deletion ($t(32) = 8.83$, $p < .0001$, $d = 3.12$). This skill was not taught as part of the classroom PA programme, suggesting that children with TD were more readily able to transfer PA knowledge to an untrained task. Children with SLI demonstrated significantly more growth on measures of rhyme oddity ($t(32) = 3.11$, $p = .004$, $d = 1.10$) and initial phoneme identity ($t(32) = 8.43$, $p < .0001$, $d = 2.98$) compared to children with TD. Children with TD were approaching mastery of, or had already mastered, these skills before instruction. Therefore, they had less potential for gain on these tasks compared to children with SLI.

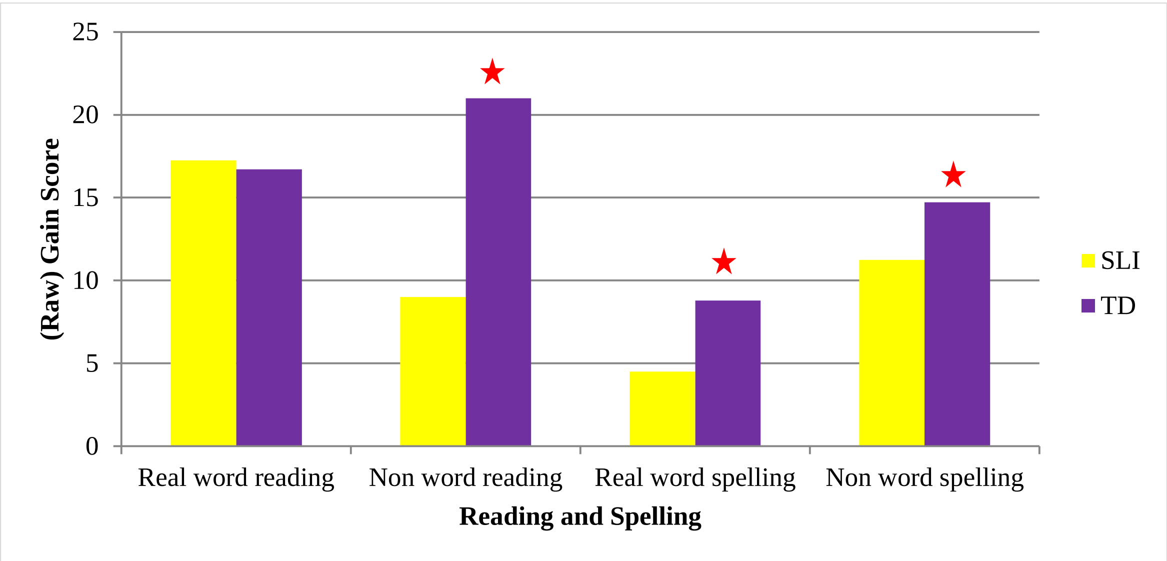
Children with TD showed significantly greater gains in reading and spelling development compared to children with SLI. Specifically, children with TD produced significantly higher gain scores on measures of non-word reading ($t(32) = 4.27$, $p = .0002$, $d = 1.51$), real word spelling ($t(32) = 9.20$, $p < .0001$, $d = 3.25$) and non-word spelling ($t(32) = 3.06$, $p = .004$, $d = 1.08$). The resulting effect sizes were large. Comparison of gain scores in real word reading revealed no significant differences between children with SLI and TD ($t(32) = 0.50$, $p = .62$, $d = 0.18$). The *Burt Word Reading Test* (Gilmore et al., 1981), which was used to measure real word reading, may have been too difficult for this age group, resulting in low and non-significant results. Further, when scoring the *Burt Word Reading Test* (Gilmore et al., 1981), a point was given for each whole word read correctly, whereas for non-word reading a point was given for each correct phoneme–grapheme conversion. This may have also contributed to low and non-significant real word reading results.

Overall, these results suggest that children with TD in this sample were more readily able to transfer their enhanced PA knowledge to reading and spelling tasks. Although children with SLI made significant improvements in reading and spelling relative to their own pre-instructional abilities, they did not demonstrate as much growth in these literacy areas as did children with TD who received the same classroom PA instruction. Figure 5.2 and 5.3 compare the gain scores for children with TD and SLI who received classroom PA instruction in the first year of school.



Note. SLI = spoken language impairment; TD = typical development; RO = rhyme oddity; IPI = initial phoneme identity; FPI = final phoneme identity; PB = phoneme blending; PD = phoneme deletion; PS = phoneme segmentation; ★ = significantly higher gain score.

Figure 5.2. Gain in phonological awareness development for children with spoken language impairment and typical development



Note. SLI = spoken language impairment; TD = typical development; ★ = significantly higher gain score.

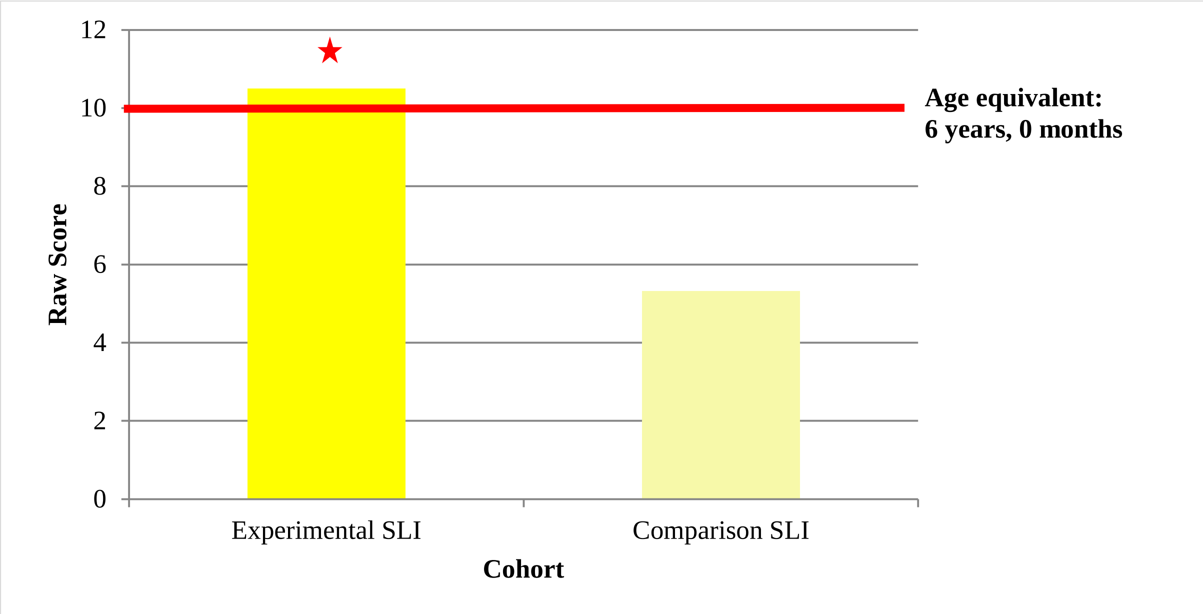
Figure 5.3. Gain in reading and spelling development for children with spoken language impairment and typical development

Although the PA abilities of children with SLI can be successfully raised in a time-efficient manner in the classroom, these children appear to require further support in transferring PA knowledge to untaught phonological skills and reading and spelling tasks if they are to benefit equally from classroom PA instruction.

5.3.3 Effect on reading outcomes and reducing inequality in reading achievement.

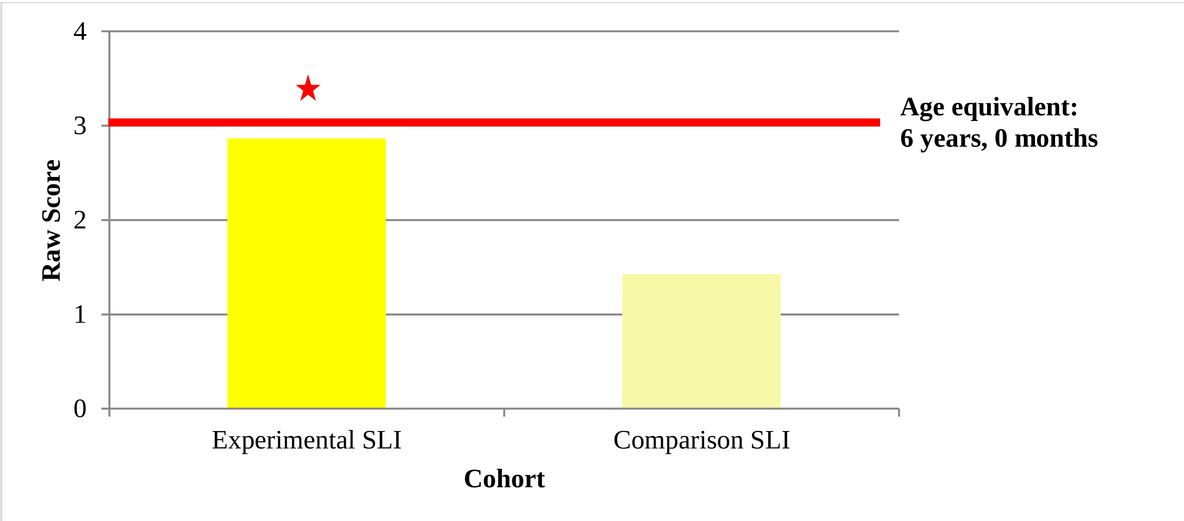
A primary objective of this thesis is to contribute empirical evidence that will support initiatives aimed at raising achievement and reducing disparities in reading outcomes for school-aged children. Children with SLI are one cohort who often experience reading disability. Therefore, raising their achievement to an age-expected level would represent a reduction in reading inequality. To evaluate this, the end-of-year reading accuracy and comprehension scores of the children with SLI who received classroom instruction (i.e., the experimental children) were compared against the score of: 1) comparison children with SLI, 2) comparison children with TD and (c) experimental children with TD.

1) **Reading accuracy and comprehension results for children with SLI in the experimental and comparison conditions:** Children with SLI who received classroom PA instruction performed significantly higher on measures of reading accuracy ($t(26) = 3.77$, $p = .0008$, $d = 1.98$) and reading comprehension ($t(26) = 3.27$, $p = .0031$, $d = 1.58$) after one year of schooling in comparison to children with SLI who continued with the 'usual' classroom literacy programme. This result suggests that a short and intensive PA programme, as part of the beginning literacy curriculum, can reduce inequalities in reading achievement between what children with SLI may currently experience and what they can potentially achieve. Effect sizes according to Cohen (1988) and Hattie's (2009) educational research are considered large.



Note. ★ = significantly higher end-of-year reading accuracy score.

Figure 5.4. End-of-year reading accuracy scores for children with spoken language impairment



Note. ★ = significantly higher end-of-year reading comprehension score.

Figure 5.5. End-of-year reading comprehension scores for children with spoken language impairment

Reading accuracy and comprehension for experimental children with SLI and comparison children with TD: Experimental SLI children performed at a similar level in reading accuracy ($t(79)= 0.89$, $p = .38$, $d = 0.47$) and comprehension ($t(79) = 1.63$, $p = .11$, $d = 0.82$) to children with TD who received the 'usual' classroom literacy instruction. This is despite performing significantly lower on measures of PA and letter-knowledge at the start of the school year (see Table 5.1) and prior to instruction. In addition, the mean performance on reading accuracy and comprehension for children with SLI in the experimental condition was age-appropriate on a standardised reading measure. These results indicate that a time-efficient 10-week period of teacher-directed PA instruction helped reduce the gap in performance between children with SLI and children with TD, when children with TD continued with the 'usual' curriculum. In accordance with Cohen (1988) and Hattie (2009), desirable effect sizes were achieved.

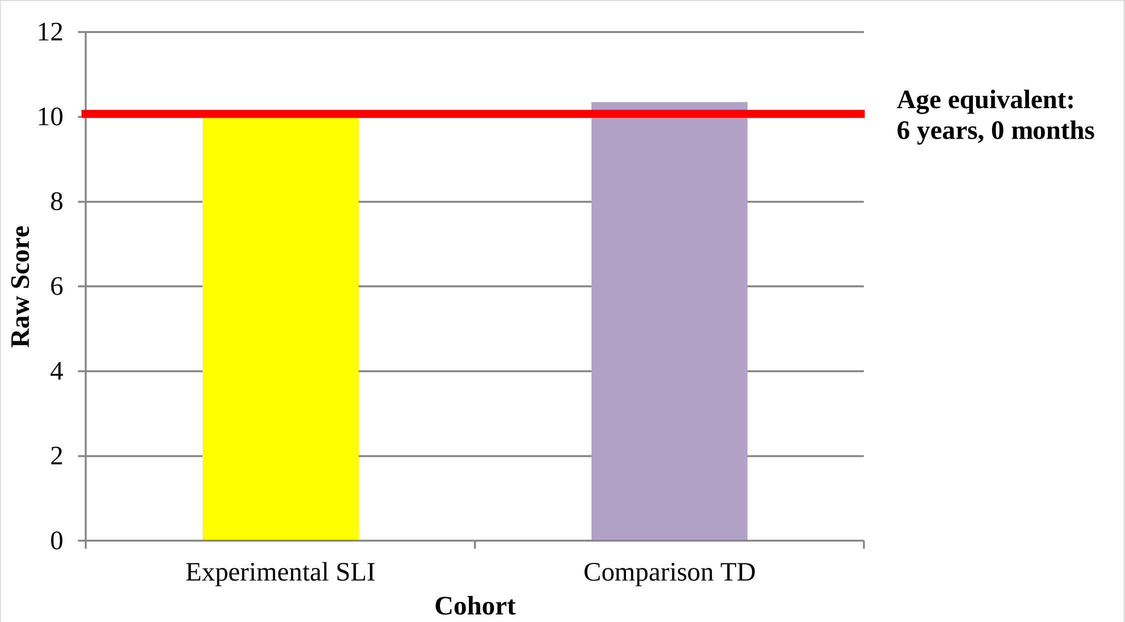


Figure 5.6. End-of-year reading accuracy for experimental children with spoken language impairment and comparison children with typical development

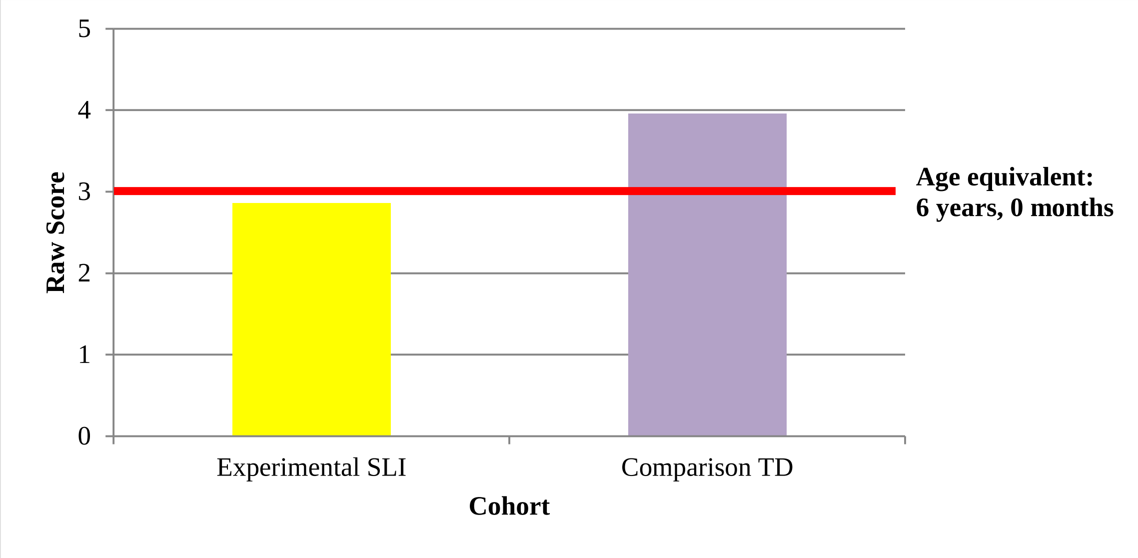
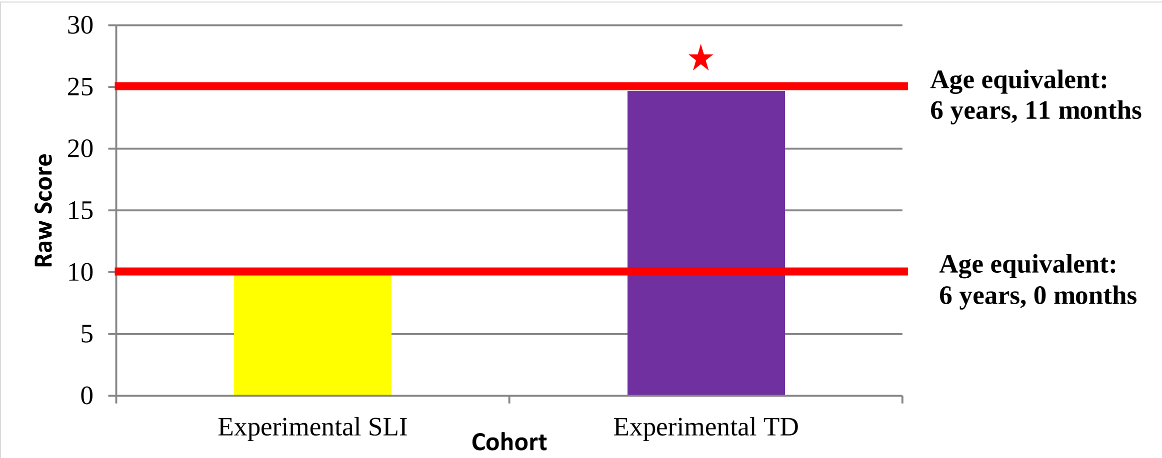


Figure 5.7. End-of-year reading comprehension for children with spoken language impairment and comparison children with typical development

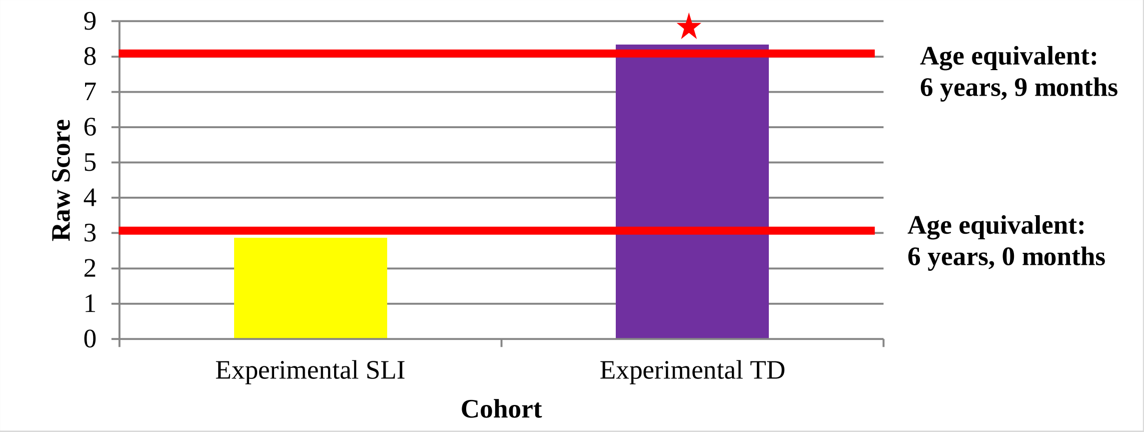
3) **Reading accuracy and comprehension for experimental children with SLI and experimental children with TD:** Experimental children with TD performed significantly higher than experimental children with SLI on end-of-year measures of reading accuracy ($t(32) = 9.74$, $p < .0001$, $d = 5.11$) and comprehension ($t(32) = 9.88$, $p < .0001$, $d = 4.93$). After one year of schooling, experimental children with SLI performed at an age-appropriate level in reading accuracy, and slightly below an age-appropriate level in reading comprehension (i.e., a raw score of three is equivalent to an age-appropriate level for children ages six years and zero months—the children with SLI obtained a mean of 2.89). In contrast, children with TD performed at a six years and 11 months age level in reading accuracy, and at a level of six years and nine months in reading comprehension.

Experimental children with TD performed significantly better than experimental children with SLI on baseline measures of language, PA and speech development (see Table 5.1). It is possible that these differences in spoken language ability at school-entry moderated the extent to which children with TD and children with SLI benefitted from classroom PA instruction, thereby having a subsequent effect on reading ability. Overall, this result suggests that although the reading achievement of children with SLI can be raised to an age-expected level when exposed to explicit and intensive classroom PA instruction, inequalities in reading outcomes may be maintained when children with TD are also exposed to the same instruction. Effect sizes were considered large.



Note. ★ = significantly higher end-of-year reading accuracy score.

Figure 5.8. End-of-year reading accuracy for experimental children with spoken language impairment and typical development



Note. ★ = *significantly higher end-of-year reading comprehension score.*

Figure 5.9. End-of-year reading comprehension for experimental children with spoken language impairment and typical development

These results demonstrate that inequality in reading achievement was influenced in three different ways as a function of classroom reading instruction. First, children with SLI who received classroom PA instruction performed significantly higher in reading accuracy and comprehension than children with SLI who received the 'usual' literacy curriculum. This suggests that phoneme-focused instruction was successful in raising reading achievement. Second, children with SLI who were exposed to classroom PA performed at an age-appropriate level in reading fluency and slightly under an age-appropriate level for reading comprehension after one year of schooling. This is significant in that reading performance was raised to a typically developing level. Finally, when children with SLI and TD were both exposed to the same classroom PA instruction, inequalities in reading outcomes were present. This inequality was characterised by children with TD showing advances of almost one year above chronological age in reading achievement, and children with TD achieving within the average range. Importantly, this result demonstrated that children without risk for reading problems can also benefit from classroom PA instruction; not just vulnerable readers.

In addition, it is also important to evaluate PA performance after one year of schooling. Table 5.4 illustrates the mean performance for each group of children on the computer-based PA screening and monitoring tool described in Chapter 2. The table shows a similar pattern of results to those obtained for end-of-year reading accuracy and comprehension outcomes. For example, experimental children with SLI performed significantly higher than comparison children with SLI and at a similar level to comparison children with TD. Experimental children with TD performed significantly higher than experimental children with SLI on higher-level phoneme awareness skills including phoneme blending, deletion and segmentation.

Table 5.4

End-of-year phonological awareness abilities for experimental and comparison children

PA Task	Experimental SLI (*n* = 7)	Comparison SLI (*n* = 21)	Comparison TD (*n* = 74)	Experimental TD (*n* = 27)
RO				
M	9.29	7.26*	8.63	9.74
SD	0.95	1.63	2.0	0.59
Range	8–10	2–9	1–10	8–10
IPI				
M	9.57	5.47*	9.29	9.89
SD	0.53	2.32	1.42	0.32
Range	9–10	3–10	2–10	9–10
FPI				
M	6.57	2.79*	7.42	9.48*
SD	1.27	2.57	2.36	0.70
Range	5–8	0–8	2–10	8–10
PB				
M	12.86	7.95*	12.75	14.93*
SD	1.07	3.42	3.09	0.27

Range	12–15	5–11	1–15	14–15
PD				
M	10.14	5.21*	9.81	13.40*
SD	1.35	3.01	2.89	1.05
Range	9–12	0–10	5–15	12–16
PS				
M	10.71	5.11*	9.41	13.33*
SD	1.25	2.88	3.68	1.39
Range	10–13	0–11	2–16	11–17

Note. * indicates a statistically significant difference in average group performance in comparison to experimental children with SLI; RO = rhyme oddity; IPI = initial phoneme identity; FPI = final phoneme identity; PB = phoneme blending; PD = phoneme deletion; PS = phoneme segmentation.

5.3.4 Poor responders to classroom phonological awareness instruction.

Reading accuracy and comprehension results from the NARA (Neale, 1999) were analysed after one year of schooling to investigate the proportion of children with SLI who responded poorly to classroom PA. As mentioned in Chapter 4, the NARA (Neale, 1999) provides Reading Age Levels starting from six years and zero months. To be assigned a Reading Age Level children must obtain a raw score greater than 10 for reading accuracy (i.e., 10 = a Reading Age Level of six years and zero months), and greater than three for reading comprehension (i.e., 3 = a Reading Age Level of six years and zero months). After one year of schooling, two children with SLI (i.e., one child from Group A and one child from Group B) who received classroom PA instruction performed below the age-expected level in reading accuracy and reading comprehension (see Table 5.7). Although their raw scores were close to the average range, their performance was not at or above a six-year-old level. Because of this, the developmental progress of these two children from school-entry (i.e., baseline), to pre- and post-instructional assessments, to end-of-year reading outcomes was investigated and compared to the mean group performance for the experimental children with SLI. This was to determine whether any apparent differences in progress over time could have mediated reading outcomes. This analysis is intended to be descriptive in nature. The two children with SLI are referred to as Child A4 (i.e., originally from Group A) and Child B3 (i.e., originally from Group B). Table 5.5 profiles the baseline performance of Child A4 and Child B3 in comparison to the mean group performance of experimental children with SLI.

Table 5.5

Comparison of baseline scores for Child A4, Child B3 and experimental children with spoken language impairment

	CELF-P2		PIPA			PCC	PTONI
	RLI	ELI	RO	IPI	LS		
Child A4	93	84	3	0	0	88.2%	101
Child B3	94	83	3	6	4	70.4%	102
Experimental							
SLI group							
(*n* = 7)	90.57	84	3.14	2.57	3.86	75.33%	97.43
M	4.24	2.83	1.77	3.21	1.86	12.77	3.82
SD	85–95	83–89	0–6	0–6	0–6	63.3–89	92–102
Range							

Note. CELF-P2 RLI = Clinical Evaluation of Language Fundamentals—Preschool, Second Edition, Receptive Language Index (Wiig et al., 2006); CELF-P2 ELI = Clinical Evaluation of Language Fundamentals—Preschool, Second Edition, Expressive Language Index (Wiig et al., 2006); PIPA = Preschool and Primary Inventory of PA, where RO = rhyme oddity, IPI = initial phoneme identity and LS = letter-sound knowledge (Dodd et al., 2000); PCC = percentage of consonants correct, as measured by the NZAT (MOE, 2004) and analysed using PROPH (Long et al., 2002); PTONI = Primary Test of Non-verbal Intelligence (Ehrler & McGhee, 2008).

In comparison to the mean group performance of experimental SLI children, Child A4 and Child B3 performed slightly higher in receptive language ability and at a similar level in expressive language. In PA, both children performed on par with the group mean in rhyme oddity, but differed in initial phoneme identity, with Child A4 performing at lower than the group average and Child B3 performing higher than that average. Child A4 performed lower than the group mean in letter-sound ability at school-entry. The speech sound development of both Child A4 and Child B3 varied around the group mean, with Child A4 performing above and Child B3 performing below the group average. Both children performed slightly higher than the group mean in non-verbal intelligence. Overall, Child A4 and Child B3 varied in relation to each other (e.g., one child scored high and the other scored low) and to the group mean, predominantly in the areas of initial phoneme identity and speech sound development.

To evaluate their response to instruction, pre- to post-instructional scores for Child A4 and Child B3 were extracted and compared to the experimental SLI group mean. Table 5.6 presents the raw scores obtained for performance in PA, reading and spelling from pre- to post-instruction.

Table 5.6

Pre- to post-instructional performance for Child A4 and Child B3 with spoken language impairment

	RO*		IPI*		FPI*		PB*		PD*		PS*		LN*		LS*	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Child A4	1	5	1	10	0	6	1	10	1	6	1	8	3	12	1	10
Child B3	8	10	2	5	1	5	3	6	3	3	4	8	3	13	1	10
Experimental SLI group (*n* = 7)	4.43	7.14	2.29	7.86	0.57	5.71	4.14	10.14	1.86	5.29	1.86	8.71	7.86	14	3.57	11.86

	Reading				Spelling			
	Real-word		Non-word		Real-word		Non-word	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Child A4	1	17	0	6	0	5	0	10
Child B3	9	15	3	6	2	6	1	5
Experimental SLI group (*n* = 7)	3.86	17	1	4	1	5.43	1.29	9.86

Note. * measures from the computer-based PA under investigation in this thesis project; RO = rhyme oddity; IPI = initial phoneme identity; FPI = final phoneme identity; PB = phoneme blending; PD = phoneme deletion; PS = phoneme segmentation; LN = letter-sound; LS = letter-sound; Real-word reading = The Burt Word Reading Test (Gilmore et al., 1981); Non-word reading = Calder non-word reading probes (Calder, 1992); Real-word spelling = Schonell (Schonell, 1932); Non-word spelling = TOPA; Non-Word Spelling = Pseudoword Spelling Subtest of the Test of PA—Second Edition (TOPA-2+) (Torgesen & Bryant, 2004).

As indicated in Table 5.6, Child A4 performed on par with or higher than the group mean for experimental SLI children on measures of PA, reading and spelling following instruction. Child B3 generally performed lower than the group mean on measures PA, reading and spelling post-instruction. Therefore, it was anticipated that Child B3, but not Child A4—who had shown greater gain in response to class PA instruction—would perform below an age-expected level in reading fluency and comprehension after one year of schooling. Table 5.7 illustrates the reading fluency and comprehension raw scores for Child A4, Child B3 and experimental SLI children as a group at the end of the school year.

Table 5.7

End-of-year reading accuracy and comprehension ability for children with spoken language impairment

	Accuracy	Comprehension	Reading Age Level
Child A4	9	2	Not specified
Child B3	8	2	Not specified
Experimental SLI group (*n* = 7)	10.5	2.86	6 years, 0 months

Notes. Accuracy = Reading accuracy standard score from the NARA; Comprehension = Reading comprehension standard score from the NARA; Reading Age Level = comparative age-level score from the NARA.

Child A4 and Child B3 both showed improvements in PA, reading and spelling in response to classroom PA instruction. However, neither child reached an age-expected level in reading accuracy and comprehension after one year of schooling. This could be due to a number of variables, not specifically measured in this thesis, including deficits in phonological memory, difficulties processing orthographic information, paucity of attention, behavioural problems, lower intellectual abilities (Al Otaiba & Fuchs, 2002; Nelson et al., 2003) or distractions on the day of testing. These results may not necessarily warrant the provision of further specialised support for these children, but ongoing monitoring and exposure to high quality scientifically driven classroom literacy instruction is necessary to ensure they do not fall significantly behind their peers in reading acquisition.

5.4 Discussion

The experiment reported in this chapter investigated the effect of teacher-delivered classroom PA instruction focused at the phoneme-level on reading outcomes for children with SLI. As described in Chapter 4, 34 children received 10 weeks of classroom PA instruction in the first year of school, while 95 children received the 'usual' literacy curriculum. Data from children with SLI who participated in this study was extracted, analysed and reported above. In total, seven children with SLI and 21 children with TD received classroom PA instruction, while 21 children with SLI and 74 children with TD received the 'usual' literacy curriculum. Data was analysed to examine 1) pre- to post-instructional change in PA for children with SLI, 2) the extent to which instructional benefit was equal between children with SLI and TD and 3) the effect on end-of-year reading outcomes as an indicator of reading inequality after one year of formal schooling.

5.4.1 Significant improvement in phonological awareness and literacy for children with spoken language impairment.

The first hypothesis proposed that children with SLI would show significantly higher scores on phoneme awareness and early literacy measures following a short and intensive period of teacher-directed PA instruction. This hypothesis was supported by data analyses on the small sample size in this study. Children with SLI showed statistically significant improvements on all measures of phoneme awareness, reading and spelling following 10 weeks of classroom PA instruction from their teacher. Improvement in rhyme oddity performance was observed, but not to a statistically significant level. These findings suggest that the benefits of PA instruction focused at the phoneme-level in one-to-one or small group settings for at-risk populations can be realised in the classroom environment when class teachers implement the programme. Moreover, this can be accomplished in a relatively short period of time.

Research has demonstrated that children with SLI are four to five times more likely to experience difficulties in reading development due to deficits in underlying spoken language skills that support written language acquisition (Catts et al., 2001). Further, PA deficiencies are often prevalent in the spoken language profiles of children who are classified as poor readers in the early school years (Catts et al., 1999). The results of the experiment reported in this chapter suggest that when classroom teachers are equipped with short and intensive programmes aimed at raising phoneme awareness, children with SLI are more likely to achieve positive literacy outcomes.

5.4.2 Equality of instructional benefit between children with spoken language impairment and typical development.

The second hypothesis predicted that although children with SLI would show significant improvements in phoneme awareness and early literacy following teacher PA instruction, they would also demonstrate variability in the extent to which they benefitted from instruction, in comparison to children with TD. Statistical analysis supported this hypothesis. Gain score analysis revealed that children with SLI showed significantly more development in initial phoneme identity and rhyme awareness compared to children with TD. This is most likely because children with SLI had more scope for growth in these early PA skills, while children with TD had already been approaching ceiling levels prior to instruction. Further, children with TD and SLI obtained similar gain scores on measures of final phoneme identity, phoneme blending, and phoneme segmentation. These results suggest that children with SLI showed greater or equal benefit on measures of PA that were targeted in classroom instruction when compared to children with TD.

Equal benefit in response to classroom PA instruction is in contrast to a recent study by Justice and colleagues (2010), which found that children with poor language skills appeared to benefit less in the development of phoneme-level knowledge. It is likely that the high-priority focus on phoneme-level skills in the current investigation supported the elevation of phoneme awareness; something not achieved by the broad focus on PA knowledge (e.g., syllables, rhyme and phonemes in addition to other language areas) employed by Justice et al. (2010). Noteworthy, however, is that children with SLI obtained a significantly lower mean gain score than children with TD on the phoneme deletion measure. Phoneme deletion was not targeted as part of the 10-week classroom PA programme. This result suggests that children with SLI are less able to transfer phonological knowledge to untrained PA tasks, an issue that warrants detailed attention in classroom practice.

Moreover, improvement in PA is only useful if it supports children in using a decoding strategy to recognise words and comprehend the meaning of written text (Catts & Kamhi, 2012; Gillon, 2004). Although children with SLI showed equal or greater gain on phoneme-level tasks, they appeared to benefit less from instruction on measures of reading and spelling when compared to children with TD. For example, children with TD achieved significantly higher gain scores in non-word reading, real word spelling and non-word spelling compared to children with SLI. That is, children with poor language skills appeared less able to transfer their enhanced PA knowledge to the process of learning to read and spell.

Previous research indicates that children with lower spoken language capabilities often demonstrate slower and more variable reactions to language and literacy instruction (Fuchs et al., 2002; Justice et al., 2010; O’Connor et al., 1996; Penno et al., 2002). The results from the current investigation are consistent with these findings. Children in this study with SLI showed a variable response to classroom PA instruction in that they achieved significant pre- to post-instructional improvements in phoneme awareness, reading and spelling (relative to their own pre-instructional scores) and equal or greater gain in the phoneme awareness skills that were taught as part of the programme. However, when compared to children with TD, those with SLI showed poorer transfer of enhanced PA skills to an untrained phoneme awareness task and to the reading and spelling process. These results suggest that classroom teachers should place emphasis on ensuring children with SLI are transferring phonological knowledge to written language tasks.

Further, two children with SLI in the experimental condition achieved at slightly below the age-appropriate reading level after one year of formal schooling. Research demonstrates that there is typically a group of children (i.e., approximately 3–5 per cent) who will respond poorly to instruction. Additional factors such as attention, behavioural problems, poor orthographic processing and lower intellectual skills may restrict how well children respond to instruction and the level of individual evaluation they will require (Al Otaiba & Fuchs, 2002; Nelson et al., 2003). Variations in reading outcomes demonstrate that no single approach to reading instruction will work for every child in the classroom. Rather, educators need to be aware of individual strengths and weaknesses and adjust classroom reading programmes accordingly to achieve greater equality in reading outcomes.

5.4.3 Effect of teacher-led phonological awareness instruction on reading inequalities after one year of schooling.

The third hypothesis held that the majority of children with SLI who received classroom PA instruction would approach an age-appropriate level in reading accuracy after one year of schooling, but may still perform lower than children with TD who received the same instruction. End-of-year reading outcomes for experimental children with SLI were compared to comparison children with SLI, comparison children with TD and experimental children with TD. Results showed that the gap in reading outcomes profiled in three different ways between the different cohorts of children.

First, experimental children with SLI showed significantly higher end-of-year reading accuracy and comprehension scores compared to children with SLI who had received the 'usual' literacy curriculum. Second, experimental children with SLI showed similar reading accuracy and comprehension scores to comparison children with TD and were within or close to the average range on a standardised reading measure. Third, inequalities in reading outcomes between experimental children with SLI and experimental children with TD were present after one year of formal schooling.

Although end-of-year inequalities were present between experimental children with SLI and TD, this result indicates that the majority of children in the classroom benefitted from instruction. Children with SLI approached an age appropriate reading range by the end of the year, while children with TD, who were already on a normal trajectory, accelerated beyond the typical reading range. These results provide support for including a short and intensive period of classroom PA instruction focused at the phoneme-level as part of a comprehensive approach to reducing inequalities in reading outcomes.

5.4.4 Classroom implications.

A number of implications for the classroom at both national and international levels can be derived from the research presented in this chapter. Nationally, New Zealand is a country that continues to prefer a whole language approach to reading instruction (Tunmer et al., 2006), and presents with one of the OECD's largest variation in scores between high- and low-ability readers (Martin et al., 2007). Although the results reported in this chapter could be viewed as preliminary to a much larger nation-wide investigation, they provide initial support for the benefits of supplementing a predominantly meaning-based reading curriculum with PA instruction to help raise reading achievement and reduce inequalities for vulnerable readers (e.g., children with SLI). Moreover, children in New Zealand receive Reading Recovery after one year of schooling if they demonstrate poor progress in reading acquisition

(Clay, 2002, 2003). The results presented in this chapter may help promote a shift away from recovery and towards prevention, by ensuring children are on a positive trajectory towards literacy success before the end of the first year of formal schooling.

Internationally, recent studies of reading achievement reveal that many of the world's wealthiest countries exhibit substantial inequalities between the reading outcomes of good and poor readers (UNESCO, 2009; UNICEF, 2010). Children with a disability (e.g., SLI), or who come from a lower socio-economic backgrounds, an indigenous or minority group or a rural or remote community, are among those children more likely to experience reading problems (Martin et al., 2007; UNESCO, 2009; UNICEF, 2010). Boys are also highly susceptible to reading problems (Martin et al., 2007). Inequality in reading outcome was investigated for one cohort of at-risk children in this chapter—children with SLI. The results reported here indicate that reading inequalities for at-risk children can be reduced when classroom programmes are inclusive of skills that are highly predictive of early literacy success. Moreover, this can be achieved in a short period, thereby proving suitable for implementation in the time-poor classroom environment. Replication of these results with other at-risk groups, nationalities, and as an adjunct to similar classroom literacy curriculums would support international efforts aimed at the early identification and reduction of reading difficulties.

5.4.5 Limitations and future directions.

The benefits for reading outcomes for children with SLI reported in this chapter must be acknowledged in the context of study limitations. Generalising the results of this study is limited by the use of a small sample size of children with SLI. This sample included seven children from two different classrooms (i.e., Group A, $n = 4$; Group B, $n = 3$) who received the same classroom PA programme, but in different terms of the school year (i.e., Group A = term two; Group B = term three). Future investigations should ideally include a larger sample size in a delayed-treatment design or in an experimental-comparison group design using real classroom environments.

Another limitation was that children with SLI in the comparison condition presented with a more varied profile of spoken language difficulties (see Table 5.2), including deficits in receptive language abilities. Experimental children with SLI did not present with deficits in the receptive language domain. It is possible that deficits in receptive language influence how well children can process instructions and respond to feedback during the classroom programme. Further, research shows that children with more widespread language difficulties are at increased risk for literacy difficulties (Lewis et al., 2000). Thus, although results indicated that experimental children with SLI performed significantly higher than did comparison children with SLI on end-of-year reading accuracy and comprehension measures, this figure might have been inflated due to differences in receptive language skills. Future investigations can address this limitation by examining whether different spoken language deficits (e.g., receptive, expressive or mixed language deficits) can exert a significant effect on response to classroom PA instruction. Investigating strategies to assist young children with SLI to transfer phonological knowledge into a decoding strategy when reading and spelling in the classroom, or investigating ways of ensuring the maintenance of skills two to three years after a programme's conclusion would also be areas worthwhile for future research.

In summary, the results of this chapter provide evidence to support the inclusion of a short and intensive teacher-implemented period of PA instruction focused at the phoneme-level in the beginning-reading curriculum for both typically developing and at-risk children. PA skills are often deficient in vulnerable readers and are a key catalyst for poor word-recognition development, which in turn has a negative effect on reading comprehension ability (Bradley & Bryant, 983; Catts, Kamhi, & Adlof, 2012). Ensuring classroom teachers are equipped with time-efficient methods of addressing deficits in these skills not only ensures children at risk for reading difficulties approach instruction with prerequisite foundational skills, but also ensures such programmes are easily implemented within a busy classroom environment. Although no single approach to reading instruction is sufficient to meet the literacy needs of every child in the classroom, the inclusion of PA as part of a comprehensive multi-focal curriculum is a promising area for further investigation aimed at raising achievement and equalising reading outcomes between typically developing and at-risk readers in the classroom.

Chapter 6: Establishing the Reliability and Validity of the Computer-Based Phonological Awareness Screening and Monitoring Tool

6.1 Introduction

In the development of the computer-based phonological awareness (PA) screening and monitoring tool described in Chapter 2 it was necessary to ensure that the tool had an adequate degree of validity and reliability. Validity refers to the extent to which an instrument measures that which it is intended to measure (Messick, 1995), while reliability refers to how consistent an instrument is over repeated administrations of the test, either in the same or in an alternative form (Crocker & Algina, 1986). Providing evidence of validity and reliability ensures teachers can confidently use the computer-based PA tool as part of classroom practice. Doing so allows teachers to predict which children are at risk for reading difficulties, monitor progress to identify children not keeping pace with their peers, and profile strengths and weaknesses in PA that will inform curriculum design. The aim of this chapter, therefore, is to provide evidence of the validity and reliability of the computer-based PA screening and monitoring tool as a time-efficient adjunct to current classroom assessment practices.

The validity of a measurement tool can be examined using content, construct and criterion validity (Kane, 1992; Messick, 1995). Content validity refers to the systematic evaluation of the content within a test to ensure it accurately samples the trait being measured (Anastasi & Urbina, 1997). For example, does the computer-based screening and monitoring tool include tasks and items that, according to the literature, are appropriate for five-year-old children? Item analysis is a fundamental source of content validity (Lissitz & Samuelsen, 2007).

Construct validity refers to the degree to which a test measures the construct it intends to measure (Thorndike & Thorndike-Christ, 2010). For example, does the computer-based screening and monitoring tool actually measure the construct of PA? Factor analysis provides evidence as to whether tasks within a test measure the same underlying construct. Criterion validity refers to the relationship between the test and another variable (i.e., criterion variable), and includes predictive and concurrent validity (Thorndike & Thorndike-Christ, 2010). Predictive validity uses test data to predict a future independent outcome. For example, does performance on the computer-based PA tool accurately predict word-decoding ability after one year of schooling? Multiple regression analysis and evaluation of test sensitivity and specificity provide evidence of predictive validity. Concurrent validity involves the collection of the test data and criterion data at the same time (Thorndike & Thorndike-Christ, 2010). The measure used to collect criterion data is already held to be valid. For example, does the computer-based PA tool produce similar results to an existing valid and reliable PA test? Test data and criterion data can be correlated to provide evidence of concurrent validity.

The reliability of an instrument can be determined using measures of internal consistency and test-retest reliability (Field, 2009). Internal consistency refers to the degree of correlation between items within a test and the consistency of responses between items (Thorndike & Thorndike-Christ, 2010). For example, are tasks in the computer-based PA tool highly correlated and do they measure the same latent trait? Calculation of Cronbach's alpha (α) provides evidence of internal consistency. Test-retest reliability refers to how consistent a test is over repeated administrations under identical conditions (Portney &

Watkins, 2009). Test-retest reliability can be evaluated by correlating the results from successive administrations of the computer-based screening and monitoring tool throughout the first year of school.

In this chapter, evidence is presented to support the use of the computer-based PA screening and monitoring tool as a valid and reliable time-efficient measure for use in the classroom. Specifically, this chapter addresses the following questions:

6.1.1 Content validity.

1. Do the test items demonstrate an appropriate 'fit' for five-year-old children in the first year of school?
2. Do the test items sample a range of difficulties to help ensure differentiation between high- and low-ability students?

   It was hypothesised that the majority of test items would be appropriate for the abilities of five-year-old children and would sample a range of difficulty levels.

6.1.2 Construct validity.

3. Do the test items measure a uni-dimensional construct of PA?
4. Do the PA tasks represent a continuum of difficulty over the first year of school?

   It was hypothesised that test items would measure a uni-dimensional construct of PA and that tasks would represent a continuum of difficulty throughout the first year of formal schooling.

6.1.3 Criterion validity.

5. Is the computer-based PA screening and monitoring tool predictive of end-of-year word-recognition ability?
6. Are the PA tasks correlated with an existing instrument already held to be valid and reliable?

It was hypothesised that tasks in the computer-based PA tool would be highly predictive of end-of-year word-recognition ability and would correlate strongly with an existing standardised PA instrument.

6.1.4 Reliability.

7. Do the test items consistently measure the same latent trait?
8. Are the PA tasks consistent across repeated administrations?

   It was hypothesised that tasks in the computer-based PA tool would demonstrate satisfactory levels of internal consistency and test-retest reliability throughout the first year of formal schooling.

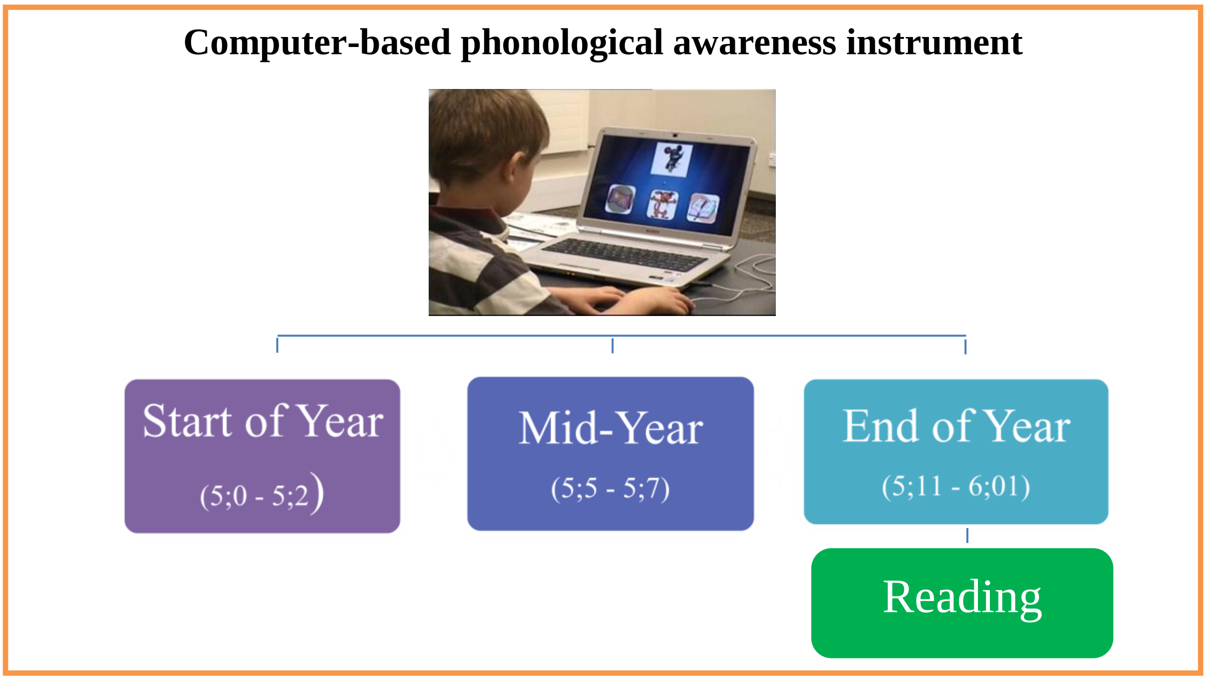
6.2 Method

6.2.1 Participants.

The 95 children who participated as comparison Group C in Chapter 4 were used as the sample from which data were extracted to examine the reliability and validity of the computer-based PA tool. These children received the 'usual' literacy curriculum in their first year of formal schooling and had a mean age of five years and zero months to five years and two months at school-entry. There were 39 boys and 56 girls in this sample. Seventy-four children presented with TD and 21 children presented with deficits in at least one spoken language domain. This cohort of children was chosen to establish the reliability and validity of the computer-based tool because they were receiving what is typically viewed as the 'usual' approach to classroom reading instruction, in a New Zealand new entrant classroom. Therefore, within the context of study limitations, this sample provided a population to which study results could be immediately applied (see Chapter 4, Section 4.2).

6.2.2 Procedure.

In Chapter 4, children in Groups A, B and C participated in a quasi-experimental design that included a delayed-treatment approach for children in Groups A and B. The PA development of children in Group C was tracked throughout the first year of school. This means that when Group C's data is examined in isolation, it is as though this cohort participated in a longitudinal research design. The computer-based PA assessment described in Chapter 2 was administered to children in Group C at the start (i.e., approximately five years and zero months), middle (i.e., approximately five years and six months) and end (i.e., approximately six years and zero months) of the first year at school. These children continued with the 'usual' literacy curriculum, which primarily focused on whole language instruction and included a secondary focus on teaching phonics. Explicit instruction in PA was not provided. Figure 6.1 illustrates the implementation of the longitudinal design.



Note. Age is represented as years; months.

Figure 6.1. Longitudinal research design to determine the validity and reliability of the computer-based PA tool

All assessment, reliability and the ‘usual’ curriculum approach were identical to those reported in Chapter 4 (see Section 4.3). Table 4.3 in Chapter 4 profiles the language, PA, speech and non-verbal abilities of children in Group C at school-entry, and Table 4.4 illustrates the performance of this group on the computer-based PA tool throughout the first year at school.

6.3 Results

The responses of 95 children to items in the computer-based PA screening and monitoring tool at the start, middle and end of the first year at school were analysed to investigate validity and reliability. In the following sections, content, construct and criterion validity are successively addressed, followed by evidence of reliability using internal consistency and test-retest reliability.

6.3.1 Content validity.

Content validity refers to the process of ensuring that test content accurately reflects the knowledge being measured (Anastasi & Urbina, 1997). According to Litssitz and Samuelsen (2007), test items are the building blocks of any measurement tool and are therefore a critical source of content validity. Construction of a test with a sufficient degree of content validity includes analysing items within a test to determine whether they are appropriate for the intended population and whether they sample a range of difficulty to enable differentiation between high- and low-ability students (Crocker & Algina, 1986). The Rasch Model (Bond & Fox, 2001) was selected as the primary measurement model to conduct a formal item review.

6.3.1.1 The Rasch Model.

The Rasch Model was used to evaluate the ‘fit’ and difficulty spectrum of PA test items in relation to the ability of five-year-old children in the first year of formal education.

This model provides a theoretical range against which test developers can compare patterns of responses to determine whether items show a 'fit' or 'misfit' to the ability of test takers. Items deviating from the ideal range (i.e., showing a misfit) require adaptation or removal from the measurement tool (Bond & Fox, 2001). The model evaluates a test item in terms of difficulty (the item parameter) and people in terms of ability (the person parameter). An underlying assumption of this model is that the ability (or latent trait) being measured is uni-dimensional (i.e., the items that comprise the test measure the same underlying construct) (Bond & Fox, 2001). The exploratory factor analysis reported in Section 6.3.2.1 confirms that PA, as measured in this thesis, represents a uni-dimensional construct. The Rasch Model is occasionally referred to as the One-Parameter Logistic Model under Item Response Theory (IRT). Despite sharing a mathematical similarity, these models differ on a conceptual level (Baker, 2001). In the Rasch Model, data must conform to the properties of the model for measurement to take place (Andrich, 2004). Items that do not conform to the model (i.e., show a model 'misfit') require careful investigation and an explanation as to why this is the case. In IRT, the importance of data-model fit is emphasised. However, additional model parameters enable the model to adjust to reflect the pattern of the data (Embretson & Reise, 2000).

In this thesis, Rasch Model analysis was conducted on the responses of the 95 children to each test item at the start, middle and end of the school year by entering them into a software programme called *Winsteps* (Version 3.70) (Linacre, 2010). This analysis provided information on which test items showed a model 'fit' or significant 'misfit', and how the test items in each task related to each other in terms of difficulty.

6.3.1.1.1 Determining which items show a 'fit' or significant model 'misfit'.

Rasch Model analysis using *Winsteps* provides several types of statistical analyses to evaluate test items in relation to the latent trait being measured. For the purposes of this thesis, the ‘outfit statistics’ of mean-square and ZSTD for each test item were used to evaluate which items showed a ‘fit’ or significant model ‘misfit’. The relevance of using the mean-square statistic and the ZSTD statistic are described below.

Mean-square statistic (MNSQ): The mean-square statistic draws attention to the accuracy of an item by providing an indication of the size of an item’s ‘misfit’ to the model. An item with a mean square close to 1.0 suggests that the item is accurate. An item with a mean square less than 1.0 is considered less accurate, but this does not cause any real problems (Linacre, 2010, p. 23). An item with a mean square greater than 2.0 is considered inaccurate and in need of attention.

ZSTD statistic: A ZSTD statistic is assigned to each mean-square statistic to indicate whether the size of the ‘misfit’ is statistically significant. The ZSTD is ‘standardized like a Z-score’ (Linacre, 2010, p. 25). An item with a ZSTD statistic between -2 and +2 indicates a statistically significant model ‘fit’.

In line with Linacre (2010), test items with a mean-square statistic greater than 2.0 and a ZSTD statistic less than -2 and greater than +2 were interpreted as showing a statistically significant model ‘misfit’. Items may show a ‘misfit’ for a number of reasons, including: 1) being too easy or difficult, 2) using confusing or ambiguous instructions or 3) lacking in image quality (i.e., animated or static graphics). Items demonstrating a ‘misfit’ may require adaption or deletion from the instrument (Bond & Fox, 2001).

The point-measure correlation is another outfit statistic. It provides an indication of an item’s discrimination ability. The point-measure correlation provides an indication of how much each item is aligned with the ability being measured. For example, do items that are more difficult require more ability to produce a correct response, and do easier items require less ability to produce a correct response. Positive correlations above 0.3 indicate that the item is well correlated to the ability being measured. Negative correlations or correlations close to zero suggest there is little relationship between the item and the ability being measured. This indicates that an item does not effectively distinguish between individuals with more or less ability, which is cause for concern. Small correlations (e.g., 0.14) necessitate further investigation.

The mean-square statistic, ZSTD statistic and point-measure correlation obtained for each test item at the start, middle and end of the first year at school are provided in Appendix C. As an example, Table 6.1 illustrates the mean-square statistic, ZSTD statistic and point-measure correlation obtained for 10 rhyme oddity test items at the start of the school year. Rhyme oddity was considered to be the easiest PA task in the computer-based screening and monitoring tool, as well as appropriate for five-year-old children (see Chapter 2). Therefore, it was anticipated that the majority of rhyme oddity items would demonstrate a model ‘fit’ and point-measure correlations above 0.3.

Table 6.1

'Fit' or 'misfit' for rhyme oddity test items at school-entry

Test Items				Outfit Statistics			Interpretation
Which word does not rhyme?				MNSQ	ZSTD	PT-measure correlation	'Fit' or 'Misfit'
1	cat	mat	bus	4.98	2.81	0.29	Misfit
2	peg	doll	leg	2.72	3.01	0.29	Misfit
3	saw	toe	bow	1.27	0.72	0.66	Fit
4	sand	hand	cup	0.63	-0.61	0.77	Fit
5	hen	car	pen	0.36	-0.81	0.51	Fit
6	dog	book	hook	0.37	-0.92	0.62	Fit
7	bun	sun	kite	0.82	-0.13	0.52	Fit
8	tent	lock	sock	1.66	1.53	0.62	Fit
9	shell	duck	bell	0.33	-1.01	0.58	Fit
10				0.36	-0.91	0.59	Fit
							8/10 items = Fit
							2/10 items = Misfit

Note. MNSQ = mean-square statistic; ZSTD = ZSTD statistic; 'Fit' or 'Misfit' indicates whether the data fit the properties of the Rasch Model; grey shading = correct response to test item; pink shading = items that show a significant model 'misfit' and may not be appropriate for measuring PA ability in five-year-old children.

Table 6.1 demonstrates that eight test items showed a model 'fit' because they had a mean-square statistic less than 2.0 and a ZSTD statistic between -2 and +2. However, items 1 and 2 demonstrated a significant model 'misfit'. Inspection of responses to item 1 revealed that 87 per cent of children obtained a correct score on this item, suggesting that the 'misfit' occurred because the item was too easy. Inspection of responses to item 2 revealed that this item was of average difficulty (see Table 6.5) and that the 'misfit' could not be attributed to the item being too easy or too difficult for five-year-old children. A difference between item 2 and the remaining rhyme oddity items is that one of the distractor options and the correct option had one phoneme in common: 'doll' (correct option) and 'leg' (distractor option) both contain the /l/phoneme. In no other rhyme oddity test item did the distractor option contain an identical phoneme to the correct option. This distractor option (i.e., ‘leg’) may have been too strong and may therefore offer an explanation as to why item 2 showed a ‘misfit’. It is also possible that image quality played a role in the item ‘misfit’. For example, the image of a ‘leg’ was of a leg only, as opposed to being in the context of the body of a person. Further, it may have been too visually similar in length and diameter to the image of the ‘peg’. Detailed inspection of rhyme oddity item 2 and newly developed items containing phonologically similar correct and distractor options will be required in future investigations of the computer-based instrument. In addition, the point-measure correlations for items 1 and 2 were just below 0.3, indicating that these items do not differentiate well between high-and low-ability students.

As another example, Table 6.2 illustrates the mean-square statistic, ZSTD statistic and point-measure correlation for items in the phoneme blending task during the middle of the school year. This table provides an example of a situation where all test items demonstrate an appropriate ‘fit’ to the ability five-year-old participants.

Table 6.2

'Fit' or 'misfit' for phoneme blending test items during the middle of the school year

Test Items				Outfit Statistics			Interpretation
What word am I saying?				MNSQ	ZSTD	PT-measure correlation	'Fit' or 'Misfit'
1	dog	dot	man	0.28	-1.51	0.66	Fit
2	mouse	mouth	ring	1.22	0.61	0.70	Fit
3	seal	seat	duck	0.15	-1.32	0.83	Fit
4	bug	sun	bun	0.21	-0.62	0.88	Fit
5	cat	cap	lock	1.96	1.90	0.82	Fit
6	flip	drum	flag	1.70	1.82	0.65	Fit
7	crab	crane	snake	0.54	-0.61	0.78	Fit
8	bread	spade	space	1.13	0.41	0.73	Fit
9	tray	clown	train	1.10	0.43	0.69	Fit
10	stop	star	plane	0.68	-0.14	0.85	Fit
11	point	fast	pond	0.93	0.12	0.61	Fit
12	bank	band	toast	0.49	-1.01	0.69	Fit
13	desk	lamp	lamb	0.37	-1.32	0.73	Fit
14	wand	mask	world	0.46	-1.13	0.64	Fit
15	cast	cost	jump	0.44	-1.12	0.64	Fit
							15/15 items = Fit
							0/15 items = Misfit

Note. MNSQ = mean-square statistic; ZSTD = ZSTD statistic; 'Fit' or 'Misfit' indicates whether the data fit the properties of the Rasch Model; grey shading = correct response to test item.

At the start of the school year, it was expected that a number of test items may prove challenging for five-year-old children. This would allow for the identification of high-ability students, progress monitoring throughout the school year, and the creation of an adaptive version of the test in future studies. Hence Table 6.3 illustrates the mean-square statistic, ZSTD statistic and point-measure correlation for items in the phoneme segmentation task at the start of the school year. Phoneme segmentation was considered the hardest task in the computer-based screening and monitoring tool and, according to the literature, it is challenging for five-year-old children (Adams, 1990; Chard & Dickson, 1999). Therefore, it was expected that a number of phoneme segmentation items would demonstrate a significant model 'misfit' at the start of the school year. Rasch analysis supported this assumption.

Table 6.3

'Fit' or 'misfit' for phoneme-segmentation test items at school-entry

Test Items		Outfit Statistics			Interpretation
How many sounds in the word …?		MNSQ	ZSTD	PT-measure correlation	'Fit' or 'Misfit'
1	moon (3)	0.63	-0.41	0.79	Fit
2	tooth (3)	0.56	-0.41	0.87	Fit
3	cow (2)	2.08	-2.82	0.29	Misfit
4	cup (3)	0.08	-1.41	0.45	Fit
5	soap (3)	Winsteps could not analyse items 4 to 18 due to lack of variance in the dataset (i.e., the majority of responses were incorrect)			Misfit
6	saw (2)				Misfit
7	flush (4)				Misfit
8	crab (4)				Misfit
9	sew (2)				Misfit
10	step (4)				Misfit
11	plate (4)				Misfit
12	star (3)				Misfit
13	bank (4)				Misfit
14	lock (3)				Misfit
15	jump (4)				Misfit
16	pond (4)				Misfit
17	bear (2)				Misfit
18	cast (4)				Misfit
					3/18 items = Fit 15/18 items = Misfit

Note. (number) = the number in brackets indicates the number of phonemes in the target word according to New Zealand English; pink shading = items demonstrating a significant model 'misfit'.

Table 6.3 illustrates that items 1, 2 and 4 show a model 'fit', while items 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 and 18 demonstrate a significant model 'misfit'. Items 1 ('moon'), 2 ('tooth') and 4 ('cup') were simple CVC words containing either long medial vowels, earlier developing sounds (e.g., /k/, /p/, /m/), or both. This may have increased the salience of the phonemes in these items, allowing five-year-old children to segment them correctly. Item 3 ('cow') was thought to be easy during the construction of the test. However, over 85 per cent of children indicated that this item had three sounds, as opposed to two. From video observations, it was revealed that children often added a final schwa phoneme when segmenting 'cow' verbally (i.e., /kaʊ**ə/**). Only nine per cent of children scored correctly on this item. The majority of phoneme segmentation items demonstrating a significant 'misfit' had a CCVC or CVCC syllable structure. Less than 10 per cent of participants provided a correct response to phoneme segmentation items 5 to 18, indicating that these items showed a 'misfit' because they were extremely difficult.

Table 6.4 summarises which test items demonstrated a 'fit' or significant model 'misfit' at the start, middle and end of the first year at school. This table illustrates that the majority of items 'fit' at all three points in the school year, while a small number of items demonstrated a 'fit' at only one or two assessment points.

Table 6.4

Summary of items by task demonstrating a 'fit' or significant model 'misfit' at the start, middle and end of the school year

PA Task	Start of Year (Age: 5;0 – 5;2)		Middle of Year (Age: 5;5 – 5;7)		End of Year (Age: 5;11 – 6;01)	
	Fit	Significant Misfit	Fit	Significant Misfit	Fit	Significant Misfit
Rhyme Oddity	*3,4,5,6,7,8,9,10	1 (easy), 2	3, 4, 5, 6, 7, 8, 9, 10	1 (easy), 2	3, 5, 6, 7, 8, 9, 10	1(easy), 2, 4 (easy)
Initial Phoneme Identity	1, 2, 3, 4, 5, 6, 7,8 ,9, 10		1, 2, 3, 45, 6, 7, 8, 9, 10		1, 2, 3, 4, 5, 6, 7, 8, 9, 10	
Final Phoneme Identity	1, 3, 4, 5, 6, 7, 8, 9, 10	2 (hard)	1, 2, 3, 4, 5, 6, 7, 8, 9, 10		1, 2, 3, 5, 6, 7, 8, 9, 10	4 (easy)
Phoneme Blending	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15		1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15		1, 2,3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15	
Phoneme Deletion	2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15	1 (easy) 6	2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15	1 (easy)	2, 3, 4, 6, 7, 8, 9, 10, 12, 13, 14, 15	1 (easy), 5 (easy), 11
Phoneme	1, 2, 4	3, 5, 6, 7, 8, 9,	2, 3, 4, 5, 6,	1 (easy),	2, 3, 4, 5, 6, 7, 8, 9,	1 (easy),

Segmentation		10, 11, 12, 13, 14, 15, 16, 17, 18 (these items were all hard)	7, 8, 9, 10, 11, 12, 13, 14, 15	16 (hard), 17 (hard), 18 (hard)	10, 11, 12, 13, 14, 15	16 (hard), 17 (hard), 18 (hard)
Letter-name	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18		1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18		1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18	
Letter-sound	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18		1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18		1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18	

Note. * numerals in the table indicate the item ‘number’ demonstrating a ‘fit’ or significant ‘misfit’; (easy) = indicates an item on which the majority of participants responded correctly, suggesting the ‘misfit’ occurred because the item was too easy; (hard) = indicates an item on which the majority of participants responded incorrectly, suggesting the ‘misfit’ occurred because the item was too hard; items showing a ‘misfit’ without an ‘easy’ or ‘hard’ description, show a ‘misfit’ for a reason besides difficulty level (see Appendix C).

Table 6.4 demonstrates that the majority of test items (i.e., 108 out of 114) show a model 'fit' at one or multiple points during the first year at school. All items demonstrating a model 'fit' also had point-measure correlations above 0.3, indicating that they discriminate well between individuals of high- and low-ability. At the start of the school year, 15 out of 18 phoneme segmentation items demonstrated a significant model 'misfit'. This suggests that phoneme segmentation items, as part of the computer-based screening and monitoring tool, are less suitable for measuring the PA ability of five-year-old children at school-entry. However, by the middle and end of the school year, 14 out of 18 phoneme segmentation items demonstrated a 'fit', indicating that these items become increasingly appropriate measures as children begin to interact with beginning classroom literacy instruction. This is consistent with research findings showing that phoneme segmentation is a difficult task for children five years of age (Adams, 1990; Chard & Dickson, 1999). Importantly, the CBA contained test items that ranged in suitability and difficulty throughout the school year. Having items that are above the ability level of five-year-old children is an important feature of the CBA because it promotes differentiation between high- and low-ability students and provides scope for the development of an adaptive form of the CBA in future investigations. This is because an adaptive CBA requires a range of test items to choose from so that it can accurately adapt the presentation of test items based on a child's responses to previous test items.

Of the 114 test items, six items demonstrated a significant model 'misfit' at all three assessment points in the school year. These items were rhyme oddity items 1 and 2, phoneme deletion item 1 and phoneme segmentation items 16, 17 and 18. Other items demonstrated a 'misfit' at only one or two assessment points in the school year. These include rhyme oddity item 4, final phoneme identity items 2 and 4, phoneme deletion items 6, 5 and 11, and phoneme segmentation item 1. Detailed explanation on why these items demonstrated a significant model ‘misfit’ is provided in Appendix C. Adapting these ‘misfit’ items in terms of linguistic complexity (i.e., word familiarity, syllable structure and manner of articulation) and presentation (i.e., animated and static graphics or verbal instructions) at the point at which the ‘misfit’ occurred will be required in future investigations.

However, it is important to note that some items demonstrating a significant ‘misfit’ may not necessarily require adaption. This is because their low level of difficulty is a purposeful part of the test construction process, to help ensure graded levels of difficulty within tasks. For example, to make it easier, rhyme oddity item 1 was developed using a simple syllable structure, a high-frequency rime unit and salient contrasts in the manner of articulation between correct and distractor options. This was done to ensure children’s success on what would usually be one of the first tasks administered as part of the computer-based PA screening and monitoring tool.

In summary, the majority of rhyme oddity, initial phoneme identity, final phoneme identity, phoneme blending, phoneme deletion, letter-name and letter-sound test items demonstrated a ‘fit’ throughout the first year of schooling. This suggests that these items are appropriate for measuring the PA ability of five-year-old children. Although phoneme segmentation items were less appropriate as measures at school-entry, they became more suitable by the middle and end of the school year. This feature of the CBA is important for the development of an adaptive form of the assessment in the future. The next step in item analysis was to determine whether test items sampled a range of difficulty levels. This is necessary to allow teachers to differentiate between high- and low-ability students in the classroom.

6.3.1.1.2 Identifying a hierarchy of item difficulty.

Rasch analysis enables comparison of item difficulty through computation of 'estimates of item difficulty' (Bond & Fox, 2001). *Winsteps* refers to 'estimate of item difficulty' as the 'measure' statistic, which in essence is a logit (log-odds) score assigned to an item to indicate its difficulty (Linacre, 2010). A logit score is plotted along an interval scale called a logit scale. The logit value of zero represents an arbitrary mean. Therefore, items with a logit score near zero are considered to be of average difficulty. Items with increasingly positive logit scores are more difficult, while items with increasingly negative logit scores are easier. In theory, a logit scale can range from negative infinity to positive infinity (Bond & Fox, 2007). Therefore, for the purposes of this thesis, the following difficulty descriptions were applied to logit values:

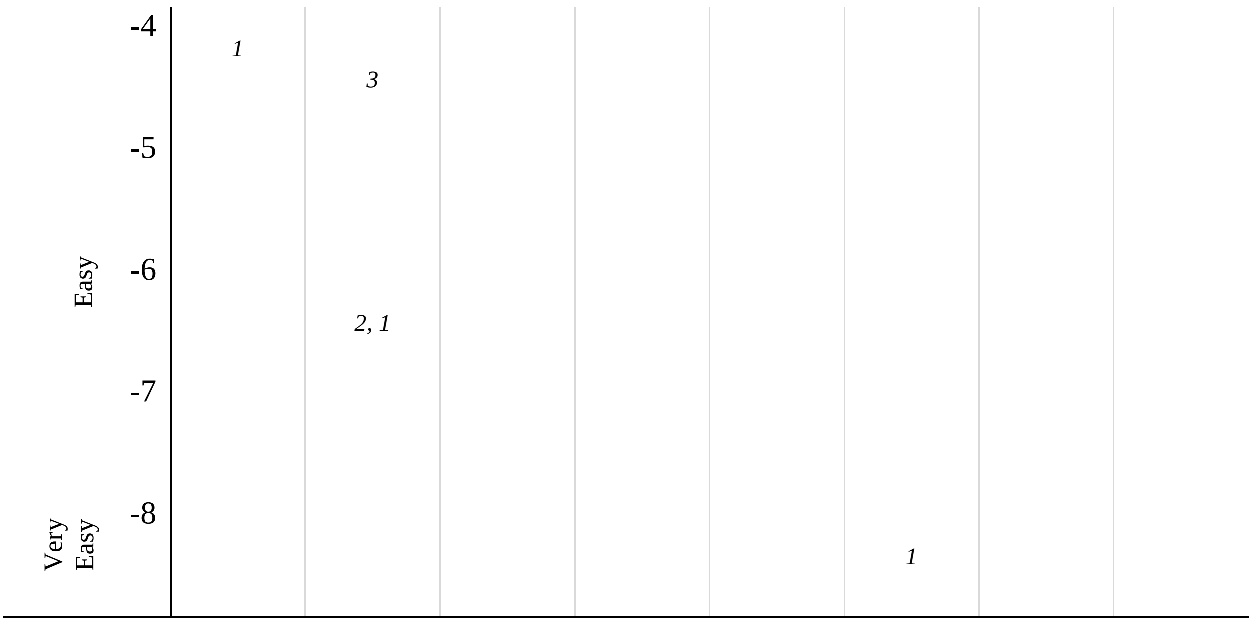
- 8 and above = very difficult
- 5 to 7 = difficult
- 2 to 4 = moderately difficult
- 1 to -1 = average difficulty
- -2 to -4 = moderately easy
- -5 to -7 = easy
- -8 and below = very easy

Table 6.5 summarises the hierarchy of difficulty for items at the start of the school year by plotting the 'measure' statistic (i.e., logit score) for each item against a logit scale. Exact 'measure' statistics (and error margins) for test items are provided in Appendix D.

Table 6.5

Hierarchy of item difficulty at school-entry

Logit Scale		RO	IPI	FPI	PB	PD	PS	LN	LS
Very difficult	8								
							3		
Difficult	7								
			6	*2*					
	6						*4*		
			9	*4, 5, 6, 7, 8, 9, 10*					
	5								
					6, 7, 8, 9, 10, 11, 12, 13, 14				
Moderately Difficult	4					*7, 8, 9, 10, 11, 12, 13, 14, 15*			
		10, 9	*10*		*15*				
	3	*7*							
	2		*8*						*13, 15, 16, 17, 18*
		3			*4, 5*	*4*			
Average Difficulty	1			*3*		*3*			*7, 8, 11, 12, 14*
		8	*7*						
	0					*5*			*5, 6, 9, 10*
		2	*5*						
	-1		*4*			*6*		*13, 14, 15, 16, 7, 18*	*1, 2, 3, 4*
		4			*2, 3*	*2*			
Moderately Easy	-2			*1*				*7, 8, 9, 19, 11, 12*	
		6			*1*	*1*	*2*		
	-3							*1, 2, 3, 4, 5, 6*	
		5							



Note. Items are plotted against the logit scale to the nearest whole number. Exact logit values (i.e. 'measure' statistic) for test items are provided in Appendix D. RO = rhyme oddity; IPI = initial phoneme identity; FPI = final phoneme identity; PB = phoneme blending; PD = phoneme deletion; PS = phoneme segmentation; LN = letter-name; LS = letter-sound.

Table 6.5 illustrates that test items in the computer-based screening and monitoring tool sample a wide range of difficulty levels at school-entry. Rhyme oddity and initial phoneme identity items provide an even spectrum of easier items, to those that are more difficult. For example, rhyme oddity items ranged from moderately easy to moderately difficult (e.g., three items were moderately difficult, four items were of average difficulty and three items were moderately easy). Initial phoneme identity items ranged from easy to difficult (e.g., two items were difficult, one item was moderately difficult, four items were of average difficulty and three items were easy). Final phoneme identity, phoneme blending and phoneme deletion items predominantly sampled the moderately difficult to difficult range. For example, eight out of 10 final phoneme identity items were classified as difficult, while nine out of 15 phoneme blending and 10 out of 15 phoneme deletion items were considered moderately difficult.

Only four out of 18 phoneme segmentation items could be analysed at school-entry; 14 items were extremely difficult and could not be analysed because the majority of respondents scored incorrectly, leading to lack of variance in the dataset. Of the four phoneme segmentation items analysed, one item was very difficult, one item was difficult, one item was moderately easy and one item was very easy. The table demonstrates a large gap between easier and more difficult phoneme segmentation items, suggesting that these items do not adequately sample a range of difficult levels at this stage of schooling. The majority of letter-name items were moderately easy, whereas letter-sound items tended to be of average difficulty.

Table 6.6 illustrates the hierarchy of item difficulty for test items by the middle of the first year at school. In comparison to Table 6.5, items in the final phoneme identity, phoneme blending, phoneme deletion and phoneme segmentation tasks begin to sample a wider range of difficulty levels. This is probably because tasks and items that were more difficult at school-entry became easier as children's PA skills developed.

Table 6.6

Hierarchy of item difficulty by the middle of the school year

	Logit Scale	RO	IPI	FPI	PB	PD	PS	LN	LS
Very Difficult	8						*16, 17, 18*		
	7								
Difficult	6								
						15	*15*		
	5					*14*	*14*		
						9	*9*		
	4					*11, 12*	*11, 12*		
		9, 10							
ModeratelyDifficult					*14, 15*				
	3				*12, 6*	*13*	*13*		
		7	*6, 9, 10*	*10,5,6*	*11,13*	*10*	*10*		
	2			*7*	*9*				
		3		*9*	*7, 8*	*8*	*8*		
	1	*8*							*16, 17, 18*
				4, 8		*7*	*7*		
Average Difficulty	0		*4, 7, 8*		*10, 5*				*15*
		2	*3*						
	-1								*9, 12, 13*
						6	*6*		
		4	*5*						
	-2					*4*	*4*	*17, 18*	*14*
Moderately Easy						*3*	*3*		
		6		*3*					

Moderately Easy	-3		*2*		*4*			*14, 15*	*8, 10, 11*
			1	*2*					
	-4	*5*			*3*			*9, 10, 11, 13*	*1, 2, 3, 4, 5, 6, 7*
				1		*5*	*5*		
	-5	*1*						*12*	
Easy	-6							*1, 2, 3, 4, 5, 6, 7, 8*	
					2				
	-7								*16, 17, 18*
					1				
Very Easy	-8								
	-9								
						1, 2	*1, 2*		

Note. Items are plotted against the logit scale to the nearest whole number. Exact logit values (i.e. 'measure' statistic) for test items are provided in Appendix D. RO = rhyme oddity; IPI = initial phoneme identity; FPI = final phoneme identity; PB = phoneme blending; PD = phoneme deletion; PS = phoneme segmentation; LN = letter-name; LS = letter-sound.

Table 6.6 illustrates that rhyme oddity and initial phoneme identity continue to sample a range of easy to moderately difficult ability levels with initial phoneme identity items being less spread than at school-entry. While the majority of final phoneme identity, phoneme blending, phoneme deletion and phoneme segmentation items continue to be of greater difficult, they begin to sample the moderately easy to easy range. For example, four out of 15 phoneme-blending items, five out of 15 phoneme deletion items and five out of 18 phoneme segmentation items are either moderately easy or easy by the middle of the school year. This is in comparison to one out of 15 phoneme blending items, one out of 15 phoneme deletion items and two out of 18 phoneme segmentation items being classified as moderately easy or easy at school-entry. Letter-name and letter-sound test items became easier to complete by the middle of the school year.

Table 6.7 profiles the hierarchy of item difficulty by the end of the school year. By this stage, items are becoming increasingly easier for children to complete as they approach six years of age. Items are more evenly spread across high to low logit scores, particularly for items in the final phoneme identity, phoneme deletion and phoneme segmentation tasks. Interestingly, phoneme deletion item 11, which was moderately difficult at the start and middle of the school year, becomes very difficult by the end of the school year.

Table 6.7

Hierarchy of item difficulty by the end of the school year

	Logit Scale	RO	IPI	FPI	PB	PD	PS	LN	LS
Very Difficult	9					*11*	*16, 17, 18*		
	8								
						15			
	7								
Difficult	6	*10*							
				9		*14*			
				10					
	5					*9*			
						13			
Moderately Difficult	4	*9*							
					14		*15, 14*		
						12			
					15				
	3				*6, 11, 12,*		*11, 6, 12*		
				6, 5	*13*	*10*	*13*		
	2	*8*	*10*	*5*			*9*		
			9	*8*	*9,8, 7*		*8*		
							7		
Average Difficulty	1		*7*	*7*					
		7	*6, 8*				*10*		
					10				
	0								
			5		*5*	*8*			
	-1	*5*	*4*			*7*			
	-2		*3*						
						6			

Moderately Easy	-3		*2*	*3*	*4*				*16,18*
							4		
	-4	*3*		*2*	*3*	*4, 5*			
			1	*1*			*3*		
	-5								*17*
Easy	-6	*2*			*2*			*16, 17, 18*	*12, 13, 14, 15*
						3	*2*		
	-7							*10,11, 12, 13, 14, 15*	*9, 10, 11*
						2	*5*		
					1				
Very Easy	-8	*6*						*4, 5, 6, 7,8, 9*	*2, 3, 4, 5, 6, 7, 8*
							1		
	-9							*1, 2*	*1*
		1, 4		*4*					
						1			

Note. Items are plotted against the logit scale to the nearest whole number. Exact logit values (i.e. 'measure' statistic) for test items are provided in Appendix D. RO = rhyme oddity; IPI = initial phoneme identity; FPI = final phoneme identity; PB = phoneme blending; PD = phoneme deletion; PS = phoneme segmentation; LN = letter-name; LS = letter-sound.

In summary, test items in the computer-based PA screening and monitoring tool sample a range of difficulty levels throughout the first year of school and thus provides evidence of content validity. At school-entry, rhyme oddity and initial phoneme identity items appear to sample an even spectrum of easier to more difficult ranges, whereas final phoneme identity, phoneme blending, phoneme deletion and phoneme segmentation items were more challenging. By the middle and the end of the school year, items in each of the six PA tasks sample the full spectrum of difficulty levels. This is essential for helping teachers differentiate between high- and low-ability students in the classroom. In addition, letter-name and letter-sound items became easier throughout the school year and generated increasingly lower point-measure correlations. This suggests that the ability of letter-name and letter-sound items to discriminate between high- and low-ability students reduces over time. Tables in Appendix D provide the exact 'measure' static and associated error margin for each test item at the start, middle and end of the school year. Table 6.8 provides an example of one of these tables by profiling the 'measure' statistic and error margin for rhyme oddity test items at school-entry. In addition, Appendix B provides a table illustrating the mean performance on each task in the CBA during the first year at school.

Table 6.8

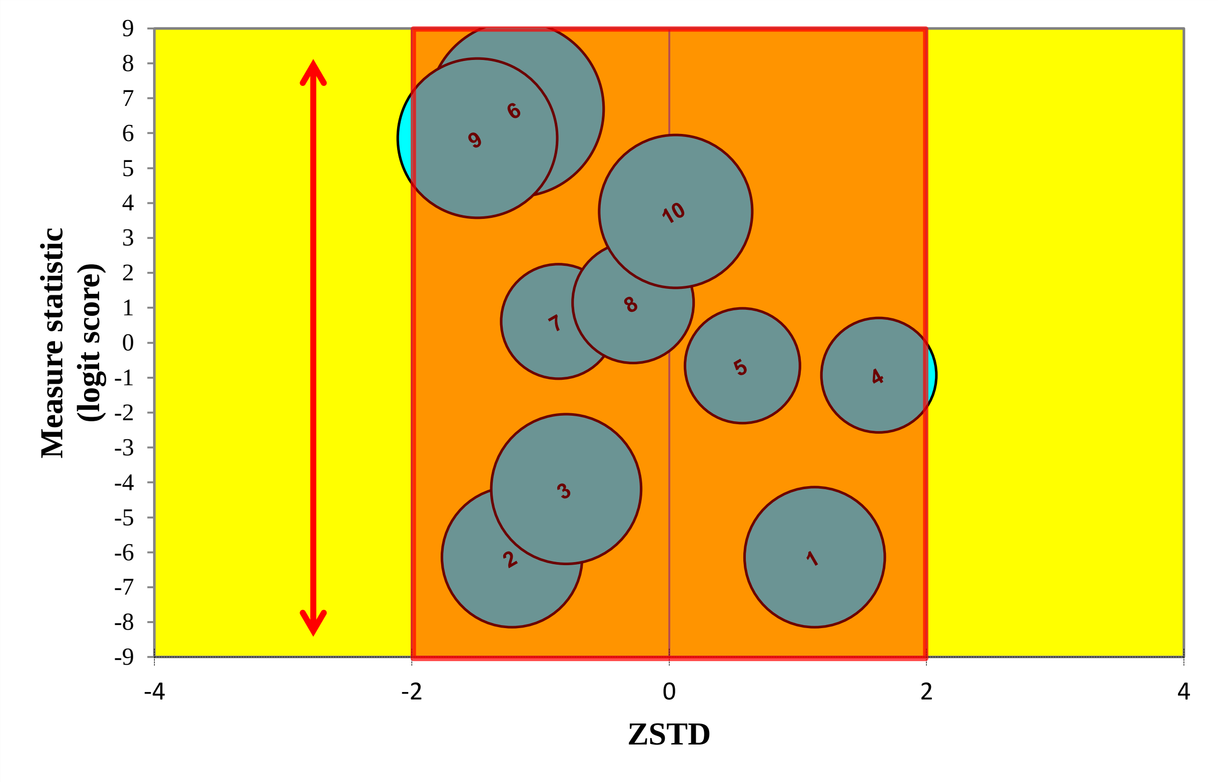
Difficulty of rhyme oddity items at school-entry using the 'measure' statistic (logit score)

Test Items				Outfit Statistics		
Which word does not rhyme?				MEASURE	ERROR	Interpretation
10	ring	sing	lamb	3.72	0.33	More Difficult
9	shell	duck	bell	3.67	0.32	
7	bun	sun	kite	2.96	0.28	↑
3	saw	toe	bow	1.77	0.25	
8	tent	lock	sock	0.89	0.25	
2	peg	doll	leg	-0.68	0.27	
4	sand	hand	cup	-1.47	0.29	↓
6	dog	book	hook	-2.81	0.31	
5	hen	car	pen	-3.96	0.38	
1	cat	mat	bus	-4.11	0.39	Less Difficult

Note. Grey shading indicates the correct response to test items.

6.3.1.1.3 Combining item 'fit' and item difficulty within phonological awareness tasks.

Bubble charts can be used to compare item 'fit' and item difficulty within each PA task. Plotting this information allows for a quick interpretation as to which items fit appropriately for five-year-old children and whether items within a task sample a range of difficulty levels (Linacre, 2010). On a bubble chart, the horizontal axis represents the ZSTD statistic, where items within the range of -2 to +2 demonstrate a model 'fit' and items outside this range demonstrate a significant model 'misfit'. The vertical axis represents item difficulty in logits. Figure 6.2 illustrates a bubble chart for initial phoneme identity items at school-entry. This figure demonstrates that all ten initial phoneme identity items showed a significant model 'fit' and sampled a range of ability levels from easy to more difficult. Bubble charts for each PA task at the start, middle and end of the school year are provided in Appendix E.



Note. Orange shading = 'fit' zone between -2 and +2 ZSTD. Red arrow = demonstrates the range of difficulty from easy to more difficult items.

Figure 6.2. Initial phoneme identity items by 'fit' and hierarchy of difficulty

6.3.1.2 Supporting Rasch analysis using Classical Test Theory.

Results from Rasch analysis were supported by a secondary item analysis using Classical Test Theory (CTT). In CTT, items are analysed in relation to the total test or task score. In contrast, the Rasch analysis evaluates items in relation to participant ability irrespective of total test or task score (Bond & Fox, 2001). Research suggests that while the Rasch Model has a stronger theoretical and mathematical basis than CTT, these two psychometric frameworks generally produce comparable results (Bond & Fox, 2007). In this thesis, items identified under CTT as being too easy, too difficult or not discriminating well between high- and low-ability students were also identified under Rasch analysis as having a mean square above 2.0, a ZSTD outside -2 to +2 and a point-measure correlation below 0.3.

Rasch analysis also identified four additional items, not identified by CTT, as demonstrating a significant ‘misfit’ during the first year of formal schooling. These included phoneme segmentation items 9 and 12 at school-entry, phoneme deletion item 1 at school-entry and phoneme deletion item 11 at the end of the school year. Formal item review using Rasch analysis, and validated using CTT, allowed appropriate and flawed items to be identified as part of the test construction process.

6.3.1.3 Chance levels.

It is important to examine how the use of a multiple-choice response format influences task performance (Thorndike & Thorndike-Christ, 2010). Test items in the rhyme oddity, initial phoneme identity, final phoneme identity, phoneme blending and phoneme deletion tasks had a multiple-choice response format comprising of three options: one correct option and two distractor options. Phoneme segmentation also involved a multiple-choice response format in which children clicked a box for each sound they heard in the target word. Five boxes could be selected, giving a total of five options: one box, two boxes, three boxes, four boxes or five boxes. Letter-name and letter-sound tasks required children to select one letter from a choice of six letter options. The use of a multiple-choice response format means that children’s performance on a task can be influenced by guessing behaviour. For example, in the initial phoneme identity task, children can score four out of 10 items correctly by selecting the first picture in the row for each item.

To evaluate the degree to which the multiple-choice format influenced task performance, a criterion chance level was applied and the percentage of children scoring above that chance level was calculated using a binomial distribution. The probability of a child scoring a certain number of test items correctly by chance was set at a criterion of $p < .05$. For tasks containing 10 test items with a choice of three options per test item (i.e., rhyme oddity, initial phoneme identity, final phoneme identity), children needed to score seven or more items correct to perform above chance level. For tasks containing 15 test items with a choice of three options per test item (i.e., phoneme blending, phoneme deletion), children needed to score nine or more items correct to perform above chance level. For the phoneme segmentation task that contained 18 items with a choice of five options per test item, children needed to score eight or more items correct to perform above chance level. Finally, the letter-name and letter-sound tasks contained 18 items with a choice of six letter options per test item, meaning that children needed to score six or more items correct to perform above chance level.

Table 6.9 profiles the percentage of participants scoring above a chance level of $p < .05$ on each PA task at the start, middle and end of the first year at school. As the table illustrates, the percentage of children performing above the criterion chance level increased over time.

Table 6.9

The percentage of participants scoring above chance level on tasks with multiple-choice test items at the start, middle and end of the school year

	School-Entry (n = 95)	Middle of Year (n = 95)	End of Year (n = 95)
Rhyme oddity	28.87%	62.5%	74.74%
Initial phoneme identity	34%	73.96%	81.05%
Final phoneme identity	0%	33.33%	60%
Phoneme blending	0%	57.29%	78.95%
Phoneme deletion	0%	39.58%	60%
Phoneme segmentation	0%	40.63%	62.12%
Letter-name	92.6%	100%	100%
Letter-sound	84.21%	94.74%	98.95%

Table 6.9 provides support for item analysis using the Rasch Model in that rhyme oddity and initial phoneme identity appear to be the most appropriate tasks to administer at school-entry. This is because 28.87 per cent and 34 per cent of participants scored above chance level on rhyme oddity and initial phoneme identity, respectively, at school-entry. In contrast, zero per cent of participants scored above chance on final phoneme identity, phoneme blending, phoneme deletion and phoneme segmentation tasks at school-entry. Using Rasch analysis, rhyme oddity and initial phoneme identity items were both appropriate (i.e., show a 'fit') at school-entry and sampled a range of difficulty levels, whereas the majority of final phoneme identity, phoneme blending and phoneme deletion items were appropriate, but tended to be of greater difficulty. Phoneme segmentation items at school-entry were largely inappropriate and extremely difficult. Confirming Rasch analysis, an increasing proportion of children scored above chance level on final phoneme identity, phoneme blending, phoneme deletion and phoneme segmentation tasks by the middle and end of the school year. These results indicate that the tasks became increasingly appropriate measures to monitor development and response to classroom instruction towards the middle and end of the school year (see Chapters 4 and 5). The majority of children scored above chance level on letter-name and letter-sound items at the start, middle and end of the school year. Collectively, these results provide strong evidence of content validity by demonstrating that test items were largely appropriate for five-year-old children, adequately sampled a range of difficulty levels, and differentiated between high- and low-ability children in the first year of formal education.

6.3.2 Construct validity.

In the literature, PA is reported to be a uni-dimensional construct (i.e., one latent trait) that generally follows a sequential trend of development (Ehri et al., 2001; Hoien et al., 1995). To evaluate whether the computer-based screening and monitoring tool accurately represents the construct of PA, three sources of validity were explored. These were: 1) factor analysis, to determine whether the six PA tasks measured a single underlying trait; 2) test characteristic curves, to examine how tasks related to each other in terms of difficulty; and 3) effect-size analysis between PA tasks throughout the school year, to determine sensitivity to growth in this construct.

6.3.2.1 Exploratory factor analysis.

Exploratory factor analysis using the principal component extraction method was conducted at the start, middle and end of the school year to evaluate whether the six PA tasks represented a uni-dimensional or multi-dimensional construct. A correlation matrix (*r*-matrix) between each of the six PA tasks at each assessment point was constructed, and eigenvalues were computed to determine factor loadings. Analysis using SPSS (Version 19.0) extracted one primary PA factor at the start, middle and end of the school year. This primary factor (i.e., first factor) accounted for 55 per cent, 73 per cent and 75 per cent of the total variance at the start, middle and end of the school year, respectively. A secondary factor only accounted for nine per cent, eight per cent and seven per cent of the total variance at the start, middle and end of the school year, respectively. According to Stevens (2002), the percentage of variance accounted for by the primary factor at each of the three assessment points is significant. Stevens (2002) advocates that only factors with a loading of 0.4, or that account for 16 per cent of the variance in a variable, require interpretation. The findings reported here provide evidence that the six PA tasks within the computer-based screening and monitoring tool measure the same underlying PA construct.

To determine how much each PA task contributed to the primary factor, component matrix tables produced by SPSS (Version 19.0) were used. These tables provide a component loading, which represents the correlation between each PA task and the primary factor. Table 6.10 aggregates the component matrix tables to profile component loadings for each PA task.

Table 6.10

Component loadings at the start, middle and end of the school year for phonological awareness tasks

	Loadings by Task on the Primary Factor		
Task	Start of Year	Middle of Year	End of Year
Rhyme Oddity	.84	.88	.85
Initial Phoneme Identity	.85	.78	.89
Final Phoneme Identity	.75	.86	.86
Phoneme Blending	.82	.90	.88
Phoneme Deletion	.73	.87	.87
Phoneme Segmentation	.74	.83	.88

At school-entry, initial phoneme identity, rhyme oddity and phoneme blending showed the highest loadings on the primary factor. By the middle and end of the school year, each of the six PA tasks showed a similar loading on the primary factor. These results suggest that the PA tasks in the computer-based tool measure the same underlying construct of PA.

According to the literature, PA and letter-knowledge are separate constructs, which both contribute to the prediction of reading difficulties (Ehri et al., 2001). Therefore, factor analysis using the principal component extraction method was also conducted for letter-name and letter-sound tasks. One primary factor was extracted and accounted for 95 per cent, 94 per cent and 96 per cent of the variance at the start, middle and end of the school year. Table 6.11 illustrates the component loading for letter-name and letter-sound knowledge on the primary factor throughout the school year.

Table 6.11

Component loadings at the start, middle and end of the school year for letter-name and letter-sound tasks

	Loadings by Task on the Primary Factor		
Task	Start of Year	Middle of Year	End of Year
Letter-name	.98	.97	.98
Letter-sound	.98	.97	.98

6.3.2.2 Test characteristic curves.

According to the literature, PA manifests itself in different skills that vary in complexity throughout the course of development (Adams, 1990; Anthony & Lonigan, 2004; Kamhi & Catts, 2012). To determine whether tasks within the computer-based screening and monitoring tool accurately address this aspect of PA, test characteristic curves (TCCs) were computed as part of Rasch analysis. TCCs are one method of comparing the difficulty of one task to another (Linacre, 2010). The TCC illustrates the relationship between the ability of participants and their expected probability of success. The ability being measured, often referred to as 'theta', is located on the *x*-axis, with positioning towards the left of the axis indicating less ability, and positioning towards the right of the axis representing more ability.

The *y*-axis represents the probability of success on a task. As ability increases along the *x*-axis, the probability of obtaining a higher task score also increases (Linacre, 2010). TCCs were calculated for the six PA tasks at the start, middle and end of the school year. TCCs were also computed for the letter-name and letter-sound tasks. Figure 6.3 illustrates the TCCs for the responses to each PA task at school-entry. Rhyme oddity is located furthest to the left on the *x*-axis, suggesting that this is the easiest PA task for children five years of age. This was followed by initial phoneme identity, phoneme blending, final phoneme identity, phoneme deletion and phoneme segmentation. The sequence of difficulty was identical for TCCs generated at the middle and end of the school year, and is consistent with findings from the Rasch analysis and calculations of chance levels. TCCs calculated for letter-name and letter-sound knowledge at the start, middle and end of the school year were consistently easier than the PA tasks.

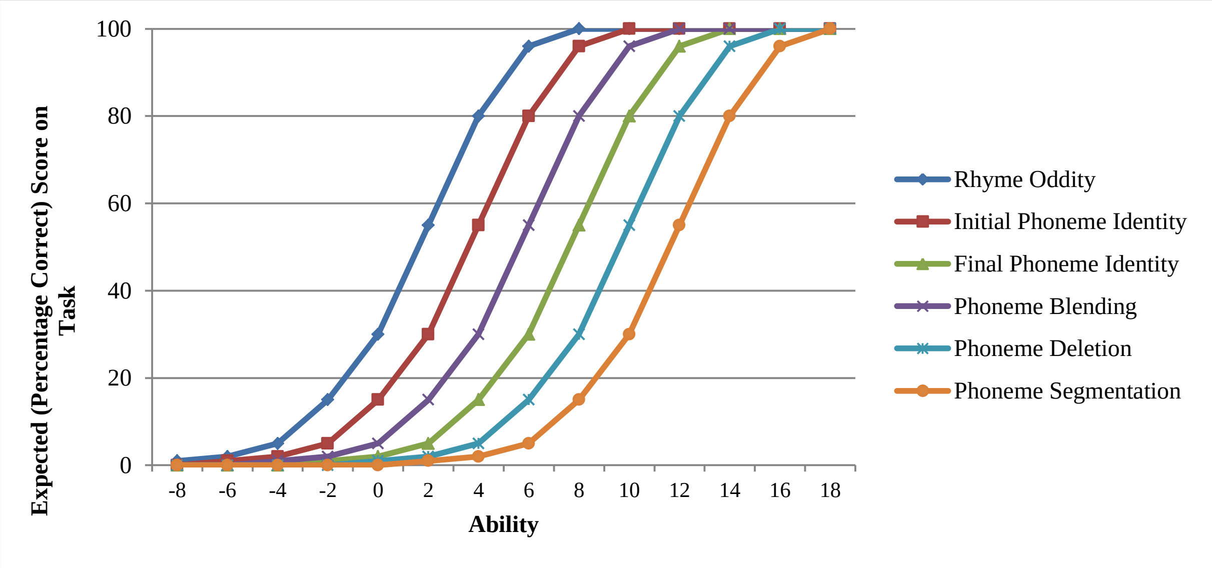


Figure 6.3. Test characteristic curves for six PA tasks at school-entry

6.3.2.3 Effect-size analysis to determine sensitivity to growth of phonological awareness as a construct.

Evidence suggests that growth in PA can account for the variance in later reading outcomes beyond that accounted for by the actual level of PA ability a child demonstrates (Byrne et al., 2000; Hindson, Byrne, Fielding-Barnsley, Newman, Hine, & Shankweiler, 2005). For this reason, effect sizes between each PA task from the start to the middle and the middle to the end of the school year were calculated. This was to help determine whether tasks comprising the computer-based tool were sensitive to growth in the PA construct being monitored. Paired t-tests between data obtained from the start to the middle and the middle to the end of the school year were conducted. This was followed by calculations of Cohen's *d*, to determine the strength of growth between tasks throughout the school year (Cohen, 1988). Results showed a significant difference in mean performance between the start and middle of the year in rhyme oddity ($t(95) = 7.99$, $p < .001$, $d = 1.63$), initial phoneme identity ($t(95) = 11.45$, $p < .001$, $d = 2.34$), final phoneme identity ($t(96) = 16.95$, $p < .001$, $d = 3.46$), phoneme blending ($t(95) = 16.45$, $p < .001$, $d = 3.36$), phoneme deletion ($t(95) = 15.05$, $p < .001$, $d = 3.07$) and phoneme segmentation ($t(95) = 14.13$, $p < .001$, $d = 2.88$). A significant difference in mean performance was also identified on all tasks from the middle to the end of the school year. For example, in rhyme oddity ($t(94) = 8.27$, $p < .001$, $d = 1.70$), initial phoneme identity ($t(94) = 6.35$, $p < .001$, $d = 1.30$), final phoneme identity ($t(94) = 12.34$, $p < .001$, $d = 2.53$), phoneme blending ($t(94) = 8.97$, $p < .001$, $d = 1.84$), phoneme deletion ($t(94) = 8.57$, $p < .001$, $d = 1.76$) and phoneme segmentation ($t(94) = 9.78$, $p < .001$, $d = 2.01$). According to Cohen (1988), an effect size of 0.80 is considered a large effect. The effect sizes calculated between mean performances on PA tasks from the start to the middle and middle to the end of the school year were all greater than 0.80, indicating that there was substantial growth in PA development during these periods. These results provide evidence that the tasks comprising the computer-based screening and monitoring tool are sensitive enough to detect growth in PA ability throughout the first year of formal schooling.

6.3.3 Criterion validity.

Determining how well the computer-based screening and monitoring tool can predict word-decoding abilities by the end of the first year at school is essential for establishing the validity of this instrument. Ensuring the instrument is aligned with existing valid and reliable paper-based measures of PA is also important for establishing validity. To achieve this, criterion validity was examined by considering concurrent and predictive validity evidence. (Thorndike & Thorndike-Christ, 2010). Predictive validity involved multiple regression analyses to determine the strength of the predictive relationship between PA and end-of-year word-decoding ability. Sensitivity and specificity calculations were also used (see Section

6.3.3.2). Concurrent validity involved correlating results on the computer-based PA tool with results on the standardised PIPA (Dodd et al., 2000).

6.3.3.1 Predictive validity using multiple regression analyses.

Multiple regression analyses were conducted to identify whether the PA tasks in the computer-based screening and monitoring tool demonstrated a strong predictive relationship with end-of-year word-decoding abilities. At the start of the school year, a number of variables were measured that were thought to contribute to the prediction of word-decoding skills and reading outcomes (see Chapter 4, Section 4.3.2). These predictor variables included PA, receptive language, expressive language, non-verbal intelligence, speech development (i.e., PCC), gender and socio-economic status. To meet the assumptions of a regression model, predictor variables must not be highly correlated (Field, 2009). Therefore, a PA factor score was calculated to represent the six PA tasks in the computer-based tool, and a letter-knowledge factor score was computed to represent letter-name and letter-sound tasks. Standard scores on a scale with 100 being the mean and $\pm$ 15 being the standard deviation were used for receptive language, expressive language and non-verbal intelligence, while PCC was used for speech development. The codes 0 and 1 were used to categorise gender (boy or girl, respectively). Similarly, socio-economic status was coded as 0 and 1 for low and high, respectively (i.e, decile rankings 1–5 = low and decile rankings 6–10 = high). End-of-year word-decoding ability was measured using the raw score from the NARA, administered at six years of age.

First, simple regression using the method of least squares was used to evaluate the individual relationship between each predictor variable and that of end-of-year decoding ability. This was to determine the significance of the contribution each variable made to end-of-year word decoding prior to its inclusion in the multiple regression analysis. Simple regression analyses revealed that PA (R^2 = 0.58; t = 11.14, p < .0001), letter-knowledge (R^2 =

0.45, t = 8.69, $p < .0001$), receptive language (R^2 = 0.61; t = 0.79, $p < .0001$), expressive language (R^2 = 0.59; t = 0.77, $p < .0001$), non-verbal intelligence (R^2 = 0.50; t = 9.62, $p <$ .0001), speech (R^2 = 0.36; t = 7.11, $p < .0001$), gender (R^2 = 0.109; t = 3.346, p = .001) and socio-economic status (R^2 = 0.07, t = 2.60, p = .01) were related to end-of-year word-decoding ability to a significant level ($p < .05$).

Multiple regression analyses were then performed, where PA, letter-knowledge, receptive language, expressive language, non-verbal intelligence, speech, gender and socio-economic status were the predictor variables and end-of-year word decoding was the independent variable. Using SPSS (Version 19.0), collinearity diagnostics revealed that receptive and expressive language demonstrated a high level of multicollinearity with both variables having a Variance Inflation Factor (VIF) greater than 10 (e.g., receptive language, VIF = 23.501; expressive language VIF = 17.37) (Myers, 1990). This is not surprising given that the receptive and expressive language scores were generated from the same standardised assessment—the CELF-P2 (Wiig et al., 2006). The core language index from the CELF-P2—which takes into account receptive and expressive standard scores—was therefore substituted into the multiple regression model as the predictor variable, ‘core language’. Resulting analysis demonstrated that PA, letter-knowledge, language, non-verbal intelligence, speech, gender and socio-economic status accounted for 71.6 per cent of the variance in end-of-year word-decoding ability (F = 29.829, $p < .001$). According to Field (2009), if a predictor has a significant effect on an outcome then its unstandardised coefficient should be significantly different to zero. Significance testing indicated that PA (t = 2.03, p = .04) and core language (t = 2.03, p = .04) produced a significant relationship with end-of-year word decoding. While considered to make a significant contribution to word decoding in isolation, letter-knowledge (t = 0.99, p = .33), non-verbal intelligence (t = 1.50, p = .14), speech (t = 0.86, p = .39), gender (t = 1.71, p = .09) and socio-economic status (t = -0.36, p = .72) did not make a unique contribution to word decoding when entered into a regression model with multiple predictors.

Finally, to produce the most efficient model possible, PA and language were the only two predictors entered into the multiple regression model. Results demonstrated that PA (t = 5.51, $p < .001$) and language (t = 2.95, p = .004) collectively accounted for 68.9 per cent of the variance in end-of-year word-decoding ability (F = 97.630, $p < .001$). To generalise the results of this regression model to the population, adjusted R^2 was calculated (Field, 2009). The R^2 value was 0.68, indicating that, when applied to the population from which the study sample was derived, we can infer that approximately 70 per cent of the variance observed in end-of-year decoding skills can be accounted for by PA and language capabilities at school-entry. This provides supporting evidence that the computer-based PA screening and monitoring tool used in this study is a strong predictor of end-of-year word-decoding ability and an important adjunct to the assessment of language ability.

6.3.3.2 Predictive validity using calculations of sensitivity and specificity.

In addition to identifying the predictive strength of the computer-based PA tool, measurement of test sensitivity and specificity were also completed. Sensitivity refers to how accurately a test can predict which children will demonstrate a particular difficulty (e.g., reading problems), while specificity refers to how accurately a test can predict which children will not demonstrate a particular difficulty (Spitalnic, 2004). The sensitivity and specificity of a test can be calculated using the Positive Prediction Value (PPV) (i.e., the proportion of positive test results that are true positives) and the Negative Prediction Value (NPV) (i.e., the proportion of negative test results that are true negatives) (Spitalnic, 2004).

At school-entry, children who performed below the 25th percentile in the computer-based rhyme oddity and initial phoneme identity tasks were considered to be at risk for end-of-year word-decoding difficulties. The 25th percentile for rhyme oddity and initial phoneme identity was less than three and six test items correct, respectively, for the sample of 95 children. Rhyme oddity and initial phoneme identity tasks were selected because at school-entry the majority of items in these tasks: 1) were appropriate for five-year-old children, 2) sampled a range of difficulty levels and 3) gave a greater likelihood that children would score above chance level. At school-entry, these two tasks were more reliable than were those for final phoneme identity, phoneme blending, phoneme deletion and phoneme segmentation (see Section 6.3.1). Further, in a study by Torgesen and colleagues (1994), it was identified that children who began the first grade with PA abilities below the 20th percentile performed, on average, three grade levels behind their peers in reading development by the fifth grade. Moreover, in New Zealand it is estimated that up to one in five children may struggle with the acquisition of early reading skills. Thus, the 25th percentile was selected as a conservative cut-off point.

From a sample of 95 children, 20 children performed below the 25th percentile on rhyme oddity and initial phoneme identity tasks at school-entry. By the end of the school year, 16 of these 20 children performed below an age-expected word-decoding level, while four of these 20 children performed at or above an age-expected level. This produced a PPV of 80 per cent (i.e., 16 children showed word-decoding difficulties out of the 20 children predicted to have word-decoding difficulties). This means that four children were over-identified as being at risk for reading difficulties. Of the 95 children in this study, 75 children performed above the 25th percentile on rhyme oddity and initial phoneme identity tasks at school-entry. Of these 75 children, 71 performed at or above an age-expected word-decoding level by the end of the school year, whereas four children performed below an age-expected level. This produced a NPV of 94.66 per cent (i.e., 71 children show age-appropriate word decoding out of the 75 children predicted to demonstrate age-appropriate word-decoding abilities). This means that four children were under-identified as being at risk for reading problems. At school-entry, the overall accuracy of the computer-based screening and monitoring tool in predicting end-of-year word-decoding ability was 91.58 per cent. That is, the end-of-year word-decoding performance for 92 per cent of children was accurately predicted using computer-based measures of school-entry PA ability.

According to the literature, one disadvantage of screening tools that measure a small range of skills is that they often have a high false-positive rate (Blachman et al., 2000). That is, they tend to over-identify the number of children at risk for reading difficulties. One way to reduce the rate of false positives is to monitor progress in PA regularly throughout the first year of schooling (Good et al., 2002). The computer-based PA screening and monitoring tool was designed for re-administration at the middle and end of the first year at school (see Chapter 2). By the middle of the school year, two of the four children who were over-identified as being at risk for reading problems scored above the 25th percentile on all six PA tasks. This meant that the number of children considered at risk for reading difficulties reduced from 20 children at school-entry to 18 children by the middle of the school year.

Combining school-entry and middle of the year PA performance produced a PPV of 88.89 per cent (i.e., 16 children exhibited end-of-year word-decoding difficulties out of 18 children predicted to show end-of-year word-decoding difficulties). By using the computer-based PA instrument as a monitoring tool, predicting the status of end-of-year word-decoding ability increased in accuracy from 92 to 94 per cent. Although there are no guidelines on what constitutes a satisfactory level of sensitivity and specificity for a screening and monitoring tool, researchers have suggested that indices above 0.80 are satisfactory (Spitalnic, 2004). Both measures of sensitivity and specificity in the CBA tool were greater than the acceptable level of 0.80.

It is important to note that while the computer-based screening and monitoring tool produced a high level of sensitivity and specificity, four children continued to be under-identified as being at risk for reading problems, even when the instrument was re-administered in the middle of the year. These children performed above the 25th percentile on measures of rhyme oddity and initial phoneme identity at school-entry and above the 25th percentile on all PA measures by mid-year. This may suggest that these children were experiencing difficulties in transferring PA knowledge to the process of learning to read.

Importantly, school-entry performance below the 25th percentile on rhyme oddity and initial phoneme identity tasks together predicted end-of-year word-decoding ability with much greater accuracy than either task in isolation. According to the literature, letter-knowledge is also a powerful predictor of early reading ability (Ehri et al., 2001). However, school-entry performance below the 25th percentile on letter-knowledge provided no additional information for the prediction of end-of-year word decoding, beyond that already provided by rhyme oddity and initial phoneme identity tasks. This outcome is consistent with results from multiple regression analyses.

6.3.4 Concurrent validity.

Concurrent validity between the PIPA and the computer-based PA assessment was demonstrated in the experiment reported in Chapter 3. Significant positive correlations were identified on tasks measuring rhyme oddity ($r(31) = 0.83$, $p < .001$), initial phoneme identity ($r(31) = 0.88$, $p < .001$) and the letter-sound knowledge ($r(31) = 0.89$, $p < .001$). These tasks were selected because they are common to both the PIPA and the computer-based PA tool. Figure 3.4 in Chapter 3 illustrates the correlation between the initial phoneme identity tasks in the PIPA and the CBA.

6.4 Reliability

Reliability provides an index as to how consistent an assessment is over repeated administrations (Crocker & Algina, 1986). Consistency over repeated administrations is important for ensuring that the measurement tool is accurate and that users can be confident that the resulting test score is precise (Thorndike & Thorndike-Christ, 2010). Unreliable assessments are ineffective for identifying a child's performance level, informing curriculum design and monitoring progress. Establishing the reliability of an instrument is therefore essential and can be achieved through measures of internal consistency and test-retest reliability (Field, 2009).

6.4.1 Internal consistency.

Internal consistency refers to the relationship between items in a test and the consistency of responses between those items (Field, 2009). Evidence of internal consistency indicates that the items in the test measure one construct. Internal consistency between items within each PA task at the start, middle and end of the first year at school were calculated using Cronbach's alpha. Cronbach's alpha scores above 0.7 indicate that the items within a task are internally consistent (Field, 2009). Table 6.12 profiles the Cronbach's alpha scores for each task throughout the school year.

Table 6.12

Cronbach's alpha scores by task at the start, middle and end of the school year

	Start of Year (5;0 – 5;2)	Middle of Year (5;5 – 5;7)	End of Year (5;1 – 6;01)
Rhyme oddity	.81	.85	.84
Initial phoneme identity	.89	.85	.85
Final phoneme identity	.14	.84	.89
Phoneme blending	.45	.94	.92
Phoneme deletion	.15	.94	.85
Phoneme segmentation	.47	.89	.89
Letter-name	.81	.82	.82
Letter-sound	.80	.81	.82

Note. Age is represented as years;months.

At school-entry, rhyme oddity and initial phoneme identity showed a high degree of internal consistency, with Cronbach's alpha scores of 0.81 and 0.89, respectively. In addition, the letter-name and letter-sound tasks achieved scores of 0.81 and 0.80, respectively. Unsatisfactory Cronbach's alpha scores were calculated for final phoneme identity, phoneme blending, phoneme deletion and phoneme segmentation at the start of the year. This is consistent with Rasch analysis findings, which indicated that these latter tasks are generally more difficult at school-entry and less reliable at this point in the school year. By the middle and end of the school year, high Cronbach's alpha scores were calculated for all six PA tasks and the two letter-knowledge tasks. This result suggests that the computer-based tool becomes increasingly reliable in measuring response to classroom reading instruction during the middle and end stages of the first year of schooling.

6.4.2 Test-retest reliability.

Test-retest reliability refers to the consistency of a test across repeated administrations under identical conditions (Thorndike & Thorndike-Christ, 2010). In this Chapter, test-retest reliability was conducted by correlating each task with itself at each of the three assessment points during the school year. The resulting correlation matrix, shown in Table 6.13, revealed significant correlations at $p < .01$ for each task.

Table 6.13

Correlations between each phonological awareness task at the start, middle and end of the school year

	2 RO	3 RO		2 PB	3 PB
1 RO	.74 **	.68 **	1 PB	.85 **	.61 **
2 RO		.84 **	2 PB		.72 **
	2 IPI	3 IPI		2 PD	2 PD
1 IPI	.78 **	.71 **	1 PD	.63 **	.44 **
2 IPI		.84 **	2 PD		.71 **
	2 FPI	3 FPI		2 PS	3 PS
1 FPI	.71 **	.67 **	1 PS	.71 **	.58 **
2 FPI		.91 **	2 PS		.79 **

Note. 1 = school-entry administration; 2 = middle-of-year administration; 3 = end-of-year administration; RO = rhyme oddity; IPI = initial phoneme identity; FPI = final phoneme identity; PB = phoneme blending; PD = phoneme deletion; PS = phoneme segmentation; ** = significant to $p < .0001$

Collectively, internal consistency and test-retest reliability provide strong evidence that the computer-based PA screening and monitoring tool is internally reliable and highly consistent over repeated administrations (Crocker & Algina, 1986). This means that teachers can confidently use the computer-based PA assessment knowing that the results they obtain are reliable. That is, regardless of when the assessment is administered (e.g., morning, afternoon, today or tomorrow) teachers can expect to obtain a consistent profile of the PA ability of five-year-old children in their classroom. This information can then be used to inform classroom reading instruction and to monitior progress throughout the first year of formal schooling.

6.5 Discussion

The study reported in this chapter examined the validity and reliability of the computer-based PA screening and monitoring tool (described in Chapter 2) for a group of children receiving the 'usual' classroom reading programme in the first year of school. The data from the 95 five-year-old children who participated as comparison Group C in Chapter 4 were extracted and subjected to a number of statistical analyses to establish the validity and reliability of the computer-based PA measure. Providing evidence of validity and reliability is essential for ensuring teachers can confidently use the computer-based tool as a time-efficient method to predict and monitor risk for reading difficulty in the classroom. Specifically, data was analysed to determine whether test items 1) are appropriate and sample a range of difficulty levels for five-year-old children, 2) measure the same underlying construct of PA, 3) are predictive of end-of-year word-decoding ability and 4) are consistent over repeated administrations throughout the first year of schooling. Several hypotheses were investigated, with results providing future direction for the development and implementation of the computer-based PA tool in the classroom.

6.5.1 Content validity.

Content validity refers to the examination of content within a test to ensure it accurately measures the target trait (Anastasi & Urbina, 1997). According to Lissitz and Samuelsen (2007), formal item analysis is an essential source of content validity because test items are the building block of a measurement tool. The first hypothesis stated that the majority of test items would demonstrate an appropriate 'fit' to the ability of five-year-old children in the first year of school and would sample a spectrum of difficulty levels. This hypothesis was supported by analysis of the data. Rasch Model analysis using *Winsteps 3.70* was the primary measurement model employed to analyse item 'fit' and hierarchy of item difficulty. Results showed that the majority of rhyme oddity (i.e., eight out of 10) and all initial phoneme identity items are appropriate for administering to children at school-entry and sample a spectrum of ability levels. Final phoneme identity, phoneme blending and phoneme deletion items, while appropriate at school-entry, were of greater difficulty.

By the middle and end of the school year, the majority of items within the six PA tasks were appropriate and sampled a spectrum of difficulty. Further, these items demonstrated good point-measure correlations, indicating that they discriminated well between high- and low-ability students. Letter-name and letter-sound items were also appropriate in the first year of school. These items became increasingly easier, but proved less able to differentiate high- from low-ability students over time. Six test items from a total of 114 did not demonstrate a ‘fit’ at any point during the first year of school and will require detailed investigation in future studies (see Appendix C for explanations regarding item ‘misfit’ and future directions).

Research suggests that there is a universal sequence of PA development, but that a quasi-parallel acquisition of skills also exists (Anthony et al., 2003; Anthony & Francis, 2005; Ziegler & Goswami, 2005). It is believed that an awareness of rhyme and early phoneme-level skills (e.g., phoneme identification) emerges between four and five years of age, with further development at the phoneme level (e.g., phoneme blending, deletion, segmentation) occurring in conjunction with the onset of formal reading instruction (Dodd & Gillon, 2001; Lonigan et al., 1998; Lonigan et al., 2008). Consistent with previous research, formal item review demonstrated that rhyme oddity and initial phoneme identity tasks were most appropriate for five-year-old children at the threshold of school-entry, and that more difficult phoneme-level skills, such as phoneme blending and phoneme segmentation, became more appropriate as children began to interact with beginning classroom reading instruction.

One focus of test development, as reported in Chapter 2, was to achieve a graded level of test item difficulty to ensure teachers could discrimination between high- and low-ability students. Item analysis demonstrated that as PA tasks became increasingly appropriate, the capability of those items to differentiate between high- and low-ability students increased, as evidenced through point-measure correlation statistics. These results provide strong evidence for the content validity of the computer-based instrument.

6.5.2 Construct validity.

Construct validity refers to the extent to which an instrument measures the trait that it intends to measure (Thorndike & Thorndike-Christ, 2010). The trait intended to be measured by the computer-based tool is firstly PA and secondly letter-knowledge. Thus, the second hypothesis stated that test items for the six PA tasks would measure a uni-dimensional construct of PA and would represent a continuum of difficulty over the first year of schooling. Data analysis supported this hypothesis. Exploratory factor analysis demonstrated that all six PA tasks loaded onto one primary factor at the start, middle and end of the first year of schooling, indicating that these tasks provide a measure of the same underlying trait; namely, PA. In a separate analysis, letter-name and letter-sound tasks also loaded onto one primary factor, suggesting that these two tasks provide a measure of the same underlying construct; namely, letter-knowledge ability.

Computation of test characteristic curves showed a graded level of difficulty between each PA task that remained consistent throughout the first year of school. Specifically, rhyme oddity appeared to be the easiest task to complete, followed by initial phoneme identity, phoneme blending, final phoneme identity, phoneme deletion and phoneme segmentation. In terms of letter-knowledge, letter-name recognition was consistently easier than letter-sound identification. Effect sizes between performances from the start to the middle and from the middle to the end of the school year were large, indicating that the computer-based PA tool is sensitive to growth in PA over the school year. Accumulatively, these results provide evidence of construct validity for the computer-based tool.

Definitions to conceptualise PA range from narrow to broad descriptions. Researchers aligned with a narrow definition propose that PA consists of separate abilities (e.g., tasks that require one cognitive operation versus two cognitive operations, or a rhyming factor and a phoneme factor) (Anthony & Lonigan, 2004; Stanovich, 1992). Researchers aligned with a broad definition propose that tasks used to measure PA represent the same single underlying trait (Anthony & Lonigan, 2004). The results reported in this chapter support a broad definition of PA because the six tasks in the computer-based tool measured the same underlying ability. A progression of difficulty, with rhyme oddity being considered the easiest task and phoneme segmentation being the hardest task, supports item analysis results, and is generally consistent with reports of PA task difficulty in the literature (Anthony & Francis, 2005; Schreiber, 2008). Importantly, sensitivity to growth in PA ensures the computer-based tool can be used in a monitoring capacity to predict risk for reading difficult throughout the first year of schooling.

6.5.3 Criterion validity.

Criterion validity refers to the relationship between an instrument and another variable and can take the form of predictive or concurrent validity (Thorndike & Thorndike-Christ, 2010). Predictive validity refers to the association between an instrument and another outcome in the future, while concurrent validity refers to the connection between one instrument and another at the same point in time (Justice et al., 2002). The third hypothesis in this chapter held that the computer-based PA tool would show a strong predictive relationship with end-of-year word-decoding ability, and be highly correlated with an existing standardised measure with established reliability and validity. This hypothesis was supported by data analyses. Multiple regression analyses showed that PA tasks in the computer-based measure and a core language measure collectively accounted for 68.9 per cent of the variance in end-of-year word-recognition ability. Consistent with previous research (MacDonald & Cornwall, 1995; Torgesen et al., 1994), PA proved to be a more powerful predictor of end-of-year reading success than measures of intelligence or socio-economic status. Further, PA was a more powerful predictor of reading outcomes than letter-knowledge, speech sound development or gender.

Calculations of sensitivity and specificity showed that the computer-based PA tool accurately predicted end-of-year reading success with 92 per cent accuracy at school-entry. This increased to 94 per cent accuracy when the computer-based tool was re-administered during the middle of the school year, at which time the number of children over-identified as being at risk for end-of-year reading problems reduced from four to two by aggregating school-entry and middle-of-year data. These results are consistent with previous research indicating that false-positive rates of instruments that measure only one or a small number of skills can be reduced through repeated administrations.

A measure of concurrent validity between the computer-based PA instrument and the PIPA is reported in Chapter 3. Collectively, measures of predictive and concurrent validity provide evidence for criterion validity of the computer-based PA screening and monitoring tool.

6.5.4 Reliability.

Reliability refers to the consistency of a measurement tool over repeated administrations (Thorndike & Thorndike-Christ, 2010). This is an essential component in ensuring teachers can confidently include the computer-based instrument developed in this thesis as part of their classroom practice. Accordingly, the final hypothesis predicted that tasks in the computer-based tool would show a high level of internal reliability and be stable across repeated administrations. Data analysis supported this hypothesis. Measures of internal consistency using Cronbach's alpha revealed high reliability coefficients for rhyme oddity and initial phoneme identity at school-entry. Further, high reliability coefficients for all six PA tasks during the middle and end phases of the school year were obtained. These results are consistent with item analysis using the Rasch Model, in that rhyme oddity and initial phoneme identity are most suitable at school-entry, with additional PA measures becoming more appropriate as children interact with formal reading instruction. These results are consistent with the literature on the developmental progression of PA task difficulty (Anthony & Francis, 2005; Schreiber, 2008).

Test-retest reliability between each PA task from the start to the middle and the middle to the end of the school year was also significant. This result, in conjunction with measures of internal consistency, provide strong evidence that classroom educators can expect the computer-based PA screening and monitoring tool to be consistent across repeated administrations throughout the first year of school.

6.5.5 Implications for classroom assessment practices

These results have important implications for classroom assessment practices in that teachers can be confident that the computer-based tool does measure PA ability, differentiates between high- and low-ability students, and is predictive of end-of-year reading performance. Specifically, the results reported in this chapter provide New Zealand educators with information that is based on the performances of New Zealand school-aged children. This is significant because there is very little data available on PA development in New Zealand children; New Zealand educators have previously relied on international data. As such, the information reported in this chapter, and in Appendices B, C, D and E, provides New Zealand teachers with a foundation to begin developing a comprehensive database of PA performance in New Zealand school-aged children. Findings reported here also have implications for the need for improved screening instruments under an RTI framework at an international level (Mastropieri & Scruggs, 2005; Scruggs & Mastropieri, 2002) in that use of computer-based technology to measure PA can serve as a time-efficient adjunct to existing instruments in the classroom.

6.5.6 Limitations and future directions.

In the context of test development, the computer-based screening and monitoring tool developed in this thesis can be seen as a pilot study before wide-scale investigation at national and international levels. According to Paul (2007), a minimum of 100 participants for each age group is considered satisfactory in test development. A sample size of 95 children from the same metropolitan city was used to establish the validity and reliability of the computer-based measure reported in this chapter. Future investigations should employ a larger sample size across multiple school-sites. Further, six items demonstrated a significant 'misfit' to the ability of five-year-old children during the first year of school. These items will require detailed examination and may require adaptation or deletion in future studies.

To establish the validity and reliability of the computer-based tool, a researcher sat alongside each participant as the participant independently administered the tasks. A limitation of this approach is that the teacher was not in control of the instrument in the classroom, meaning that results reported here may lack generalisability. Given that the validity and reliability of the computer-based tool has been established under the observation of research staff, future studies should focus on applying the CBA in classroom settings in which teachers are implementers and in which additional variables, such as professional development and ongoing support in test use, can be examined.

Given the limited sample size, the intention of this thesis is not to develop a standardised instrument, but rather to commence investigation into and dialogue on the use of computer-based screening and monitoring methods as a time-efficient way to predict and minimise risk for reading failure in the classroom. The mean scores per task and standard deviations given in Appendix B provide a foundation for future investigation, the purpose of which may be to develop normative scores, or to consider how test items can be developed further to create a computer-adaptive version of the instrument.

Future investigations into the use of the computer-based PA assessment in the classroom should ideally focus on developing an adaptive version of the assessment. Computer-adaptive tests require large test item banks so that the computer has a sufficient number of test items to choose from when adapting the test to meet the wide range of ability levels present in the classroom. The number of test items in the current investigation that demonstrate a 'fit' at some stage in the school year and range difficulty provide a strong starting point for the expansion of test items and construction of an adaptive form of the instrument. It is anticipated that an adaptive version of computer-based PA screening and monitoring tool would further enhance the time-efficiency and effectiveness of PA assessment in the classroom.

In summary, the results reported in this chapter, in addition to those reported in Chapters 3, 4 and 5 provide a strong foundation for further investigation into the use of CBA as a way to equip teachers with a time-efficient and effective method towards raising achievement and equalising positive literacy outcomes within the beginning classroom literacy programme.

Chapter 7: General Discussion

7.1 Introduction

The research reported in this thesis investigated methods to integrate phonological awareness (PA) assessment and instruction efficiently and effectively into the classroom environment. PA is widely recognised as a powerful precursor and prognostic marker for early reading success and has been researched extensively under controlled research conditions (Brady et al., 1994; Carroll & Snowling, 2004; Ehri et al., 2001; Gillon, 2000a, 2005; Gillon & McNeill, 2009). Understanding how the benefits of PA can be bridged into existing classroom curriculums will provide educators, researchers and policy makers with valuable information to support national and international initiatives focused on raising achievement and reducing inequalities in reading outcomes for identified subgroups of school-aged populations who are currently underachieving. To achieve this, experiments in this thesis integrated PA into classroom practices from two complementary perspectives: 1) investigating the use of a computer-based PA assessment and 2) examining the impact of a short and intensive teacher-implemented PA programme on raising reading outcomes in the first year of school. Specifically, three broad hypotheses were tested:

1) A computer-based assessment tool (CBA), if well developed, will be an efficient and effective method for monitoring PA development and predicting reading outcomes in the first year of formal education.
2) Teacher-implemented classroom-wide phonological awareness instruction that is time-efficient and focused at the phoneme level will significantly raise reading

achievement for children with and without spoken language difficulties in the first year of school.

3) Teacher-implemented classroom-wide phonological awareness instruction will have a positive effect on reducing the prevalence of reading difficulties and minimising inequality in reading outcomes when included as part of the beginning literacy curriculum.

To test these hypotheses, four experiments were conducted to: 1) compare computer- and paper-based PA assessment modalities (see Chapter 3), 2) examine the effect of classroom PA instruction on literacy outcomes (see Chapter 4), 3) investigate the influence of classroom PA instruction on reading outcomes for children with spoken language impairment (SLI) (see Chapter 5) and, 4) establish the validity and reliability of the computer-based PA assessment (see Chapter 6).

This chapter begins with a brief review of these four experiments followed by a discussion regarding how the research reported in this thesis addresses the aforementioned hypotheses. The final sections of this chapter outline implications for classroom reading practices, study limitations and directions for future research.

7.2 Summary of Thesis Experiments

7.2.1 Experiment 1: Comparison of computer- and paper-based assessment modalities.

The first experiment examined whether the computer-based PA screening and monitoring tool (see Chapter 3) was time-efficient to administer while maintaining congruency of scores with a paper-based version of the assessment. Thirty-three children participated in the study, 12 of whom presented with moderate-severe speech delay (MSD). A crossover research design was used whereby half the participants received the computer-based version of a PA assessment, followed two weeks later by a paper-based counterpart. The other half of participants received the same assessments, but in the reverse order of delivery. Results showed that: 1) CBA generated comparable scores to the paper-based counterpart for both children with typical development (TD) and children with MSD, 2) CBA took 31 per cent less time to administer, and 3) CBA scores showed strong evidence of test-retest reliability, in addition to concurrent validity with an existing paper-based PA measure. Having established the time-efficiency of the CBA and the congruency of scores with paper-based testing, the next step was to use the computer-based PA assessment to monitor response to reading instruction within everyday classroom environments.

7.2.2 Experiment 2: Classroom phonological awareness instruction and literacy outcomes.

This experiment investigated the impact of a short and intensive period of PA instruction implemented by classroom teachers as part of the beginning literacy programme. A quasi-experimental design was employed to measure the PA, reading and spelling development of 129 five-year-old children. Thirty-four children in two classrooms received 10 weeks of PA instruction from their teachers. Ninety-five children from 10 classrooms continued with their ‘usual’ reading programme, which included phonics instruction but did not target PA. Results showed that children who received PA instruction demonstrated superior reading and spelling outcomes compared to children who followed the ‘usual’ curriculum. Importantly, the number of children experiencing word decoding difficulties after one year of schooling declined from 26 per cent among children who followed the usual literacy curriculum to 6 per cent among children who received PA instruction. These results suggest that a short and intensive period of classroom PA instruction can help raise the literacy profiles of five-year-old children in the first year of school.

7.2.3 Experiment 3: Classroom phonological awareness instruction and spoken language impairment.

This experiment investigated the impact of classroom PA instruction on reading outcomes for children with SLI. Data from the experiment described in Section 7.2.2 was extracted and analysed to compare end-of-year literacy outcomes for children with SLI who received PA instruction (n = 7) in comparison to children with TD who received classroom PA instruction (n = 27) and children with SLI (n = 21) and TD (n = 74) who followed the 'usual' literacy programme. Children with SLI showed significant improvements in PA, reading and spelling development following classroom PA. However, in comparison to children with TD, they appeared more restricted in their ability to transfer enhanced PA skills to written language tasks. Importantly, children with SLI who received PA instruction performed significantly better than did children with SLI who followed the 'usual' literacy curriculum; and at a similar level to children with TD who followed the 'usual' programme. Children with TD who received PA instruction performed at a significantly higher level on end-of-year reading measures compared to all other cohorts in this experiment. This suggests that children both with and without increased risk for reading difficulties benefitted from classroom PA instruction. These results indicate that a short and intensive period of classroom PA instruction can have a positive impact on raising reading achievement and minimising the risk for growing inequality between high- and low-ability students in the classroom.

7.2.4 Experiment 4: Reliability and validity of a computer-based phonological awareness screening and monitoring tool.

This aim of this experiment was to establish whether the computer-based PA screening and monitoring tool was a valid and reliable instrument to include as part of classroom assessment practices in the first year of school. In a longitudinal research design, the responses of 95 children to test items in the CBA at the start, middle and end of the first year at school were analysed to establish content, construct and criterion validity, and test-retest and internal consistency reliability. Results demonstrated that the majority of test items were appropriate for five-year-old children and sampled a range of ability levels. Tasks of rhyme oddity, initial phoneme identity and letter-knowledge were most suitable at school-entry with final phoneme identity, phoneme blending, phoneme deletion and phoneme segmentation becoming most appropriate by the middle and end of the school year. Significantly, performance on the computer-based PA assessment accounted for a large proportion of the variance in end-of-year reading performance and was able to predict end-of-year reading status with 94 per cent accuracy. These results suggest that the CBA developed in this thesis has sufficient reliability and validity to be confidently used as part of classroom assessment practices.

7.3 Efficiency and Effectiveness of Computer-Based Phonological Awareness Assessment

In this thesis, the first hypothesis stated that the computer-based screening and monitoring tool (described in Chapter 2) would be a time-efficient and effective method for monitoring PA development and predicting reading outcomes in the first year of formal education. The experiments reported in this thesis support this hypothesis. Chapters 3 and 6 provide evidence of time-efficiency and effectiveness from a technical perspective and Chapters 4 and 5 provide this evidence from a practical stance in the classroom.

7.3.1 Time-efficiency of computer-based phonological awareness assessment.

Time-efficient approaches to raising reading achievement and equalising reading outcomes are aligned with the practical needs of a busy classroom environment. Increasing the time-efficiency of PA assessment holds important implications for educators by: 1) ensuring PA can be quickly and easily measured in the classroom, 2) confirming that children have important requisite skill to take advantage of reading instruction, 3) informing curriculum design, and 4) monitoring response to reading instruction to identify children not keeping pace with same-aged peers. Research suggests that in the interests of balancing a busy classroom curriculum and numerous learning objectives, teachers may forgo methods and strategies that are ineffective and time consuming to implement (McLeod, Fisher, & Hoover, 2003). Thus, a key objective of this thesis was to examine whether computer-based technology could assist teachers in screening and monitoring PA in a time-efficient manner throughout the first year of school.

Research in this thesis demonstrated that CBA of a small number of key predictors of reading success, namely phoneme awareness and letter-knowledge, offers educators an excellent time-efficient alternative to paper-based testing procedures. The properties of CBA, including preparation and administration of test items, recording and scoring of responses, and interpretation of results, removes a significant proportion of labour that the classroom teacher would otherwise be required to complete (Bjornsson, 2008; Martin, 2008; Ripley, 2008; Singleton et al, 1999). Computer-based administration of PA test content took 31 per cent less time to administer than a paper-based counterpart, thereby relieving teachers to focus their attention on maximising instructional time in the classroom. It is important to note that this percentage represents savings in the time that the child was engaged in the assessment. This does not include the time the teacher would save while the child independently completes the assessment or the time saved by having the computer record, score and interpret performance. The savings in child administration time most likely came from improved and consistent presentation of test items (i.e., an educator did not have to organise and present picture cards) and automatic scoring of responses (i.e., an educator did not have to balance scoring of responses in between presentation of test items). These results are consistent with previous investigations including research by Olsen (1990), who demonstrated that non-adaptive CBA was at least 25 per cent faster to administer than paper-based methods. Moreover, these results show that CBA of PA can significantly reduce the quantity of assessment time required in the classroom while obtaining similar results to a more conventional paper-based equivalent.

Research in this thesis provides evidence that the computer-based PA tool offers a unique contribution to the variety of PA instruments already available to educators by offering a time-efficient alternative for utilisation in the classroom. In Chapter 1, Table 1.2 compared a selection of PA assessment tools available to classroom educators based on administration time, modality of assessment and content. This table has been adapted to include the computer-based PA assessment investigated in this thesis, and is presented below as Table 7.1. It is important to note that results regarding time-efficiency were derived from a sample of children approximately five-years of age using rhyme oddity (RO), initial phoneme identity (IPI) and letter-name and letter-sound (LK) subtests, as opposed to the entire CBA. Nonetheless, these results illustrate that computer-based PA assessment is time-efficient in comparison to a sample of existing paper-based measures, and provides evidence to support the first hypothesis addressed in this thesis.

Table 7.1

Comparison of phonological awareness assessments available to teachers and the computer-based assessment developed in this thesis

	Administration Time (in minutes)		Modality		Content		Classroom Example
	Teacher	*Child*	*Paper*	*Computer*	*Broad*	*Narrow*	*Teacher time required for a class of 12*
CTOPP*	30	30	+		+		6 hours
LAC*	20–30	20–30	+		+		4–6 hours
TOPA-2+ **	30–45	30–45	+			+	6–9 hours
DIBELS	7	7	+		+		1.5 hours
PALS-K**	20	20	+		+		4 hours
PIPA	25–30	25–30	+		+		5–6 hours
PA Profile	10–20	10–20			+		2–4 hours
Yopp-Singer	5–10	5–10	+			+	1–2 hours
CBA (RO; IPI; LK)	0	13–14		+		+	0 hours (except loading program)

Note. + indicates the modality and content within each assessment; * indicates that a specialised qualification or form of training is required for administration of the assessment; ** indicates that the assessment can be administered on an individual basis or to a small group of children; CTOPP = Comprehensive Test of Phonological Processing (Wagner et al., 1999); LAC = Lindamood Auditory Conceptualisation Test, Revised Edition (Lindamood & Lindamood, 1979); TOPA- 2+ = Test of PA, Second Edition: PLUS (Torgesen & Bryant, 2004); DIBELS = Dynamic Indicators of Basic Early Literacy Skills (Good & Kiminski, 2003); PALS-K = PA Literacy Screening—Kindergarten (Invernizzi et al., 2005); PIPA = Preschool and Primary Inventory of PA (Dodd et al., 2000); PA Profile = PA Profile (Robertson & Salter, 1995); Yopp-Singer = Yopp-Singer Test of Phoneme Segmentation (Yopp, 1995); pink shading = results from the current thesis.

7.3.2 Effectiveness of computer-based phonological awareness assessment.

Assessment tools must be effective in meeting their intended objectives if they are to be used confidently by classroom teachers. In this thesis, the intended objectives of developing a computer-based PA screening and monitoring tool were to provide everyday teachers with a time-efficient method of: 1) identifying children at risk for reading problems from the outset of beginning literacy instruction, 2) monitoring progress and identifying children not keeping pace with their peers, and 3) profiling strengths and weaknesses in PA to support curriculum design and referral to special education services for children with significant difficulties. The effectiveness of the CBA in addressing these intended purposes was established through measures of validity (e.g., content, construct and criterion) and reliability (e.g., test-retest and internal consistency).

Validity refers to the extent to which an instrument measures that which it is intended to measure (Messick, 1995). Estimates of content validity demonstrated that: 1) rhyme oddity and initial phoneme identity test items were most appropriate at school-entry and sampled a spectrum of difficulty levels, 2) more challenging phoneme-level test items (e.g., final phoneme identity, phoneme blending, phoneme deletion and phoneme segmentation) became increasingly appropriate and differentiated between high- and low-ability students by the middle and end of the first year of school, and 3) letter-knowledge test items were appropriate but declined in their ability to differentiate between high- and low-ability students as the year progressed. The majority of test items in the computer-based PA tool demonstrated a ‘fit’ to the ability spectrum of five-year-old children, sampled a range of difficulty levels and discriminated well between high- and low-ability children. These results provide evidence of content validity for the computer-based PA tool, and ensures teachers can confidently use the instrument to measure PA across the spectrum of abilities typically found in the classroom.

The effectiveness of an assessment can also be determined by how accurately it measures the construct (i.e., the ability) that it is designed to measure. In this thesis, the CBA included six tasks to measure PA specifically. There were also two tasks designed to assess letter-knowledge ability. Estimates of construct validity indicated that: 1) the six PA tasks measured the same underlying construct of PA, 2) there appeared to be a continuum of difficulty progressing from rhyme awareness, to initial phoneme identity, phoneme blending, final phoneme identity, phoneme deletion and phoneme segmentation, and 3) each PA task was sensitive in detecting growth in PA from the first half (the start to the middle) to the second half (the middle to the end) of the school year. These findings align with a broad definition of PA—that PA develops along a continuum from awareness of 'shallow' phonological units (e.g., syllables and rhyme) to awareness of 'deeper' phonological units (e.g., phonemes) (Anthony et al., 2003; Stanovich, 1992). Along this continuum, mastery at one point is not a requisite for development at a more complex level to begin; rather an overlap in skill development can be expected (Anthony et al., 2003). These results demonstrate that the CBA measures PA in a manner that is consistent with definitions of PA reported in the literature. Such findings provide classroom teachers with confidence that the computer-based PA tool effectively measures the construct that it is designed to measure; that is, PA and letter-knowledge in five-year-old children.

The effectiveness of a PA assessment can also be established through its ability to predict reading outcomes, and whether or not the assessment produces similar results to existing PA tools already held to be valid and reliable. Estimates of criterion validity in the form of predictive validity provided evidence that: 1) the six PA tasks in the CBA accounted for a large proportion of the variance in end-of-year word recognition ability, and 2) rhyme oddity and initial phoneme identity predicted end-of-year reading status (i.e., performance at or below an age-appropriate level) with 92 per cent accuracy at school-entry; increasing to 94 per cent accuracy when all six PA tasks were re-administered during the middle of the school year. Criterion validity in the form of concurrent validity also provided evidence that rhyme oddity, initial phoneme identity and letter-sound tasks were highly correlated with the PIPA: an existing standardised paper-based test with satisfactory levels of validity and reliability. Consistent with previous research, results demonstrated that PA is a more robust predictor of reading outcomes than measures of letter-knowledge, speech intelligibility, non-verbal intelligence, socio-economic status or gender. Accordingly, educators can be confident that the computer-based PA assessment will provide information regarding risk for reading difficulties and will inform curriculum design in the first year of education.

Finally, the ecological validity of an assessment can also determine its effectiveness in a classroom environment. Ecological validity refers to the extent to which the methods, materials and behaviours observed in a study reflect the real-life environment to which results will be generalised (Thorndike & Thorndike-Christ, 2010). In this thesis, CBA was used to monitor response to two different types of classroom reading instruction: 1) 10 weeks of PA instruction in addition to the 'usual' literacy curriculum, and 2) the 'usual' literacy curriculum only. The experiment reported in Chapter 4 demonstrated that the computer-based PA assessment measured growth in PA, reading and spelling for children who received either form of classroom reading instruction. Further, the CBA was able to identify significant differences in quantity of growth and end-of-year reading performance between the two different types of reading instruction. Similarly, the CBA was sensitive to different outcomes in reading performance between children with and without an elevated risk for reading difficulties. The practical implementation of the CBA in classroom-based experiments (see Chapters 4 and 5) provides evidence of ecological validity, and affords everyday educators with the confidence that this tool can be used effectively as part of classroom assessment practices.

Reliability refers to how consistent and precise a measurement tool is over repeated administrations (Crocker & Algina, 1986). In this thesis, measures of reliability showed: 1) high test-retest reliability between the computer-based PA assessment and an identical paper-based version, 2) high test-retest reliability between repeated administrations of each PA task from the start to the middle and the middle to the end of the school year; and 3) satisfactory levels of internal consistency across PA and letter-knowledge tasks. These results suggest that educators can be assured that the computer-based PA assessment is reliable over repeated administrations and will generate equivalent results regardless of the modality chosen (i.e., computer- or paper-based) for administration.

In summary, the research reported in this thesis provides evidence that the computer-based PA tool is a time-efficient and effective method for monitoring growth in PA and for predicting reading outcomes. This tool will allow teachers to measure key predictors of literacy success effectively and efficiently as part of classroom assessment practices, thereby supporting global initiatives that aim to raise achievement and reduce inequality in reading outcomes.

7.4 Classroom Phonological Awareness Instruction and Literacy Outcomes

The second hypothesis stated that teacher-implemented classroom-wide phonological awareness instruction that is time-efficient and focused at the phoneme level would significantly raise reading achievement for children with and without spoken language difficulties in the first year of school. The experiments reported in this thesis support this hypothesis. Chapters 4 and 5 provide evidence that 20 hours of PA instruction over 10 weeks can result in significant improvements in PA, reading and spelling for children with and without risk for reading problems in the classroom.

7.4.1 Efficient and effective classroom phonological awareness instruction and reading outcomes.

Programmes that aim to maximise learning outcomes must to be cognisant to the needs of the classroom. Time-efficient programmes are more manageable for teachers to integrate into existing literacy curricula. In particular, time-efficient instruction in PA is important for 1) managing a busy classroom schedule and multiple literacy objectives, 2) ensuring children have adequate phoneme awareness and letter-knowledge to take advantage of beginning reading instruction, and 3) identifying and supporting children who are not making expected progress. In this thesis, experiments sought to identify how the benefits of PA instruction, as demonstrated under controlled research conditions, could be efficiently and effectively replicated in the classroom reading programme to support early reading development.

As previously mentioned, there is wide scope for investigation regarding how to best integrate PA instruction into the classroom environment. A great deal of scientific evidence supporting the importance of PA to early reading success has been conducted outside the classroom environment under controlled research conditions (Ehri et al., 2001; Gillon, 2000a, 2005; Gillon & McNeill, 2009). More recently, efforts have been directed towards transitioning the benefits of PA into the classroom environment. Classroom-based studies that include a focus on PA instruction suggest that long-duration and high-intensity instruction produces sustained benefits for reading achievement and significantly lowers the prevalence of reading difficulties (Shapiro & Solity, 2008). Conversely, classroom-based studies of short duration (i.e., less than one academic year) have generally reported less positive outcomes for maintained reading improvements in comparison to longer duration programmes. However, short-duration programmes are more cognisant of the needs of the classroom timetable and are easier to integrate into existing literacy curricula. A key point of difference between classroom-based programmes is that those of long duration tend to focus on teaching phoneme-level skills, while those of shorter duration have generally focused on a broad range of PA skills (e.g., words, syllables, rhymes and phonemes) as opposed to those at the critical phoneme level (Fuchs et al., 2001; Justice et al., 2010; Shapiro & Solity, 2008). Thus, this thesis advances knowledge by exploring whether the pairing of short-duration classroom instruction and phoneme-focused PA teaching could produce similar results to classroom programmes of longer duration. To achieve this, the PAT programme, which involves 20 hours of instruction over a 10-week period and which has proven successful in controlled clinical research, was adapted for inclusion in the classroom reading programme.

The research reported in this thesis supports the second hypothesis by providing evidence that a short-duration, high-intensity, phoneme-focused programme implemented by classroom teachers as an adjunct to the 'usual' literacy curriculum can help elevate reading and spelling performance. Children who received 20 hours of PA instruction over 10 weeks performed significantly higher on PA, reading and spelling measures immediately following and up to six months post-instruction compared to children who followed the 'usual' literacy curriculum only. By the end of the first year at school, 20 per cent fewer experimental children performed below an age-expected reading level compared to comparison children (see Section 7.5.1 for more detailed prevalence information). This represents a similar reduction in the prevalence of reading difficulties to that reported by Shapiro and Solity (2008) who studied the use of a long-duration, high-intensity, phoneme-level programme in the classroom. These results suggest that a short-duration and phoneme-focused PA programme can offer educators a time-efficient alternative, while maintaining the benefits for reading outcomes, when competing priorities and paucity of time are issues in the classroom.

The specific focus on developing skills at the phoneme level in a short-duration and high-intensity framework may have contributed to positive literacy outcomes by:

1) focusing instruction at the level of PA that is most strongly related to early reading success (Gillon, 2004),
2) having a backwards chaining effect by which instruction at the phoneme level can influence development in syllable and rhyme awareness, although the reverse is less likely to occur (Brown, 1998; Yeh, 2003), and
3) providing instruction in a skill that is critical for reading development but that is not typically included in a whole language approach to reading instruction.

Notably, rhyme awareness appeared to develop in a similar trajectory between experimental and comparison children indicating that the 'usual' classroom programme was effective in raising this aspect of PA. However, the usual curriculum proved far less effective in developing skills at the phoneme level. Consistent with theories of learning to read (see Chapter 1, Section 1.3), the results of this thesis demonstrate the importance of PA at the phoneme level in reading acquisition and that instruction in this level of PA should be included as a critical component of beginning classroom reading programmes.

In Chapter 1, Table 1.3 compared the duration, intensity and content of instruction, in addition to reading outcomes (i.e., immediate versus maintained), from four recent classroom investigations that included a focus on PA. This table has been adapted to include results from the experiments reported in this thesis and is presented as Table 7.2. This table depicts how research reported in this thesis advances existing research knowledge by providing evidence that a short intensive period of phoneme-focused instruction in the classroom can contribute to accelerated literacy gains that are maintained up to six months post-instruction.

Table 7.2

Duration, intensity and content of PA instruction on reading outcomes

	Duration		Intensity		Content		Reading Outcomes	
	Long	Short	High	Low	Broad	Narrow	Immediate	Sustained
Shapiro & Solity (2008)	+		+			+	+	+
McIntosh et al. (2007)		+	+		+		+	
Fuchs et al. (2001)		+		+	+		+	
Justice et al. (2010)		+		+	+		+	
Current Study (thesis)		+	+			+	+	+

Note. + indicates the duration, intensity and content included in each study; 'Immediate reading outcomes' refers to improvements demonstrated immediately after the programme's conclusion; 'Sustained reading outcomes' refers to reading improvements still evident when measured up to five months post-instruction.

7.4.2 Efficient and effective classroom phonological awareness instruction for children with spoken language impairment.

Ensuring a classroom PA programme can raise reading achievement for children who are at risk for reading problems is critical in establishing the effectiveness of the programme. Children with SLI formulated the at-risk comparison group in this thesis. These children are reported to be four to five times more likely to experience reading and spelling difficulties due to deficits in the underlying language systems, including PA, which support written language development (Catts et al., 2001). Previous investigations in the classroom demonstrate that children with SLI present with a slowed and more variable response to teacher instruction that includes a focus on PA, and appear to benefit less in the acquisition of phoneme-level skills (Al Otaiba & Fuchs, 2002; Justice et al., 2010; Penno et al., 2002). Indeed Fuchs et al. (2002) reported that an equal number of children with language difficulties showed no reading improvements to those children who showed some improvement when exposed to a classroom PA programme.

The research reported in this thesis demonstrates that children with SLI can benefit from classroom reading instruction that is tailored to the phoneme level and is highly intensive over a short period. These children demonstrated significant improvements in phoneme awareness, reading and spelling relative to their pre-instructional performance. By the end of the first year at school, five out of seven children with SLI in the experimental condition performed at an age-appropriate level in reading fluency and comprehension. This performance was on par with the end-of-year reading outcomes demonstrated by children with TD in the comparison classrooms (see Section 7.5.2 for further details). Two children with SLI in the experimental condition attained results slightly below the raw score required to achieve an age-appropriate reading level of six years. These results demonstrate that phoneme-level skills can be stimulated in children with SLI when specifically targeted in instruction and that reading levels can be elevated to a typical level for the majority of children with SLI in a classroom environment.

A specific focus on phoneme-level skills may have contributed to positive reading outcomes for children with SLI. Research demonstrates that this cohort often presents with deficits in the phonological knowledge that supports word-decoding accuracy and may over-compensate for these deficits by relying on context during reading tasks (Kamhi & Catts, 2012). In the country in which this doctoral research was undertaken (New Zealand), school-aged children typically receive whole language literacy instruction. This form of literacy instruction does not include an explicit focus on developing word-level skills such as PA. Rather, children learn to extract meaning from the context. Results from this thesis suggest that children with SLI benefit from a more balanced approach to reading that combines instruction in PA and word-level knowledge alongside context-based instruction typically used in New Zealand classrooms.

Importantly, experimental children with SLI demonstrated reduced generalisation of PA knowledge to an untaught PA task and to written language tasks when compared to experimental children with TD. Although children with SLI demonstrated significant improvements from pre- to post-instruction, they were less able to transfer PA knowledge to an untaught skill (i.e., phoneme deletion), in comparison to children with TD. In addition, children with SLI showed significantly less gain in reading and spelling development compared to children with TD. These findings indicate that children at risk for reading problems require support to ensure their enhanced PA knowledge is generalised and applied to reading and spelling development. This is an area of research that requires careful consideration in future studies.

In summary, research reported in this thesis provides evidence that teacher-implemented classroom-wide PA instruction that is time-efficient and focused at the phoneme level can significantly raise reading achievement for children with and without SLI in the first year of school. Providing teachers with methods to efficiently and effectively teach key predictors of literacy success as a component of the beginning classroom reading programme may support initiatives that pursue elevated reading achievement for school-aged children.

7.5 Reducing the Prevalence of Reading Difficulties and Reading Inequality

The third hypothesis stated that teacher-implemented classroom-wide phonological awareness instruction would have a positive effect on reducing the prevalence of reading problems and achieving greater equality in reading outcomes when included as part of the beginning literacy curriculum. The experiments reported in this thesis support this hypothesis. Specifically, Chapters 4 and 5 provide evidence that the number of children experiencing reading problems and large gaps between high- and low-ability students can be positively influenced by teacher-implemented instruction in PA in the classroom.

7.5.1 A positive influence on the prevalence of reading difficulties.

International prevalence statistics suggest that as many as one in three children may experience hardship in acquiring foundational literacy skills (NAEP, 2003; Nicolson, 2009). Unresolved reading difficulties often persist and precipitate ongoing disparities in literacy performance and academic underachievement (Gillon, 2004). The quasi-experimental study reported in Chapter 4 provides insight into how classroom PA instruction can influence the number of children presenting with reading problems after one year of school. Results demonstrated that at the end of the school year, approximately 6 per cent of children who received classroom PA instruction performed below an age-expected level in reading accuracy compared to 26 per cent of children who followed the ‘usual’ literacy curriculum only. Further, approximately 6 per cent of children who received classroom PA instruction performed below an age-expected level in reading comprehension compared to 32 per cent of children in the comparison condition. It is important to recognise that longer-term follow up is required to understand more precisely the effect of classroom PA instruction on later reading outcomes. Nonetheless, these initial results hold promise that modification to the classroom curriculum to include PA can have a significant effect on raising reading achievement for the majority of children.

7.5.2 Achieving greater equality in reading outcomes in the first year of school.

International studies of literacy achievement demonstrate that large inequalities between good and poor readers are prevalent among many of the OECD's wealthiest countries. This has led international associations (e.g., the United Nations, UNICEF) to call for a stronger global commitment towards reducing inequality in reading outcomes, particularly for children who are vulnerable for literacy difficulties (UNESCO, 2009; UNICEF, 2010). The research reported in this thesis contributes to this global commitment by examining the influence of classroom PA instruction on end-of-year reading inequality between children with and without risk for reading problems. As mentioned, the at-risk comparison group in this thesis was children with SLI.

Research presented in Chapter 5 provides evidence that the gap between children with high and low risk for reading difficulties was influenced in several ways as a function of the type of classroom reading programme received (e.g., the 'usual' literacy curriculum and 10 weeks of PA, or the 'usual' literacy curriculum only). End-of-year reading outcomes for children with SLI who received classroom PA instruction were compared to: 1) children with SLI who followed the 'usual' programme (i.e., comparison condition), 2) children with TD who followed the 'usual' programme (i.e., comparison condition), and 3) children with TD who received classroom PA (i.e., experimental condition). Three different types of gaps in reading performance were identified. First, children with SLI in the experimental condition performed significantly higher than children with SLI in the comparison condition. Second, children with SLI in the experimental condition performed on par with children with TD in the comparison condition. Third, children with SLI in the experimental condition performed significantly lower than children with TD who were also in the experimental condition. Most importantly, these results show that the reading outcomes for a group of children who are traditionally at high risk for reading problems can be raised to an age-appropriate level by the end of the first year at school. In a New Zealand context, this means that children who are traditionally more likely to be candidates for Reading Recovery after one year of school may no longer require this service, thereby reducing demands on resources and allowing children most in need to receive supplementary support outside the classroom.

Although inequalities remained between children with SLI and TD who received classroom PA instruction, this result demonstrates that the integration of PA instruction into an existing reading programme can benefit the majority of children in the classroom, not just vulnerable readers. This is because children with TD performed on average 11 months and 9 months, respectively, in advance of their chronological age for reading accuracy and comprehension after one year at school. Conversely, children with TD who did not receive classroom PA instruction performed at or only slightly above their chronological age on end-of-year reading measures. These initial results indicate that modification to the classroom reading programme to include key predictors of literacy success can have a positive influence on the gap between children with high and low risk for reading problems. Moreover, this can be achieved in a time-efficient way by everyday classroom teachers.

7.6 Summary of Findings

The aim of this thesis was to investigate how PA can be efficiently and effectively integrated into beginning classroom literacy programmes to improve reading outcomes. This was achieved from two complementary perspectives: 1) by providing teachers with a time-efficient computer-based method to screen and monitor PA development in the classroom, and 2) by providing teachers with a short-duration PA programme that is effective in raising reading achievement and equalising literacy disparities in the first year of school. Based on evidence reported in this thesis, the following conclusions can be drawn:

1. Computer-based PA screening and monitoring can reduce child assessment time by over 30 per cent in comparison to paper-based administration. This percentage does not include the time the teacher saves while the child is independently administering the CBA or that saved by not having to organise, record and aggregate results.
2. Computer-based PA screening and monitoring in the first year of school predicted end-of-year reading status (i.e., performance at, below or above an age-expected reading level) with 94 per cent accuracy. Repeated administration during the middle of the school year reduced the false positive rate by 50 per cent. Only four per cent of children were unidentified by the tool as being at-risk for reading problems.
3. Short-duration, high-intensity and phoneme-focused PA instruction delivered by teachers as part of the beginning classroom reading programme can significantly raise the PA, reading and spelling abilities of children with and without SLI.
4. Twenty hours of PA instruction focused at the phoneme level for 10 weeks in the first year of school was found to reduce the prevalence of reading difficulties to a similar level found for classroom PA programmes of longer duration (e.g., Shapiro & Solity, 2008).
5. Classroom PA instruction that is short in duration, highly intensive and focused at the phoneme level can have a positive effect on minimising potential inequalities

in reading outcomes. Children with SLI who received PA instruction performed at an age-appropriate reading level after one year of schooling. This was significantly higher than children with SLI who followed the 'usual' literacy curriculum only. Despite both groups reading at or above a typical level, a significant reading gap existed between experimental children with SLI and TD. This gap illustrates that both children with and without elevated risk for reading problems can benefit from classroom-wide PA instruction.

7.7 Advancing understanding about theories of reading through thesis findings

Understanding how children learn to read is critical for teachers who aim to implement effective literacy curriculums in their classrooms. Research reported in this thesis advances current understanding about theories of learning to read, particularly for: 1) the modified dual-route model of word recognition, 2) the self-teaching hypothesis and 3) the interactive-compensatory model of reading. Each of these frameworks for conceptualising reading acquisition highlights the critical role of phonological knowledge for raising achieving and equalising reading outcomes for school-aged children.

7.7.1 Support for the modified dual-route model of word recognition.

The modified dual-route model of word recognition (Ehri, 1992) proposes the use of either a phonological or a visual-phonological route to the recognition of single words. The phonological route involves segmenting words into letters and then matching these to phonemes to deduce word meaning. To use the phonological route, the reader must have an understanding of how graphemes are represented by phonemes, as well as the ability to segment and blend individual phonemes (i.e., PA) to construct accurate phonological representations. A visual-phonological route involves the aggregation of visual and phonological information to decode unfamiliar words, in particular irregular words. In the visual-phonological route, readers decode parts of an irregular word that are in fact regular using phonological information. Irregular parts of an irregular word are then recognised using visual information. For the most part, children first learn to read a word via the phonological route, using knowledge of phoneme-grapheme connections and PA. With increasing practice, children begin to recognise printed words more quickly by combining visual and phonological information.

Research reported in this thesis highlights the importance of phonological skills within the modified dual-route model of word recognition. This is achieved by demonstrating that children who received classroom instruction in PA and how this knowledge applies to word decoding became significantly better readers than children who received whole language instruction. Teaching children the underlying phonological-mechanics of word decoding from the outset of beginning literacy instruction may enrich the quality of the phonological and visual-phonological routes available to children when learning to read. This in turn may allow access to a greater number of successful decoding experiences leading towards enhanced reading fluency as indicated on the Neale Analysis of Reading Ability (Neale, 1999) after one year of school. These results demonstrate that the integration of PA into an existing curriculum, which primarily focuses on using contextual cues to recognise printed words, may support enhanced reading outcomes for all children.

7.7.2 Support for the self-teaching hypothesis.

The self-teaching hypothesis states that children begin to self-teach new grapheme-phoneme connections in the reading process through increased exposure and practice at converting letters into sounds (Share, 1995). Each successful attempt at decoding printed words enhances knowledge about the relationship between speech and print, thereby strengthening children’s access to a written world. Children who approach reading instruction with a strong ability to convert graphemes into phonemes are more able to engage in self-teaching. Conversely, children who approach reading instruction with a poor ability to translate graphemes into phonemes are less able to self-teach in the reading process. These initial differences may contribute to inequalities in reading outcomes in the later school years.

The research reported in this thesis suggests that explicit instruction in PA and letter-knowledge supports self-teaching in word recognition which in turn may help establish equality in reading outcomes. Children who received instruction in phonological skills which support self-teaching were significantly better readers by six-years-of-age compared to children who received instruction focused on context and visual word learning. Moreover, children at-risk for reading problems who received class PA instruction performed at an age-appropriate level in reading development after one year of school. Comparatively, at-risk children who continued with the 'usual' reading programme performed significantly lower than an age-expected level. These results suggest that classroom reading instruction that includes a focus on teaching children PA and letter-knowledge skills may help both typically developing and at-risk readers to begin self-teaching, thereby enhancing reading outcomes and reducing inequality for the majority of children.

7.7.3 Support for the interactive-compensatory model of reading.

The interactive-compensatory model of reading emphasises the integration of phonological, orthographic, syntactic, semantic and lexical knowledge for accessing the meaning of connected text (Stanovich, 1980, 1984). Intrinsic to this model is the notion that deficiencies in any of these processes can be compensated for by another skill area. Children at-risk for reading difficulties, in particular those with SLI, are known to compensate for deficits in phonological knowledge by over-relying on the context afforded by connected text to guess word meaning. Using this strategy may be successful when beginning-reading books are predictive in structure and contain pictures and high-frequency words. However, this strategy becomes less successful as text difficulty increases, pictures are removed, and word choice is increasingly constrained. Further, the cognitive resources required to access meaning using context, as opposed to phonological information, can reduce the amount of resources available for achieving skilled reading comprehension. Consequently teaching children to use a phonological decoding strategy to support successful reading development, as opposed to compensatory strategies, can be viewed as a critical component of beginning reading instruction.

The research reported in this thesis demonstrates the importance of teaching children at-risk for reading difficulties to apply a phonological decoding strategy to word recognition in connected text; thereby minimising the need to over-rely on context. Children with SLI are one cohort who are four to five times more likely to experience reading difficulties (Catts et al., 2001), and who are reported to over-rely on context to guess word meaning because of deficits in PA knowledge (Kamhi & Catts, 2012). In this thesis, children with SLI who received class PA instruction obtained significantly higher reading fluency scores in connected text after one year of schooling in comparison to children with SLI who received whole language instruction. These children also obtained significantly higher reading comprehension scores than those children with SLI who did not receive PA instruction. Teaching children at-risk for reading problems to use a phonological decoding strategy to recognise printed words can, therefore, significantly raise not only word reading fluency but may also help increase the cognitive resources available to comprehend written information at a connected text-level. These findings suggest that the inclusion of PA instruction as part of beginning classroom reading programmes can help children at-risk to use a phonological decoding strategy when learning to read; thereby reducing their need to rely on context to compensate for deficiencies in phonological knowledge.

7.8 Implications for Classroom Reading Practices: National and International Needs

A specific aim of this thesis was to bridge theoretical perspectives on reading and PA into the classroom environment as a means towards raising achievement and reducing inequalities in reading outcomes. The research reported in this thesis demonstrates that PA can be effectively and efficiently integrated into beginning reading curriculums by combining computer-based PA assessment with short and intensive phoneme-focused instruction. The findings reported in this thesis have a number of implications at both national and international levels.

7.8.1 Implications for classroom reading practices in New Zealand.

Research reported in this thesis has implications for classroom reading practices in New Zealand. In New Zealand, a significant barrier to the routine screening and monitoring of PA is the lack of instruments that have been designed, trialled and used with New Zealand school-aged children (Klee & Tillard, 2010). Assessments commonly used in New Zealand classrooms to measure reading-related skills either do not assess PA or provide information on 'shallow' levels of PA only (e.g., syllables and rhyme) as opposed to development at the critical phoneme level (Carson, Gillon, & Boustead, 2010). A second barrier is time-efficiency, with up to 70 per cent of New Zealand teachers reporting that they do not use the widely available SEA, for example, to measure school-entry numeracy and literacy because it is time consuming (Dewar & Telford, 2003). The research reported here reduces these two barriers by contributing information towards one of the first databases on PA development in New Zealand school-aged children, and demonstrating that PA data can be collected in a time-efficient manner.

In terms of classroom reading instruction, New Zealand educators are currently faced with the challenge of reducing one of the OECD's largest gaps between high- and low-achieving readers (Martin et al., 2007). Currently, a whole language approach to teaching reading is the method of choice in New Zealand schools, although the teaching of phonics as an adjunct is becoming increasingly popular. Children struggling with reading acquisition after one year of schooling are provided with Reading Recovery, which essentially provides whole language instruction in a more intensive manner outside the classroom (Clay, 2002). Explicit and systematic teaching in PA is not common in either classroom or one-to-one reading support programmes. It has been argued that a whole language approach to reading instruction has done little to support equality in reading outcomes in New Zealand classrooms (Tunmer et al., 2006).

The research reported in this thesis holds important implications for classroom reading instruction in New Zealand. This is because evidence in this thesis demonstrates that integration of PA instruction into existing classroom reading programmes can 1) raise reading achievement, 2) help reduce the prevalence of reading difficulties and 3) have a positive effect on reducing the reading ability gap between children with and without elevated risk for literacy problems. Integrating PA instruction into existing classroom practices can be time-efficient, requiring only 20 hours of instruction over a 10-week period. It can therefore be said to be cognisant to the needs of a busy classroom schedule. Moreover, the ability to raise reading achievement and reduce inequality by the end of the first year of formal schooling may help stimulate a shift in New Zealand literacy practices away from the concept of ‘recovery’ from reading difficulties (i.e., the Reading Recovery Programme) to that of ‘preventing’ reading problems in the early school years.

In addition, the research reported in this thesis holds promise for raising achievement for young Māori children under the *‘Ka Hikitia - Managing for Success: The Māori Education Strategy 2008 – 2012’* (MOE, 2012a). This strategy is designed to improve the New Zealand education system to meet the learning needs of an indigenous Māori population. One of the four focus areas of this strategy is the *'foundation years'* in which achieving a strong platform for future learning during the early childhood and schooling years is emphasised. This thesis contributes to this focus area by demonstrating that reading achievement is likely to be raised in response to a short and intensive period of PA instruction in a classroom environment that is multi-cultural and includes a cohort of young Māori children. Table 4.2 in Chapter 4 illustrates the ethnic composition of the 129 five-year-old children who participated in either classroom PA instruction or the 'usual' reading curriculum in their first year of school. Inclusive in Table 4.2 is the percentage of Māori children who participated in experimental Groups A and B and comparison Group C. Detailed analysis of reading outcomes for this subgroup of children was not included in this thesis; but is however an important area of investigation in future studies.

7.8.2 Implications for classroom reading practices internationally.

Internationally, large inequalities between the abilities of good and poor readers have been identified. Countries such as Finland, Denmark, Ireland and Canada demonstrate a small gap between high- and low-ability readers, whereas countries such as America, the United Kingdom, France and New Zealand present with much greater variability between good and poor readers (see Table 1.1, Chapter 1; UNICEF, 2010). Since the research reported in this thesis was conducted in a country in which large discrepancies between high- and low-achieving readers have been reported, the findings reported here have important implications for countries faced with a similar literacy challenge; that is, a challenge to reduce the gap between good and poor readers.

Countries presenting with large inequalities in reading outcomes may benefit from computer-based methods to screen and monitor PA development from school-entry, in addition to time-efficient yet effective teacher-implemented PA programmes as part of classroom literacy practices. As mentioned, little is known about how to best integrate PA instruction into the heterogeneous classroom environment to raise reading achievement internationally. Replication of the results reported in this thesis to evaluate the effect of time-efficient PA assessment and measurement in classrooms throughout the wider international community are required. Such future investigations should consider: 1) cultural differences (e.g., pictures and words would need to be culturally appropriate), 2) transparency of the language (e.g., the regularity with which phonemes match graphemes and vice versa), 3) bi- or multi-lingualism (e.g., the effect of measuring and instructing PA in one or multiple languages), and 4) differences in educational systems (e.g., country, state and school objectives and methods used to achieve these). Capitalising on a key predictor of early literacy success, namely PA, and ensuring it can be measured and instructed upon in an easy and efficient manner by everyday classroom teachers may contribute to global initiatives aimed at raising achievement and reducing inequalities in school-aged reading outcomes.

Further, Response to Intervention (RTI) is receiving much attention as a method of preventing or minimising reading problems in everyday classrooms (see Chapter 1, Section 1.4.1.2). RTI aims to prevent academic difficulties, including reading problems, through research-based classroom programmes, screening and monitoring of student progress, and the implementation of increasingly intensive layers of support for struggling children (Nelson, 2010). Within this framework, researchers have recently called for improved screening instruments for educators and a greater understanding regarding the duration of classroom instruction to help determine when more intensive support is required (Kamhi & Catts, 2012). The research reported in this thesis has implications for these two areas of research need by providing evidence that 1) a computer-based method of screening and monitoring PA is both time-efficient and effective in the classroom environment for predicting reading difficulties, and 2) 20 hours of PA instruction over 10 weeks focused at the phoneme level can have a positive influence on raising reading achievement and providing information regarding which children are not responding to classroom instruction and may require ongoing support in one-to-one or small group settings. Replication of the experiments reported in this thesis within a RTI framework abroad using multiple sites across multiple countries will help validate the role that the findings in this thesis can play in supporting classroom reading programmes internationally.

7.9 Limitations and Directions for Future Research

The positive results reported throughout the experiments in this thesis must be interpreted within the context of study limitations. One limitation across the four experiments is the small number of participants; particularly for children presenting with MSD (see Chapter 3) and SLI (see Chapters 4 and 5). Larger sample sizes provide greater certainty regarding the ability to generalise experimental findings to other contexts (Field, 2009). This limitation can be overcome in future investigations by employing larger sample sizes and engaging in research that aims to replicate the results reported here.

Moreover, evidence regarding the time-efficiency of PA assessment was based on four subtests deemed most appropriate for children at school-entry, namely rhyme oddity, initial phoneme identity, letter-name and letter-sound knowledge. This limitation can be addressed in future research by investigating the time-efficiency of all six PA tasks in comparison to paper-based counterparts at not only the start but also at the middle and at the end point of the first year of school. This would provide confirming evidence that the computer-based PA assessment developed in this thesis is indeed more time-efficient than paper-based testing. Further, the savings in time of using the CBA are representative of savings in child administration time. These results do not take into consideration the quantity of time the teacher saves by letting the child independently administer the assessment or the time saved by not having to organise, record and score results. Future investigations that examine these additional time savings are required. Examining the time-efficiency of the computer-based PA assessment when delivered as an adaptive assessment would also be of interest in future investigations. This would require the development of a large test-item bank and software programming so that the computer modifies presentation of test stimuli based on children's responses. Results from such investigations may demonstrate that computer-based PA assessment is in actual fact significantly more time-efficient than has been reported in this thesis.

A number of additional variables may mediate how well the computer-based PA assessment can be used in the classroom environment. These include educators' own knowledge of PA, confidence in using a computer-based medium for assessment, availability of technology in the classroom and population demographics (e.g., socio-economic status, gender and ethnic grouping). Each of these areas requires in-depth investigation in future studies to validate the use of the CBA across a wide number of populations and educational settings. The focus of such investigations may be on professional development for teachers in how to measure PA using a computer-based medium, adaptation to test stimuli and instructions to enhance cultural appropriateness (e.g., for Māori and Pasifika students), and on increasing efficiency of delivery in a paper-based form for localities that are limited in their access to computer technology.

In this thesis, an important finding was accelerated reading outcomes for children with and without SLI in response to classroom PA instruction. It is important to note that some research design features may mediate generalisation of this result. The use of a quasi-experimental design meant that teachers and child participants were not randomly assigned at an individual level to each instructional condition. This limitation represented a trade-off between examining PA instruction in a real-life environment (i.e., the classroom) and using a controlled research context within which the effectiveness of PA instruction had already been comprehensively established. From a controlled research perspective, a quasi-experimental design may limit the causal relationships implied in this thesis, yet from an ecological validity standpoint, the quasi-experimental design enhances the practical application of research findings. In addition, it is plausible that generalisation of research findings may be confounded by variables of teacher (i.e., number of years' experience), child (e.g., gender and cultural background) and educational setting (e.g., socio-economic locality or language of the school). Further research can overcome uncertainties in study design and extraneous variables through replication studies across multiple intervention sites both nationally and internationally.

Of particular interest in New Zealand is the replication of these thesis results with a large cohort of young Māori children. This would support the '*Ka Hikitia - Managing for Success: The Māori Education Strategy'* (MOE, 2012a) and may require the adaptation of PA assessment and instruction resources to include Māori pictures and Māori words. Similar research methods to those employed this this thesis can then be implemented by educators across full and partial Māori immersion settings, in addition to mainstream education settings, where young cohorts of Māori children reside.

An additional variable that may have influenced the effectiveness of classroom PA instruction is the provision of professional development to experimental teachers and the reversal back to the 'usual' literacy curriculum. Experimental teachers received eight hours of professional development, including in-class support, whereas teachers continuing with the 'usual' literacy programme did not receive any form of professional development. It was noted that although experimental teachers no longer had specific blocks of time allocated to teaching PA or specific PA resources when reverting back to the 'usual' curriculum, they did have a tendency to spontaneously use phoneme blending or segmentation strategies in classroom literacy-based tasks. Even if these behaviours did have an influence on study findings, they can be interpreted to suggest that professional development had a positive and long-lasting effect on improving teachers' instructional reading practices in the classroom. Future investigations regarding the effect of professional development in PA instruction techniques on the reading outcomes of children with and without risk for literacy problems is required.

Finally, in this thesis, investigation into the maintenance of enhanced reading and spelling skills following classroom PA instruction was only conducted until six months post-instruction. At this point, a significant reduction in the number of children presenting with reading problems was identified between those who received classroom PA instruction and those who followed the 'usual' reading programme. Future follow-up investigations to examine the reading abilities of these participants would add validity to the claims of a reduction in the prevalence of reading problems tentatively stated in this thesis.

7.10 Conclusion

Investigating methods that ensure children read proficiently in their own classrooms is a critical issue in reading education. The number of children experiencing reading difficulties, in addition to inequalities between good and poor readers among the world's wealthiest countries, raises concerns regarding how far children are falling behind in literacy acquisition (UNESCO, 2009; UNICEF, 2010). A large number of initiatives are currently underway to determine how reading achievement can be raised within everyday classrooms. One method investigated in this thesis was the efficient and effective integration of a key predictor of literacy success, namely PA, into beginning classroom reading programmes. To achieve this, the research reported here demonstrates that PA can be integrated into existing reading programmes efficiently and effectively using two complementary methods. First, computer-based technology can be used to screen, monitor and predict risk for reading difficulties and, second, classroom-wide but short term PA instruction focused at the phoneme level can be implemented to help ensure children have the requisite skills to take advantage of beginning reading instruction. Equipping classroom teachers with the tools to efficiently and effectively screen, monitor and instruct in PA as part of a comprehensive multi-focal literacy curriculum offers an important contribution towards international initiatives that seek to raise achievement for every child. This will help ensure all children have the opportunity to prosper on their journey towards becoming proficient readers.

References

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Adams, M. J., Foorman, B., Lundberg, I., & Beeler, C. (1997). *Phonemic awareness in young children: A classroom curriculum*. Baltimore: Brookes.

Adolf, S. M., Catts, H. W., & Little, T. D. (2006). Should the Simple View of Reading include a fluency component? *Reading and Writing, 19*, 933-958.

Al Otaiba, S., & Fuchs, D. (2002). Characteristics of children who are unresponsive to early literacy intervention: A review of the literature. *Remedial and Special Education, 23*, 300-316.

Al Otaiba, S., Kosanovich, M. L., & Torgesen, J. K. (2012). Assessment and instruction in phonemic awareness and word recognition skills. In A. G. Kamhi & H. W. Catts (Eds.), *Language and reading disabilities* (3rd ed., pp. 112-145). Upper Saddle River, NJ: Pearson Education Inc.

Al Otaiba, S., Puranik, C., Zilkowksi, R., & Curran, T. (2009). Effectiveness of early phonological awareness interventions for students with speech or language impairments. *Journal of Special Education, 43*, 107-128.

Al Otaiba, S., & Torgesen, J. (2007). Effects from intensive standardized kindergarten and first grade interventions for the prevention of reading difficulties. In S. R. Jimerson, M. K. Burns & A. M. Van Der Heyden (Eds.), *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 212-222). NY: Springer.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

Anderson, P., Lindsey, J., Schulz, W., Monseur, C., & Meiers, M. (2004). *An evaluation of the technical and methodological aspects of the School-Entry Assessment Kit*. Victoria, Australia: Australian Council for Educational Research Ltd.

Andrich, D. (2004). Controversy and the rasch model: A characteristic of incompatible paradigms? *Medical Care, 42*, 1-16.

Anthony, J. L., & Francis, D. J. (2005). Development of phonological awareness. *Current Directions in Psychological Science, 14*, 255-259.

Anthony, J. L., & Lonigan, C. J. (2004). The nature of phonological awareness: Converging evidence from four studies of preschool and early grade school children. *Journal of Educational Psychology, 96*, 43-55.

Anthony, J. L., Lonigan, C. J., Driscoll, K., Phillips, B. M., & Burgess, S. R. (2003). Phonological sensitivity: A quasi-parallel progression of word structure units and cognitive operations. *Reading Research Quarterly, 38*, 470-487.

Australian Government. (2005). *National inquiry into the teaching of reading: Teaching recommendations*. Victoria, Australia: Department of Education, Science and Training. Retrieved 9th March 2009 from: www.dest.gov.au/nitl/report.htm.

Baddeley, A. (1986). *Working memory*. NY: Oxford University Press.

Baker, F. (2001). *The basics of item response theory: ERIC clearinghouse on assessment and evaluation*. University of Maryland, College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.

Barnes, S. K. (2010). *Using computer-based testing with young children*. Paper presented at the Northeastern Educational Research Association (NERA) Conference. Retrieved 10th January 2012 from: http://digitalcommons.uconn.edu/nera_2010/22

Barnett, S. (2003). *Low wages = low quality: Solving the real preschool teacher crisis (Preschool Policy Brief No. 3)*. New Brunswick, NJ: National Institute for Early Education Research Rutgers University.

Bernthal, J. E., Bankson, N. W., & Flipsen, P. (2009). *Articulation and phonological disorders* (6th ed.). Boston, MA: Allyn & Bacon.

Bird, J., Bishop, D., & Freeman, N. (1995). Phonological awareness and literacy development in children with expressive phonological impairments. *Journal of Speech and Hearing Research, 36*, 446-462.

Bishop, D., & Norbury, C. F. (2008). Speech and language disorders. In M. Rutter, D. Bishop, D. Pine, S. Scott, J. Stevenson, E. Taylor & A. Thapar (Eds.), *Rutter's child and adolescent psychiatry* (pp. 782-801). Oxford: Blackwell.

Bishop, D. V., & Adams, C. (1990). A prospective study of the relationship between specific language impairment, phonological disorders, and reading retardation. *Journal of Child Psychology and Psychiatry, 31*, 1027-1050.

Bjornsson, J. (2008). Changing Icelandic national testing from traditional paper and pencil based tests to computer based assessment: Some background, challenges and problems to overcome. In F. Scheuermann & A. G. Pereira (Eds.), *Towards a research agenda on computer-based assessment* (pp. 10-14). Luxembourg: Office for Official Publications of the European Communities.

Blachman, B. A., Ball, E. W., Black, R. S., & Tangel, D. M. (2000). *Road to code: A phonological awareness program for young children*. Baltimore: Paul Brookes.

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch Model*. Mahwah, NJ: Erlbaum.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model: Fundamental measurement in the human sciences* (2nd ed.). London, UK: Erlbaum.

Boudreau, D. M., & Hedberg, N. L. (1999). A comparison of early literacy skills in children with specific language impairment and their typically developing peers. *American Journal of Speech-Language Pathology, 8,* 249-260.

Bowen, C. (2009). *Children's speech sound disorders*. Oxford: Wiley-Blackwell.

BPS. (1999). *Guidelines for the development and use of computer-based assessment*. Leicester, UK: British Psychology Society.

Bradley, L., & Bryant, P. (1983). Categorizing sounds and learning to read - A causal connection. *Nature, 301,* 419-421.

Bradley, L., & Bryant, P. (1985). *Rhyme and reason in reading and spelling*. Ann Arbor, MI: University of Michigan Press.

Bridgeman, B. (2009). Experiences from large-scale computer-based testing in the USA. In F. Scheuermann & J. Bjornsson (Eds.), *The transition to computer-based assessment. New approaches to skills assessment and implications for large-scale testing* (pp. 39-44). Luxembourg: Office for Official Publications of the European Communities.

Briscoe, J., Bishop, D. V. M., & Norbury, C. F. (2001). Phonological processing, language, and literacy: A comparison of children with mild-to-moderate sensorineural hearing loss and those with specific language impairment. *Journal of Child Psychology and Psychiatry, 42,* 329-340.

Brown, G. D. A. (1998). The end point of skilled word recognition: The ROAR model. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 121-138). Mahwah, NJ: LEA.

Bryant, D., Clifford, R. M., Salvja, G., Planta, R. C., Early, D. M., Barbarin, O., … Burchinal, M. (2002). *Diversity and directions in state pre-kindergarten programs*. Chapel Hill, NC: The University of North Carolina, FPG Child Development Institute, NCEDL.

Burgess, S. R., & Lonigan, C. J. (1998). Bidirectional relations of phonological sensitivity and prereading abilities: Evidence from a preschool sample. *Journal of Experimental Child Psychology, 70*(2), 117-141.

Bus, A., & Van Ijzendoorn, M. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology, 91*, 403-411.

Byrne, B., Fielding-Barnsley, R., & Ashley, L. (2000). Effects of preschool phoneme identity training after six years: Outcome level distinguished from rate response. *Journal of Educational Psychology, 92*, 659-667.

Calder, H. (1992). *Reading freedom teacher's manual*. New South Wales, Australia: Pascal Press.

Carroll, J., Gillon, G., & McNeill, B. (2011). *Phonological Awareness: Early Childhood Teachers' Knowledge*. Paper presented at the American Speech, Language, and Hearing Association Convention (ASHA), 17-19 Nov 2011, San Diego, CA, USA.

Carroll, J., & Snowling, M. (2004). Language and phonological skills in children at high risk of reading difficulties. *Journal of Child Psychology and Psychiatry, 45*(3), 631-645.

Carroll, J. M., Snowling, M. J., Hulme, C., & Stevenson, J. (2003). The development of phonological awareness in preschool children. *Developmental Psychology, 39*(5), 913-923.

Carson, K., Gillon, G., & Boustead, T. (2011). Computer-administered versus paper-based assessment of school-entry phonological awareness ability. *Asia Pacific Journal of Speech, Language, and Hearing, 14*(2), 85-101.

Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition, 47*, 149-180.

Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition, 91*, 77-111.

Cataldo, S., & Ellis, N. (1988). Interactions in the development of spelling, reading and phonological skills. *Journal of Research in Reading, 11*, 86-109.

Catts, H. W., Bridges, M. S., Little, T. D., & Tomblin, J. B. (2008). Reading achievement growth in children with language impairment. *Journal of Speech, Language, and Hearing Research, 51*(6), 1569-1579.

Catts, H. W., Fey, M. E., Tomblin, J. B., & Zhang, Z. (2002). A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language, and Hearing Research, 45*, 1142-1157.

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (1999). Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. *Scientific Study of Reading, 3*, 331-361.

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (2001). Estimating the risk of future reading difficulties in kindergarten children: A research-based model and its clinical implementation. *Language, Speech, and Hearing Services in Schools, 32*, 38-50.

Catts, H.W., & Hogan, T. P. (2002). *At what grades should we assess phonological awareness?* Paper presented at the American Speech, Language, and Hearing Association Convention (ASHA), San Francisco, CA.

Catts, H. W., Hogan, T. P., Adolf, S. M., & Barth, A. E. (2003). *The simple view of reading: Changes over time*. Retrieved 10th January 2012 from: http://www.ku.edu/~splh/Catts/poster7.pdf

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

Catts, H. W., & Kamhi, A. G. (1999). Causes of reading disabilities. In H. W. Catts & A. G. Kamhi (Eds.), *Language and reading disabilities* (pp. 95-127). Boston: Allyn & Bacon.

Catts, H. W., Kamhi, A. G., & Adlof, S. (2012). Defining and classifying reading disabilities. In A. G. Kamhi & H. W. Catts (Eds.), *Language and reading disabilities* (3rd ed.). Upper Saddle River, NJ: Pearson Education Inc.

Catts, H. W., Petscher, Y., Schatschneider, C., Bridges, M. S., & Mendoza, K. (2009). Floor effects associated with universal screening and their impact on the early identification of reading disabilities. *Journal of Learning Disabilities, 42*, 163-176.

Catts, H. W., Wilcox, K. A., Wood-Jackson, C., Larrivee, L. S., & Scott, V. G. (1997). Toward an understanding of phonological awareness. In C. K. Leong & R. M. Joshi (Eds.), *Cross-language studies of learning to read and spell: Phonologic and orthographic processing* (pp. 31-52). Dordecht: Kluwer Academic Press.

Chapman, J. W., & Tunmer, W. E. (1997). A longitudinal study of beginning reading achievement and reading self-concept. *British Journal of Educational Psychology*, *67*, 279-291.

Chard, D. J., & Dickson, S. (1999). Phonological awareness: Instructional and assessment guidelines. *Interventions in School and Clinic, 34*, 261-170.

Clay, M. M. (2002). *An observation survey of early literacy achievement*. Portsmouth: Heinemann.

Clay, M. M. (2003). Afterword. In S. Forbes & C. Briggs (Eds.), *Research in reading recovery* (Vol. 2). Portsmouth, NH: Heinemann.

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

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing* (pp. 151-216). London: Academic Press.

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

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

Connelly, V., Johnston, R., & Thompson, G. B. (2001). The effects of phonics instruction on the reading comprehension of beginning readers. *Reading and Writing, 14,* 423-457.

Conti-Ramsden, G., & Botting, N. (1999). Characteristics of children attending language units in England: A national study of 7-year-olds. *International Journal of Language and Communication Disorders, 34*(4), 359-366.

Conti-Ramsden, G., & Durkin, K. (2007). Phonological short-term memory, language and literacy: Developmental relationships in early adolescence in young people with SLI. *Journal of Child Psychology and Psychiatry, 48*(2), 147-156.

Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Philadelphia: Harcourt Brace Jovanovich College Publishers.

Croft, C., Strafford, E., & Mapa, L. (2000). *Stocktake/evaluation of existing diagnostic tools in literacy and numeracy, in English: A report to the Ministry of Education*. Wellington, New Zealand: NZCER.

Csapo, B., Molnar, G., & Toth, K. (2009). Comparing paper-and-pencil and online assessment of reasoning skills. A pilot study for introducing electronic testing in large-scale assessment in Hungary. In F. Scheuermann & J. Bjornsson (Eds.), *The transition to computer-based assessment. New approaches to skills assessment and implications for large-scale testing* (pp. 113-118). Luxembourg: Office for the Official Publications of the European Communities.

Dale, E., & O'Rourke, J. (1981). *The living word vocabulary*. Chicago: World Book-Childcraft International.

de Graaff, S., Hasselman, F., Bosman, A. M. T., & Verhoeven, L. (2007). Cognitive and linguistic constraints on phoneme isolation in Dutch kindergartners. *Learning and Instruction*. doi: 10.1016/j.learninstruc.2007.08.001

Denton, C. A., Ciancio, D., & Fletcher, J. (2006). Validity, reliability, and utility of the observation survey of early literacy achievement. *Reading Research Quarterly, 41*, 8-34.

Dewar, S., & Telford, M. (2003). *A study of the use of the School Entry Assessment (SEA) in schools*. Wellington, New Zealand: Research Division Ministry of Education.

Dickinson, D. K. (2006). Toward a toolkit approach to describing classroom quality. *Early Education and Development, 17*, 177-202.

Dodd, B. (1995). *Differential diagnosis and treatment of children with speech disorder*. London: Whurr.

Dodd, B., Crosbie, S., McIntosh, B., Teitzel, T., & Ozanne, A. (2000). *Preschool and primary inventory of phonological awareness (PIPA)*. London, UK: Psychological Corporation.

Dodd, B., & Gillon, G. (2001). Exploring the relationship between phonological awareness, speech impairment and literacy. *Advances in Speech Language Pathology, 3*(2), 139-147.

Dunn, L., & Dunn, M. (2007). *Peabody Picture Vocabulary Test - Fourth Edition (PPVT-4).* Minneapolis, MN: Pearson.

Eggen, T., & Straetmans, G. (2009). Computerised adaptive testing at the entrance of primary school teacher training college. In F. Scheuermann & J. Bjornsson (Eds.), *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing* (pp. 134-144). Luxemburg: Office for Official Publications of the European Communities.

Ehren, B. J., & Nelson, W. N. (2005). The responsiveness to intervention approach and language impairment. *Topics in Language Disorders, 25*(2), 120-131.

Ehri, L. (1991). Development of the ability to read words. In R. Barr, M. Kamil, P. Mosenthal & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 384-417). NY: Longman.

Ehri, L. C. (1992). Reconceptualizing the development of sight word reading and its relationship to recoding. In L. Gough & R. Treiman (Eds.), *Reading acquisition* (pp. 107-143). Hillsdale, NJ: Erlbaum.

Ehri, L. C. (1998). Grapheme-phoneme knowledge is essential for learning to read words in English. In J. Metsala & L. Ehri (Eds.), *Word recognition in beginning reading* (pp. 3-40). Hillsdale, NJ: Erlbaum.

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the national reading panel's meta-analysis. *Reading Research Quarterly, 36*(3), pp. 250-287.

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

Ehrler, D. J., & McGhee, R. L. (2008). *Primary Test of Nonverbal Intelligence (PTONI)*. Austin, TX: Pro-Ed.

Elbaum, B., Vaughn, S., Hughes, M. T., & Moody, S. W. (1999). Grouping practices and reading outcomes for students with disabilities. *Exceptional Children 65*, 399-415.

Elbro, C., Borstrom, I., & Peterson, D. K. (1998). Predicting dyslexia from kindergarten: The importance of distinctness of phonological representations of lexical items. *Reading Research Quarterly, 33*, 36-60.

Elbro, C., & Pallesen, B. R. (2002). The quality of phonological representations and phonological awareness: A causal link. In L. Verhoeven, C. Elbro & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 17-32). Amsterdam: John Benjamins.

Embretson, S., & Reise, S. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London, UK: Sage Publications Ltd.

Fletcher, J., Parkhill, F., & Gillon, G. (Eds.). (2010). *Motivating literacy learners in today's world*. Wellington, New Zealand: NZCER Press.

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2007). *Learning disabilities: From identification to intervention*. NY: Guilford.

Frith, U. (1985). Beneath the surface of developmental dyslexia. In K. E. Patterson, J. C. Marshall & M. Coltheart (Eds.), *Surface dyslexia* (pp. 301-330). Hillsdale, NJ: Erlbaum.

Fuchs, D., Fuchs, L., Thompson, A., Al Otaiba, S., Yen, L., Yang, N., . . . O'Connor, R. (2001). Is reading important in reading-readiness programs? A randomized field trial with teachers as program implementers. *Journal of Educational Psychology, 93*(2), 251-267.

Fuchs, D., Fuchs, L., Thompson, A., Al Otaiba, S., Yen, L., Yang, N., . . . O'Connor, R. (2002). Exploring the importance of reading programs for kindergartners with disabilities in mainstream classrooms. *Exceptional Children, 68*, 295-311.

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2004). Monitoring early reading development in first grade: Word identification fluency versus nonsense word fluency. *Exceptional Children, 71*, 7-21.

Gallagher, A. M., Laxon, V., Armstrong, E., & Frith, U. (1996). Phonological difficulties in high-functioning dyslexics. *Reading and Writing*, *8*, 499-509.

Gaskins, I. W. (2011). Interventions to Develop Decoding Proficiencies. In R. Allington & A. McGill-Franzen (Eds.), *Handbook of reading disability research*. NY: Taylor & Francis.

Gillon, G. (2009). *A global literate society begins with early reading success for all children*. Paper presented at the Dyslexia Conference, 9-10 January 2009. Wellington, New Zealand.

Gillon, G., & Dodd, B. (1997). Enhancing the phonological processing skills of children with specific reading disability. *European Journal of Disorders of Communication, 32*, 67-90.

Gillon, G. T. (2000a). The efficacy of phonological awareness intervention for children with spoken language impairment. *Language, Speech, and Hearing Services in Schools, 31*, 1126-141).

Gillon, G. T (2000b). *The Gillon Phonological Awareness Training Programme (PAT)* (2nd ed.). Christchurch, New Zealand: Canterprise, University of Canterbury. Available at: http://www.education.canterbury.ac.nz/people/gillon/gillon_phonological_awareness_training_programme.shtml

Gillon, G. T. (2002). Follow-up study investigating the benefits of phonological intervention for children with spoken language impairment. *International Journal of Language and Communication Disorders, 37*(4), 381-400.

Gillon, G. T. (2004). *Phonological awareness: From research to practice*. NY: Guilford Press.

Gillon, G. T. (2005). Facilitating phoneme awareness development in 3- and 4-year old children with speech impairment. *Language, Speech, and Hearing Services in Schools, 36*, 308-324.

Gillon, G. T., & McNeill, B. C. (2009). Effective practices in phonological awareness assessment and intervention. *ACQuiring Knowledge in Speech, Language, and Hearing, 11*(2), 72-76.

Gillon, G. T., & Schwarz, I. E. (2001). *An international literature review of best practices in speech and language therapy (2001 Update)*. Wellington, New Zealand: Ministry of Education.

Gilmore, A., Croft, C., & Reid, N. (1981). *Burt Word Reading Test - New Zealand Revision*. Wellington, New Zealand: NZCER Press.

Good, R. H., III, & Kaminski, R. A. (2003). *DIBELS: Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Longmont, CO: Sopris West.

Good, R. H., Simmons, D. C., & Kame' enui, E. J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third grade high-stakes outcomes. *Scientific Studies of Reading, 5,* 257-288.

Good, R. H., Simmons, D. C., Kame' enui, E. J., Kaminski, R. A., & Wallin, J. (2002). *Summary of decision rules for intensive, strategic, and benchmark instructional recommendations in kindergarten through third grade (Technical Report No. 11)*. Eugene: University of Oregon.

Goswami, U. (2000). Early phonological development and the acquisition of literacy. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of early literacy research* (pp. 111-125). NY: Guilford Press.

Goswami, U., & Bryant, P. (1990). *Phonological skills and learning to read*. Hove, UK: Erlbaum.

Gough, P. B. (1996). How children learn to read and why they fail. *Annals of Dyslexia*, *46,* 3-20.

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

Green, S. R. (2001). Closing the achievement gap: Lessons learned and challenges ahead. *Teaching and Change, 8*(2), 215-224.

Griffiths, A. J., Parson, L. B., Burns, M. K., Van Der Heyden, A., & Tilly, W. D. (2007). *Response to intervention: Research for practice*. Alexandria, VA: National Association of State Directors of Special Education.

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, *106*(3), 491-528.

Hatcher, P. J., & Hulme, C. (1999). Phonemes, rhymes, and intelligence as predictors of children's responsiveness to remedial reading instruction: Evidence from a longitudinal intervention study. *Journal of Experimental Child Psychology, 72,* 130-154.

Hatcher, P. J., Hulme, C., Miles, J. N. V., Carroll, J. M., Hatcher, J., Gibbs, S., . . . Snowling, M. J. (2006). Efficacy of small group reading intervention for beginning readers with reading-delay: A randomized controlled trial. *Journal of Child Psychology and Psychiatry, 47*, 820-827.

Hattie, J. (2005). The paradox of reducing class size and improving learning outcomes. *International Journal of Educational Research, 43*(6), 387-425.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.

Hay, J., Macglagan, M., & Gordon, E. (2008). *New Zealand English*. Edinburgh, UK: Edinburgh University Press.

Haycock, K. (2001). Closing the achievement gap. *Educational Leadership, 58*(6), 6-11.

Hedges, L. V., & Nowell, A. (1999). Changing the Black-White gap in achievement test scores. *Sociology of Education*, *72*(2), 111-135.

Hindson, B. A., Byrne, B., Fielding-Barnsley, R., Newman, C., Hine, D. W., & Shankweiler, D. (2005). Assessment and early instruction of preschool children at risk for reading disability. *Journal of Educational Psychology, 97*, 687-704.

Hoien, T., Lundberg, I., Stanovich, K. E., & Bjaalid, I. (1995). Components of phonological awareness. *Reading and Writing: An Interdisciplinary Journal, 7*, 171-188.

Hoover, G., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal, 2*, 127-160.

Hubbard, C., & Mahanna-Boden, S. (2002). *Phonological awareness: Assessment and training activities*. Paper presented at the Annual American Speech, Language and Hearing Association convention (ASHA), Louisville, KY.

Hulme, C., Hatcher, P. J., Nation, K., Brown, A., Adams, J., & Stuart, G. (2002). Phoneme awareness is a better predictor of early reading skill than onset-rime awareness. *Journal of Experimental Child Psychology, 82*, 2-28.

Invernizzi, M., & Hayes, L. (2011). Developmental Patterns of Reading Proficiency and Reading Difficulties. In R. Allington & A. McGill-Franzen (Eds.), *Handbook of Reading Disability Research*. NY: Taylor & Francis.

Invernizzi, M., Juel, C., Meier, J. D., & Swank, L. (2005). *Phonological awareness literacy screening - K (PALS-K)*. Charlottesville, VA: University of Virginia.

IPA. (1999). *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.

Jenkins, J. R. (2003). *Candidate measures for screening at-risk students*. Paper presented at the National Research Center on Learning Disabilities Responsiveness-to-Intervention symposium, Kansas City, MO. Retrieved 10th January 2012 from: http://www.nrcld.org/symposium2003/jenkins/index.html

Jeynes, W. H. (2007). A meta-analysis of the relationship between phonics instruction and minority elementary school student academic achievement. *Education in Urban Society*. Retrieved 7th February 2009 from: http://eus.sage.pub.com/cgi/content/abstract/40/2/51

Johnson, K. L., & Roseman, B. A. (2003). *The source for phonological awareness*. East Moline, IL: Linguisystems Inc.

Johnson, R., & Watson, J. (2005). *The effects of synthetic phonics teaching on reading and spelling attainment: A seven year longitudinal study*. Retrieved 6th October 2001 from: www.scotland.gov.uk/library5/education/ins17-00.asp

Justice, L. M. (2006). *Clinical approaches to emergent literacy intervention (Emergent and early literacy)*. San Diego, CA: Plural Publishing Inc.

Justice, L. M., Invernizzi, M. A., & Meier, J. D. (2002). Designing and implementing an early literacy screening protocol: Suggestions for the speech-language pathologist. *Language, Speech, and Hearing Services in Schools*, *33*, 84-101.

Justice, L. M., McGinty, A. S., Cabell, S. Q., Kilday, C. R., Knighton, K., & Huffman, G. (2010). Language and literacy curriculum supplement for preschoolers who are academically at risk: A feasibility study. *Language, Speech, and Hearing Services in Schools, 41*, 161-178.

Kamhi, A. G., & Catts, H. W. (2012). *Language and reading disabilities* (3rd ed.). Upper Saddle River, NJ: Pearson Education Inc.

Kamhi, A. G., Catts, H. W., & Mauer, D. (1990). Explaining speech production deficits in poor readers. *Journal of Learning Disabilities, 23*, 632-636.

Kane, M. (1992). An argument-based approach to validity. *Psychological Bulletin*, *112*, 527-535.

Kern, M. L., & Friedman, H. S. (2008). Early educational milestones as predictors of lifelong academic achievement, midlife adjustment, and longevity. *Journal of Applied Developmental Psychology, 30*(4), 419-430.

Kim, Y., & Goetz, E. (1994). Context effects on word recognition and reading comprehension of poor and good readers: A test of the interactive-compensatory hypothesis. *Reading Research Quarterly, 29*, 178-188.

Klee, T., & Tillard, G. (2010). *Evidence-based assessment of child language disorders*. Paper presented at the New Zealand Speech Language Therapists' Association Conference: A practice worth spreading, Wellington, New Zealand.

Konza, D. (2006). *Teaching children with reading difficulties* (2nd ed.). Melbourne: Thompson Learning.

Levine, T. R., & Hullett, C. R. (2002). Eta squared, partial eta squared, and misreporting of effect size in communication research. *Human Communication Research, 28*(4), 612-625.

Lewis, B. A., Freebairn, L. A., & Taylor, H. G. (2000). Follow-up of children with early expressive phonological disorders. *American Journal of Speech-Language Pathology, 2*, 38-43.

Liberman, I. Y., Liberman, A. M., Mattingly, I., & Shankweiler, D. (1980). Orthography and the beginning reader. In J. F. Kavanagh & R. L. Venezky (Eds.), *Orthograhpy, reading and dyslexia* (pp. 137-153). Baltimore: University Park Press.

Liberman, I. Y., Shankweiler, D., Fisher, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology, 18*, 201-212.

Linacre, M. (2010). A user's guide to Winsteps. Retrieved 6th December 2011 from: http://www.winsteps.com/winman/index.htm?guide.htm

Lindamood, C., & Lindamood, P. (1979). *Lindamood Auditory Conceptualization Test*. Austin, TX: DLM Teaching Resources.

Lissitz, R. W., & Samuelsen, K. (2007). A suggested change in terminology and emphasis regarding validity and education. *Educational Researcher, 36*, 437-448.

Lloyd, S. (1992). *The Jolly Phonics Handbook*. Essex, UK: Jolly Learning Ltd.

Long, S., Fey, M., & Channell, R. (2002). Profile of Phonology (PROPH). Computerized Profiling [Computer Software]. Retrieved 9th January 2009 from: www.computerizedprofiling.org.

Lonigan, C., Burgess, S. R., Anthony, J., & Barker, T. A. (1998). Development of phonological sensitivity in 2- to 5-year-old children. *Journal of Educational Psychology, 90*(2), 294-311.

Lonigan, C., Schatschneider, C., & Westberg, L. (2008). Results of the national early literacy panel research synthesis: Identification of children's skills and abilities linked to later outcomes in reading, writing, and spelling. In *Developing early literacy: Report of the National Early Literacy Panel* (pp. 55-106). Washington, DC: National Institute for Literacy.

Lonigan, C. J. (2006). Conceptualizing phonological processing skills in prereaders. In D. K. Dickinson & S. B. Neuman (Eds.), *Handbook of early literacy research* (Vol. 2, pp. 77-89). NY: Guilford Press.

Lukatela, K., Carello, C., Shankweiler, D., & Liberman, I. Y. (1995). Phonological awareness in illiterates: Observations from Serbo-Croatian. *Applied Psycholinguistics, 16*, 463-487.

Lundberg, I., Olofsson, A., & Wall, S. (1980). Reading and spelling skills in the first school years predicted from phonemic awareness skills in kindergarten. *Scandinavian Journal of Psychology, 21*, 159-173.

MacDonald, G. W., & Cornwall, A. (1995). The relationship between phonological awareness and reading and spelling achievement eleven years later. *Journal of Learning Disabilities, 28*, 523-527.

Maguire, K. B., Knobel, M. L. M., Knobel, B. L., & Sedlacek, L. G. (1991). Computer-adapted PPVT-R: A comparison between standard and modified versions within an elementary school population. *Psychology in the Schools*, *28*, 199-205.

Martin, M. O., Mullis, I. V. S., & Kennedy, A. M. (Eds.). (2007). *PIRLS 2006 technical report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Martin, R. (2008). New possibilities and challenges for assessment through the use of technology. In F. Scheuermann & A. G. Pereira (Eds.), *Towards a research agenda on computer-based assessment* (pp. 6-9). Luxembourg: Office for the Official Publications of the European Communities.

Mastriopieri, M. A., & Scruggs, T. E. (2005). Feasibility and consequences of response to intervention: Examination of the issues and scientific evidence as a model for the identification of individuals with learning disabilities. *Journal of Learning Disabilities, 38*, 525-531.

Mathes, P. G., Denton, C. A., Fletcher, J. M., Anthony, J., Francis, D. J., & Schatschneider, C. (2005). The effects of theoretically different instruction and student characteristics on the skills of struggling readers. *Reading Research Quarterly, 40*, 148-182.

Mayringer, H., & Wimmer, H. (2000). Pseudoname learning by German-speaking children with dyslexia: Evidence for a phonological learning deficit. *Journal of Experimental Child Psychology, 75*(2), 116-133.

McIntosh, B., Crosbie, S., Holm, A., Dodd, B., & Thomas, S. (2007). Enhancing the phonological awareness and language skills of socially disadvantaged preschoolers:

An interdisciplinary programme. *Child Language Teaching and Therapy, 23*(3), 267-286.

McLeod, J., Fisher, J., & Hoover, G. (2003). *The key elements of classroom management: Managing time and space, student behaviour and instructional strategies*. Alexandria, VA: ASCD.

McNeill, B. C., Gillon, G. T., & Dodd, B. (2009). Effectiveness of an integrated phonological awareness approach for children with childhood apraxia of speech (CAS). *Child Language Teaching and Therapy 25*(3), 341-366.

McTighe, J., & Ferrara, S. (1998). *Assessing learning in the classroom*. Washington, DC: National Education Association.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741-749.

Moats, L. C. (2000). *Whole language lives on: The illusion of 'Balanced Reading' instruction*. Washington, DC: Thomas, B. Fordham Foundation.

MOE. (2004). *New Zealand Articulation Test (NZAT)*. Wellington, New Zealand: Ministry of Education.

MOE. (2007). *Literacy learning progressions: Draft for consultation*. Wellington, New Zealand: Ministry of Education.

MOE. (2008). *Ngā Haeata Mātauranga - The Annual Report on Māori Education 2007/08*. Retrieved 11th March 2011 from: http://www.educationcounts.govt.nz/publications/series/5851/35307/8

MOE. (2010). *Education counts: School roll summary reports*. Retrieved 8th April 2011 from: www.educationcounts.govt.nz/publications/schooling/2259.

MOE. (2011). *School decile rating*. Retrieved 1st December 2011 from: http://minedu.govt.nz/Parents/AllAges/EducationInNZ/SchoolsInNewZealand/SchoolDecileRatings.aspx.

MOE. (2012a). *Ka Hikitia - Managing for success: The Māori Education Strategy 2008-2012*. Retrieved 11th April 2012 from: www.minedu.govt.nz/theMinistry/PolicyAndStrategy/KaHikitia.aspx

MOE. (2012b). *The New Zealand Education System: An Overview*. Auckland, New Zealand: New Zealand Ministry of Education.

Morais, J. (1991). Phonological awareness: A bridge between language and literacy. In D. Sawyer & B. Fox (Eds.), *Phonological awareness and reading acquisition* (pp. 31-71). NY: Springer-Verlag.

Morgan, P., Farakas, G., & Hibel, J. (2008). Matthew effects for whom? *Learning Disability Quarterly, 31*(4), 187-198.

Morton, J., & Patterson, K. (1980). A new attempt at an interpretation, or, an attempt at a new interpretation. In M. Coltheart, K. Patterson & J. Marshall (Eds.), *Deep dyslexia* (pp. 91-118). London: Routledge & Kegan Paul.

Muter, V., Hulme, C. K., Snowling, M., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology, 40*, 665-681.

Muter, V., Hulme, C.K., Snowling, M., & Taylor, S. (1997). Segmentation, not rhyming, predicts early progress in learning to read. *Journal of Experimental Child Psychology, 65*, 370-396.

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

NAEP. (2003). *The nation's report card.* Retrieved 22nd February 2011 from: http://nces.ed.gov/nationsreportcard

Nathan, L., Stackhouse, Goulandris, & Snowling, M. J. (2004). The development of early literacy skills among children with speech difficulties: A test of the 'Critical Age Hypothesis'. *Journal of Speech, Language, and Hearing Research, 42*(2), 337-391.

Neale, M. D. (1999). *Neale Analysis of Reading Ability* (3rd ed.). Victoria, Australia: ACER Press.

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 Disability Research & Practice, 18*, 255-267.

Nelson, W. N. (2010). *Language and literacy disorders: Infancy through adolescence*. Boston, MA: Pearson Education Inc.

Nicholson, T. (2000). *Reading the writing on the wall: Debates, challenges and opportunities in the teaching of reading*. Palmerston North, New Zealand: Dunmore Press.

Nicholson, T. (2003). Risk factors in learning to read. In B. Foorman (Ed.), *Preventing and remediating reading difficulties: Bringing science to scale* (pp. 165-193). Timonium, MD: York Press.

Nicholson, T. (2009). *Should beginner readers rely on context clues or not: The case for and against.* Paper presented at the Literacy Research Symposium, 1-2 October 2009, Christchurch, New Zealand.

O'Connor, R. E., & Jenkins, J. R. (1999). The prediction of reading disabilities in kindergarten and first grade. *Scientific Study of Reading, 3*, 59-197.

O'Connor, R.E., Notari-Syverson, A., & Vadasy, P. F. (1996). Ladders to Literacy: The effects of teacher-led phonological activities for kindergarten children with and without disabilities. *Exceptional Children, 63*, 117-130.

O'Connor, R. E., Notari-Syverson, A., & Vadasy, P. F. (2005). *Ladders to literacy: A kindergarten activity book* (2nd ed.). Baltimore: Brookes.

OECD. (2010). *PISA 2009 at a glance*. Retrieved 7th November 2001 from: http://idx.doi/org/10.1787/9789264095298-en

Olsen, J. B. (1990). Applying computerized adaptive testing in schools. *Measurement and Evaluation in Counselling and Development, 23*, 31-38.

Paul, R. (2007). *Language disorders from infancy through adolescence: Assessment & intervention* (3rd ed.). St Louis, MO: Mosby Inc.

Penno, J. F., Wilkinson, I. A., & Moore, D. W. (2002). Vocabulary acquisition from teacher explanation and repeated listening to stories: Do they overcome the Matthew effect? *Journal of Educational Psychology, 94*, 23-33.

Perfetti, C. A. (1985). *Reading ability*. NY: Oxford University Press.

Perfetti, C. A., Beck, I., Ball, L. C., & Hughes, C. (1987). Phonemic knowledge and learning to read are reciprocal: A longitudinal study of first grade children. *Merrill-Palmer Quarterly, 33*, 283-319.

Plaut, D. (2007). Connectionist approaches to reading. In M. J. Snowling & C. Hulme (Eds.), *The science of reading* (pp. 24-39). Malden, MA: Blackwells.

Portney, L. G., & Watkins, M. P. (2009). *Foundations of clinical research: Applications to practice* (3rd ed.). Upper Saddle River, NJ: Pearson Education Inc.

Pratt, A. C., & Brady, S. (1988). Relation of phonological awareness to reading disability in children and adults. *Journal of Educational Psychology, 80*(3), 319-323.

Pressley, M. (2006). *Reading instruction that works: The case for balanced teaching*. NY: Guilford Press.

Raitano, N. A., Pennington, B. F., Tunick, R. A., Boada, R., & Shriberg, L. D. (2004). Pre-literacy skills of a subgroup of children with speech sound disorders. *Journal of Child Psychology and Psychiatry, 45*(4), 821-835.

Reitsma, P. (2002). Precursors of phonemic awareness. In L. Verhoeven, C. Elbro & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 33-48). Amsterdam: John Benjamins.

Ripley, M. (2008). Technology in the service of 21st century learning and assessment: A UK perspective. In F. Scheuermann & A. G. Pereira (Eds.), *Towards a research agenda on computer-based assessment* (pp. 22-29). Luxembourg: Office for Official Publications of the European Communities.

Robertson, C., & Salter, W. (1995). *The Phonological Awareness Profile*. East Moline, IL: LinguiSystems.

Robertson, C., & Salter, W. (1997). *Phonological Awareness Test*. East Moline, IL: LinguiSystems.

Rodrigues, M. C. (2005). Three options are optimal for multiple-choice items: A meta-analysis of 80 years of research. *Educational Measurement: Issues and Practice, 24*(3), 3-13.

Rose, J. (2006). *Independent review of the teaching of early reading - Final report*. Retrieved 3rd March 2009 from: http://www.standards.dfes.gov.uk/phonics/report.pdf

Rumelhart, D. (1977). Toward an interactive model of reading. In S. Dornic (Ed.), *Attention and performance* (Vol. 6, pp. 573-603). Hillsdale, NJ: Erlbaum.

Rvachew, S., Ohberg, A., Grawberg, M., & Heyding, J. (2003). Phonological awareness and phonemic perception in 4-year-old children with delayed expressive phonology skills. *American Journal of Speech–Language Pathology, 12*, 463-471.

Scanlon, D. M., & Vellutino, F. R. (1996). Prerequisite skills, early instruction and success in first grade reading: Selected results from a longitudinal study. *Mental Retardation and Developmental Disabilities, 2*, 54-63.

Schonell, F. (1932). *Schonell Essential Spelling List*. Gloucestershire, UK: Thornes Ltd.

Schreiber, L. R. (2008). *Development of phonological awareness skills*. Retrieved 3rd February 2009 from: http://www.speechpathology.com/article_detail.asp?article_id=342#ceuarticle

Schreuder, R., & van Bon, W. H. J. (1989). Phonemic analysis: Effects of word properties. *Journal of Research in Reading, 12*, 59-78.

Schuele, C. M., & Boudreau, D. (2008). Phonological awareness intervention: Beyond the basics. *Speech, Language, and Hearing Services in Schools, 39*, 3-20.

Scruggs, T. E., & Mastriopieri, M. A. (2002). On babies and bathwater: Addressing the problems of identification of learning disabilities. *Learning Disability Quarterly, 25*, 155-159.

Seidenberg, M.S. (1995). Visual word recognition. In J.L. Miller and P.D. Eimas (Eds.) *Handbook of Perception & Cognition, Volume 11, Speech, Language & Communication.* San Diego: Academic Press.

Seidenberg, M., & McCelland, J. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*, 523-568.

Shapiro, L. R., & Solity, J. (2008). Delivering phonological and phonics training within whole-class teaching. *British Journal of Educational Psychology, 78,* 597-620.

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

Shriberg, L. D., Austin, D., Lewis, B. A., McSweeny, J. L., & Wilson, D. L. (1997). The speech disorders classification system (SDCS): Extensions and lifespan reference data. *Journal of Speech, Language, and Hearing Research*, *40*, 723-740.

Singleton, C. H., Horne, J. K., & Thomas, K. V. (1999). Computerised baseline assessment. *Journal of Reading Research, 22*(1), 67-80.

Singleton, C. H., Horne, J. K., & Vincent, D. (1995). *Implementation of a computer-based system for the diagnostic assessment of reading (Project Report to the National Council for Educational Technology)*. Hull: Department of Psychology, University of Hull.

Singleton, C. H., Thomas, K. V., & Leedale, R. C. (1996). *Lucid CoPS Cognitive Profiling System*. Beverley, East Yorkshire: Lucid Research Ltd.

Smith, J. W., & Elley, W. (1994). *Learning to read in New Zealand*. Auckland: Longman Paul.

Snider, V. E. (1995). A primer on phonemic awareness: What it is, why it's important, and how to teach it. *School Psychology Review, 24*, 443-455,

Snow, C. E., Burns, S., & Griffith, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Snow, C. E., & Juel, C. (2005). Teaching children to read: What do we know about how to do it? In M. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 501-520). Malden, MA: Blackwell.

Snowling, M. J., Bishop, D. V. M., & Stothard, S. E. (2000). Is preschool language impairment a risk factor for dyslexia in adolescence? *Journal of Child Psychology and Psychiatry, 41*, 587-600.

Sodoro, J., Allinder, R. M., & Rankin-Erickson, J. L. (2002). Assessment of phonological awareness: A review of methods and tools. *Educational Psychological Review, 14*, 223-260.

Spitalnic, S. (2004). Test properties I: Sensitivity, specificity, and predictive values. *Hospital Physician*, 27-31.

Stackhouse, J. (1982). An investigation of reading and spelling performance in speech disordered children. *British Journal of Disorders of Communication, 17,* 53-60.

Stage, S., Abbott, R. D., Jenkins, J., & Berninger, V. (2003). Predicting response to early reading intervention using verbal IQ-word reading related language abilities, attention ratings, and verbal IQ-word reading discrepancy: Failure to validate discrepancy. *Journal of Learning Disabilities*, *36*, 24-33.

Stahl, S. A., & Murray, B. A. (1994). Defining phonological awareness and its relationship to early reading. *Journal of Educational Psychology, 86*, 221-234.

Stanovich, K. E. (1980). Towards an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly, 16*, 32-71.

Stanovich, K. E. (1984). The interactive-compensatory model of reading: A confluence of developmental, experimental and educational psychology. *Remedial and Special Education, 5*(3), 11-19.

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

Stanovich, K. E. (1988). Explaining the differences between the dyslexic and the garden-variety poor reader: The phonological-core variable-difference model. *Journal of Learning Disabilities, 21*(10), 590-604.

Stanovich, K. E. (1992). Speculations on the causes and consequences of individual differences in early reading acquisition. In P. B. Gough, L. C. Ehri & R. Treiman (Eds.), *Reading acquisition* (pp. 307-342). Hillsdale, NJ: Erlbaum.

Stanovich, K. E. (1994). An exchange: Reconceptualizing intelligence: Dysrationalia as an intuition pump. *Educational Researcher, 23*(4), 11-22.

Stanovich, K. E., Cunningham, A. E., & Cramer, B. B. (1984). Assessing phonological awareness in kindergarten children: Issues of task comparability. *Journal of Experimental Child Psychology, 38,* 175-190.

Stanovich, K. E., Nathan, R. G., & Zolman, J. E. (1988). The developmental lag hypothesis in reading: Longitudinal and matched reading-level comparisons. *Child Development, 59,* 71-86.

Stanovich, K. E., Siegel, L. S., & Gottardo, A. (1997). Converging evidence for phonological and surface subtypes of reading disability. *Journal of Educational Psychology, 89,* 114-127.

Sterling-Orth, A. (2004). *Go-to guide for phonological awareness*. Eau Claire, WI: Thinking Publications.

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Hillsdale, NJ: Erlbaum.

Stuart, M., & Coltheart, M. (1988). Does reading develop in a sequence of stages? *Cognition, 30*, 139-181.

Thorndike, R. M., & Thorndike-Christ, T. (2010). *Measurement and evaluation in psychology and education* (8th ed.). Boston, MA: Pearson Education Inc.

Torgesen, J. K (1998). Catch them before they fall: Identification and assessment to prevent reading failure in young children. *American Educator*, *22*, 32-39.

Torgesen, J. K. (2004). Lessons learned from the last 20 years of research on interventions for students who experience difficulty learning to read. In P. McCardle & V. Chhabra (Eds.), *The voice of evidence in reading research*. Baltimore: Brookes Publishing.

Torgesen, J.K., & Bryant, B. (1994). *Test of Phonological Awareness*. Austin, TX: Pro-Ed.

Torgesen, J.K., Wagner, R., & Rashotte, C. (1994). Longitudinal studies of phonological processing and reading. *Journal of Learning Disabilities, 27,* 276-286.

Torgesen, J.K., Wagner, R., & Rashotte, C. (1997). The prevention and remediation of reading disabilities: Keeping the end in mind. *Scientific Studies of Reading, 1*, 217-234.

Torgesen, J. K., Rashotte, C. A., Alexander, A., Alexander, J., & MacPhee, K. (2003). Progress towards understanding the instructional conditions necessary for remediating reading difficulties in older children. In B. Foorman (Ed.), *Preventing and remediating reading difficulties: Bringing science to scale* (pp. 275-298). Parkton, MD: York Press.

Treiman, R., & Bourassa, D. C. (2000). The development of spelling skill. *Topics in Language Disorders, 20*(3), 1-18.

Treutlein, A., Zoller, I., Roos, J., & Scholer, H. (2008). Effects of phonological awareness training on reading achievement. *Written Language & Literacy, 11*(2), 147-166.

Troia, G. A. (1999). Phonological awareness intervention research: A critical review of the experimental methodology. *Reading Research Quarterly, 34*(1), 28-52.

Tunmer, W. E., & Chapman, J. W. (1999). Teaching strategies for word identification. In G. B. Thompson & T. Nicholson (Eds.), *Learning to read: Beyond phonics and whole language* (pp. 74-102). NY: Teachers College Press & International Reading Association.

Tunmer, W. E., & Chapman, J. W. (2002). The relation of beginning readers' reported word identification strategies to reading achievement, reading-related skills, and academic self-perceptions. *Reading and Writing, 15*, 341-358.

Tunmer, W. E., & Chapman, J. W. (2004). Reading Recovery: Distinguishing myth from reality. In R. M. Joshi (Ed.), *Dyslexia: Myths, misconceptions, and some practical applications* (pp. 99-114). Baltimore, MD: International Dyslexia Association.

Tunmer, W. E., Chapman, J. W., & Prochnow, J. E. (2002). Preventing negative matthew effects in at-risk readers: A retrospective study. Retrieved 9th February 2009 from: http://www.educationcounts.govt.nz/publications/curriculum/5803.

Tunmer, W. E., Chapman, J. W., & Prochnow, J. E. (2004). Why the reading achievement gap in New Zealand won't got away: Evidence from the PIRLS 2001 international study of reading achievement. *New Zealand Journal of Educational Studies*, *39*, 127-145.

Tunmer, W. E., Chapman, J. W., & Prochnow, J. E. (2006). Literate cultural capital at school-entry predicts later reading achievement: A seven-year longitudinal study. *New Zealand Journal of Educational Studies, 41*(2), 183-204.

Tymms, P. T. (1999). Baseline assessment, value-added and the prediction of reading. *Journal of Research in Reading, 22*(1), 27-36.

Tymms, P. T. (2001). The development of a computer-adaptive assessment in the early years. *Educational and Child Psychology, 18*(3), 20-30.

UNESCO. (2009). *The next generation of literacy statistics: Implementing the literacy assessment and monitoring programme (LAMP)*. Montreal, Canada: UNESCO Institute for Statistics.

UNICEF. (2010). *The children left behind: A league table of inequality in child well-being in the world's rich countries* (Vol. Innocenti Report Card 9). Florence: Innocenti Research Centre.

van Bon, W. H. J., & van Leeuwe, J. F. J. (2003). Assessing phonemic awareness in kindergarten: The case for the phonemic recognition task. *Applied Psycholinguistics, 24*, 195-219.

Vellutino, F. R., Scanlon, D. M., & Chen, R. S. (1995). The inextricable relationship between orthographic and phonological coding in learning to read: Some reservations about current methods of operationalizing orthographic coding. In V. W. Berninger (Ed.), *The varieties of orthographic knowledge:Relationships to phonology, reading, and writing* (Vol. 2, pp. 47-111). Norwell, MA: Kluwer Academic.

Vellutino, F. R., Scanlon, D. M., & Lyon, G. R. (2000). Differentiating between difficult-to-remediate and readily remediated poor readers: More evidence against the IQ - Achievement discrepancy definition of reading disability. *Journal of Learning Disabilities*, *33*(3), 223-238.

Vellutino, F. R., Scanlon, D. M., Small, S. G., & Fanuele, D. P. (2006). Response to intervention as a vehicle for distinguishing between children with and without reading disabilities: Evidence for the role of kindergarten and first grade interventions. *Journal of Learning Disabilities*, *39*, 157-169.

Vloedgraven, J. M. T., & Verhoeven, L. (2007). Screening of phonological awareness in the early elementary grades: An IRT approach. *Annals of Dyslexia*, *57*, 33-50.

Wagner, R., & Torgesen, J. (1987). The nature of phonological process and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101*, 192-212.

Wagner, R., Torgesen, J., & Rashotte, C. (1999). *Comprehensive Test of Phonological Processing. CTOPP*. Austin, TX: Pro-Ed.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1994). Development of reading-related phonological processing abilities: New evidence of bidirectional causality from a latent variable longitudinal study. *Developmental Psychology, 30*, 73-87.

Walker, B. J. (2008). History of Phonics Instruction. In M. J. Fresch (Ed.), *An essential history of current reading practices*. London, UK: International Reading Association.

Wang, S., Jiao, H., Young, M., Brooks, T., & Olsen, J. (2007). A meta-analysis of testing mode effects in grade K-12 mathematics tests. *Educational and Psychological Measurement*, *67*(2), 219-238.

Wang, S., Jiao, H., Young, M., Brooks, T., & Olsen, J. (2008). Comparability of computer-based and paper-and-pencil testing in K-12 reading assessment: A meta-anaysis of testing mode effects. *Education and Psychological Measurement, 68*(1), 5-24.

Wanzek, J., & Vaughn, S. (2007). Research-based implications from extensive early reading interventions. *School of Psychology Review, 36*, 541-561.

Webster, P., & Plante, A. (1992). Effects of phonological impairment on word, syllable and phoneme segmentation and reading. *Language, Speech, and Hearing Services in Schools, 23*, 176-182.

Webster, P., Plante, A., & Couvillion, L. (1997). Phonologic impairment and pre-reading: Update on a longitudinal study. *Journal of Learning Disabilities, 30*, 365-375.

Wiig, E., Secord, W., & Semel, E. (2006). *Clinical Evaluations of Language Fundamentals Preschool - Australian and New Zealand* (2nd ed.). San Antonio, TX: Psychological Corporation.

Yeh, S. S. (2003). An evaluation of two approaches for teaching phonemic awareness in children in head start. *Early Childhood Research Quarterly, 18*(4), 513-529.

Yopp, H. K. (1988). The validity and reliability of phonemic awareness tests. *Reading Research Quarterly, 23*, 159-177.

Yopp, H. K. (1995). A test for assessing phonemic awareness in young children. *The Reading Teacher, 49*(1), 20-29.

Zabell, C., & Everatt, J. (2002). Surface and phonological subtypes of adult developmental dyslexia. *Dyslexia, 8*(3), 160-177.

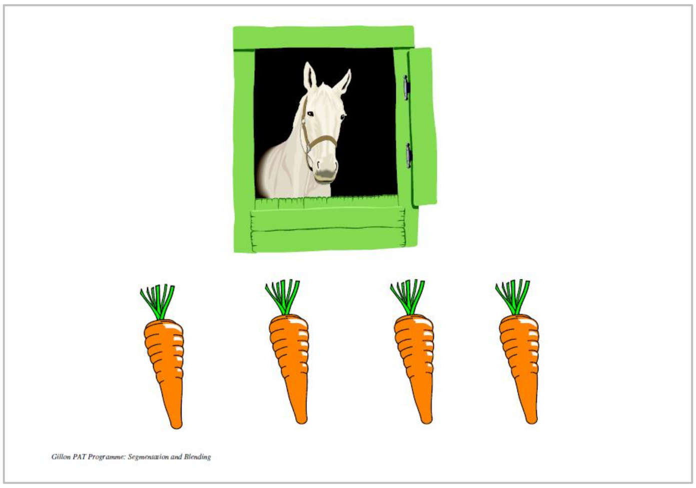
Ziegler, J., C, & Goswami, U. C. (2005). Reading acquisition, developmental dyslexia and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*, 3-29.

Appendix A: Classroom Phonological Awareness Programme Examples

This appendix provides three examples of classroom phonological awareness (PA) activities used in the experiments reported in Chapters 4 and 5. The classroom PA programme was based on the Gillon Phonological Awareness Training Programme (PAT) (Gillon, 2000b) and was modified for the classroom by drawing on classroom curriculum topics, enlarging resources and providing teachers with activity adaptation charts to address a wide range of PA abilities in the classroom. The original version of the PAT programme is available from:

http://www.education.canterbury.ac.nz/people/gillon/gillon_phonological_awareness_training_programme.shtml

Classroom PA Activity: *Phoneme Segmentation*



Procedure: A segmentation sheet is placed on the white board in front of the children. The teacher holds a small collection of coloured tokens. The following instructions, modified from original PAT programme, are provided:

> '……We have been reading books and learning a lot about animals this week, including farm animals. One animal we can find on a farm is a horse. Here is a picture of a horse. He'd like some carrots to eat. I'll say a word and I want you to show me how many sounds are in the word. We'll give the horse one carrot for each sound we hear. This horse lives in a barn. How many sounds can you hear in the word barn? *B-ar-n* (segmenting the word).' (p. 18, Gillon, 2000b)

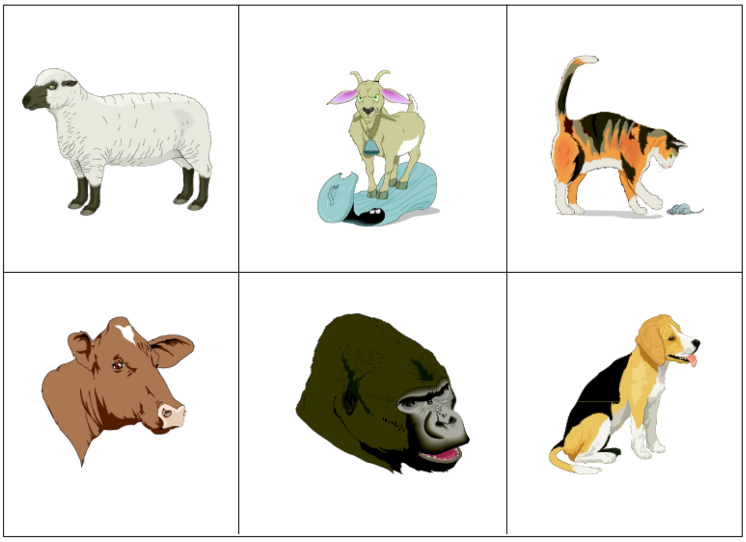
The teacher places one coloured token (with velcro attached) per sound onto one carrot below the horse. Each token represents one sound in the word *'Barn'*. The teacher discussed the number of sounds in the target word *'Barn'*. The teacher also writes the word 'Barn' on the white board to help make the link between speech and print. At least ten target words are presented per activity.

Activity Adaptations: To adapt this activity for a wide range of abilities in the classroom, experimental teachers were provided with an activity adaption chart (see following page). This chart lists the suggested target words for the activity (i.e., words from the original PAT programme and some added for the purposes of this thesis) and the types of questions teachers can ask to elicit PA knowledge for children of different ability levels. For example, the teacher can ask the child who is learning to identify the first sound in words, 'What is the first sound in *Barn*?' Likewise, the teacher can ask the child who is learning to segment sounds in words to, 'Tell me the sounds you hear in the word *Barn*'.

Activity Adaptation Chart: *Phoneme Segmentation*

	Phonological Awareness Stage				
	Easier ⟵			⟶	Harder
Child's level	**Identifying the first sound**	**Identifying the last sound**	**Blending sounds together to form a word**	**Breaking a word into sounds**	**Manipulating sounds to form a new word** (with blocks and letter tiles)
Instructions	*What is the first sound in ______?*	*What is the last sound in _____?*	*Guess what word I am saying?*	*Break up all the sounds in _____.*	*If this says _____, show me______.*
VC					
eat	/ea/	/t/	ea-t	eat (ea-t)	eat = meat, seat, seap, sop
CV					
hay	/h/	/ay/	h-ay	hay (h-ay)	hay = bay, may, kay
fur	/f/	/ur/	f-ur	fur (f-ur)	fur = tur, turp, purp
shoe	/sh/	/oe/	sh-oe	shoe (sh-oe)	shoe = shu, shup, pup, pups
CVC					
barn	/b/	/n/	b-ar-n	barn (b-ar-n)	barn = parn, parm, farm, arm
nose	/n/	/z/	n-o-se	nose (n-o-se)	nose = hose, pose, poke, moke, smoke
food	/f/	/d/	f-oo-d	food (f-oo-d)	food = zood, zoop, loop, koop, kup
hoof	/h/	/f/	h-oo-f	hoof (h-oo-f)	hoof = hoob, soob, koob, koop
horse	/h/	/s/	h-or-se	horse (h-or-se)	horse = dorse, porse, port, porb
run	/r/	/n/	r-u-n	run (r-u-n)	run = sun, bun, bunt, bunp
rein	/r/	/n/	r-ei-n	rein (r-ei-n)	rein = tein, mein, lein, leip
farm	/f/	/m/	f-ar-m	farm (f-ar-m)	farm = tarm, parm, part, parts
race	/r/	/s/	r-a-ce	race (r-a-ce)	race = rake, rate, mate, make
CCVC					
trot	/t/	/t/	t-r-o-t	trot (t-r-o-t)	trot = trop, trops, rops, grops
grass	/g/	/s/	g-r-a-ss	grass (g-r-a-ss)	grass = trass, rass, pass, pat
CCVCC					
drink	/d/	/k/	d-r-i-n-k	drink (d-r-i-n-k)	drink = rink, mink, minks, manks

Classroom PA Activity: *Phoneme Blending Bingo*



Gillon PAT Programme: Phoneme Segmentation Bingo

Procedure: A phoneme blending bingo sheet is placed on the white board in front of the children. The teacher holds six bingo word cards to be matched to the bingo board. As the teacher says the name of each bingo word (in a segmented manner), the children are required to blend each word back together. The following instructions, modified from the original PAT programme, are provided:

> '……This week we have been learning a lot about different types of animals. Some animals live in the zoo, some animals live at home and some animals live in the wild. I have some pictures of different types of animals. I'm going to say one animal name very slowly. See if you can guess the animal: c-a-t.' (p. 20, Gillon, 2000b)

Once the children blend a word correctly, a child with the correct answer attaches the word bingo card (with velcro) to the bingo board. The teacher writes the animal name on the board to draw children's attention to the link between speech and print. The teacher can adapt the difficulty by using the activity adaption chart for this task (see the following pages). At least two bingo boards are used to ensure children are exposed to at least twelve phoneme blending opportunities.

Activity Adaptations: The following activity adaption chart was provided to experimental teachers to help adapt this activity for a spectrum of PA abilities in the classroom. This chart lists the suggested target words for the activity (i.e., some words were taken from the original PAT programme and some have been added for the purposes of this thesis) and the types of questions that can be asked to elicit PA knowledge for children of different ability levels. For example, the teacher can ask the child who is learning to identifying the final sound in words, 'What is the last sound in the word *Cat*?' Similarly, the teacher can ask the child who is learning to segment words into sounds to, 'Tell me the sounds you hear in the word *Cat*?'

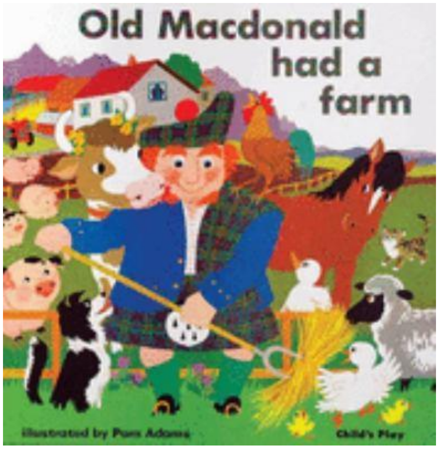
Activity Adaptation Chart: *Phoneme Blending Bingo*

	Phonological Awareness Stage				
	Easier ⟷				**Harder**
Child's level	**Identifying the first sound**	**Identifying the last sound**	**Blending sounds together to form a word**	**Breaking a word into sounds**	**Manipulating sounds to form a new word** (with blocks and letter tiles)
Instructions	*What is the first sound in ______?*	*What is the last sound in _____?*	*Guess what word I am saying?*	*Break up all the sounds in _____.*	*If this says _____, show me______.*
VC					
ape	/a/	/p/	a-pe	ape (a-pe)	ape = mape, cape, cake, make
CV					
cow	/c/	/ow/	c-ow	cow (c-ow)	cow = pow, zow, zoo, zoop
CVC					
sheep	/sh/	/p/	sh-ee-p	sheep (sh-ee-p)	sheep = leep, meep, meet, keet
goat	/g/	/t/	g-oa-t	goat (g-oa-t)	goat = boat, moat, moap, moat, mut
cat	/c/	/t/	c-a-t	cat (c-a-t)	cat = mat, sat, sap, fap, fat
dog	/d/	/g/	d-o-g	dog (d-o-g)	dog = pog, pot, mot, sot, sut

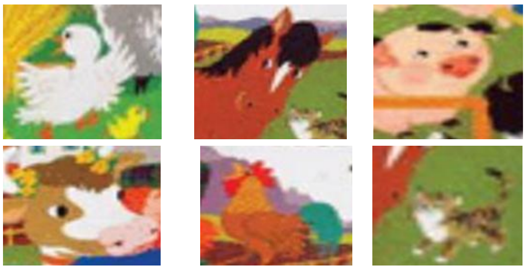
Classroom PA Activity: *Large book activity and songs*

Teachers were encouraged to select and adapt existing classroom resources to target PA skills. In particular, teachers were encouraged to use large books to help illustrate the link between speech and print. The first experimental teacher (i.e., Group A) selected the book 'Old McDonald' and the accompanying song to target phoneme blending and segmentation skills. This activity was taught to the second experimental teacher in Group B.

Book



Song cards (examples)



Procedure: The teacher reads aloud the story of 'Old McDonald' to the entire classroom and purposefully stops at each animal name to ask the children, 'How many sounds do you hear in the word: duck/horse/pig/hen/cow/cat/sheep/goat/bird/snake/tree/truck/nest?' The teacher encourages children to segment the sounds in the animal name. An activity adaption chart is used to scaffold the segmentation attempts of children with differing PA abilities. Alternatively, the teacher can engage children in singing the 'Old McDonald' song using animal picture cards. For each animal name in the song, the teacher asks children to, 'Guess what animal name I am saying? *C-a-t*.' The children are required to blend the individual sounds together to identify the animal name. At least ten opportunities are provided to segment and blend the phonemes in words.

Activity Adaptations: The following activity adaption chart was used to adjust task difficulty for the large book or singing activity. This chart lists the potential target words and the types of questions that can be asked to elicit PA knowledge for children of different ability levels.

Activity Adaptation Chart: *Large book or singing activity (phoneme blending and segmentation)*

	Phonological Awareness Stage				
	Easier ⟵			⟶	**Harder**
Child's level	**Identifying the first sound**	**Identifying the last sound**	**Blending sounds together to form a word**	**Breaking a word into sounds**	**Manipulating sounds to form a new word**
Instructions	*What is the first sound in ______?*	*What is the last sound in _____?*	*Guess what word I am saying?*	*Break up all the sounds in _____.*	*If this says _____, show me______.*
CV					
cow	/c/	/ow/	c-ow	cow (c-ow)	cow = pow, zow, zoo, zoop
CVC					
sheep	/sh/	/p/	sh-ee-p	sheep (sh-ee-p)	sheep = leep, meep, meet, keet
goat	/g/	/t/	g-oa-t	goat (g-oa-t)	goat = boat, moat, moap, moat, mut
cat	/c/	/t/	c-a-t	cat (c-a-t)	cat = mat, sat, sap, fap, fep
dog	/d/	/g/	d-o-g	dog (d-o-g)	dog = pog, pot, mot, kot, kots
duck	/d/	/k/	d-u-ck	duck (d-u-ck)	duck = muck, muz, puz, pez
horse	/h/	/s/	h-or-se	horse (h-or-se)	horse = dorse, porse, port, porb
pig	/p/	/g/	p-i-g	pig (p-i-g)	pig = pag, bag, bap, zap, zaps
hen	/h/	/n/	h-e-n	hen (h-e-n)	hen = pen, pet, get, gez
bird	/b/	/d/	b-ir-d	bird (b-ir-d)	bird = birg, girg, zirg, kirg
CCV(C)					
snake	/s/	/k/	s-n-a-ke	snake (s-n-a-ke)	snake = nake, make, smake, smate
tree	/t/	/ee/	t-r-ee	tree (t-r-ee)	tree = ree, kee, skee, snee, sneeze
truck	/t/	/k/	t-r-u-ck	truck (t-r-u-ck)	truck = druck, ruck, muck, muz
CVCC					
nest	/n/	/t/	n-e-s-t	nest (n-e-s-t)	nest = pest, pests, bests, best

Classroom PA Activity: *Tracking speech sounds with letter cards*

Procedure: Each child is provided with a large letter card that contains either a consonant sound or a vowel sound that is familiar to the child. The teacher has a list of simple three to four phoneme words selected from classroom curriculum activities or from shared classroom book activities. The teacher says a word and children with the corresponding letter cards are required to come to the front of the class and stand with their letter cards in front of them to form the word. The following instructions, modified from the original PAT programme, are provided.

> '… this week we have been learning a lot about different animals – at home, in the zoo, and in the wild. Let's see if we can spell the names of some of these animals using the letters we have in front of us. The first name is '*cat'*. Let's say all the sounds in the word '*cat* /c – a – t/' (segmenting the word). Put your hand up if you think you have a letter that makes one of the sounds in the word '*cat*'.

Children with a letter in the target word are encouraged to come up to the front of the classroom and form a line holding each letter in the correct order to formulate the word. The teacher then says:

> ‘ …. together these three letters say '*cat* /c-a-t/' (segmenting the word). If this says, *'cat'* lets spell the word *'mat'*. What sound and letter needs to change?

The child with the correct letter (i.e., m) changes places with the child whose letter was removed from the sequence (i.e., c). The game continues until approximately 10 sound changes have taken place.

Activity Adaptations: The activity is adapted to meet a range of PA abilities by providing children with letters that they are familiar with or providing children with letters they are having difficulty using in classroom reading and spelling tasks. Children who are learning to identify the first sound in words can be provided with letters that will be involved in initial sound changes to target words. Similarly, children who are learning to identify the final sound in words can be provided with letters that will be used in final sound changes. Children with more advanced skills can be provided with long and short vowels in addition to consonant digraphs (e.g., ck, th) as part of the manipulation game. Children with more advanced skills can also be encouraged to provide a sound change to the existing sequence using the phrase, 'If this says ___, show me _____'.

Appendix B: Performance in Phonological Awareness and Reading Outcomes in the First Year of School

Mean (SD) performance on the computer-based PA assessment and literacy measures in the first year of school for Groups A (n=18), B (n =16) and C (n=95)

	School-Entry (5;0-5;2)*			Middle of Year (5;5-5;7)*			End of Year (5;11-6;01)*		
	Group A	**Group B**	**Group C**	**Group A**	**Group B**	**Group C**	**Group A**	**Group B**	**Group C**
Phonological Awareness**									
- RO	5.39 (2.40)	5.69 (2.41)	5.26 (2.46)	8.28 (1.36)	8.31 (1.30)	7.46 (2.50)	9.44 (0.92)	9.50 (0.97)	8.57 (1.84)
- IPI	5.44 (2.20)	4.81 (1.83)	5.29 (3.03)	9.67 (0.49)	6.94 (2.21)	7.55(2.64)	9.94 (0.24)	9.69 (0.48)	8.47(2.27)
- FPI	0.34 (0.51)	0.51 (0 .71)	0.61 (1.71)	8.67(1.50)	4.81 (2.26)	4.75(2.86)	9.00 (1.37)	8.75 (1.57)	6.35(3.12)
- PB	1.32 (1.10)	1.61 (1.21)	1.31 (1.22)	13.17 (1.98)	9.57 (2.84)	8.43 (5.23)	14.56 (1.04)	14.44 (.96)	11.69 (3.74)
- PD	0.32 (0.52)	0.51 (0.52)	0.53 (0.81)	11.11 (3.20)	7.00 (2.42)	6.16 (4.19)	13.06 (2.13)	12.38 (1.09)	8.76 (3.48)
- PS	0.33 (0.49)	0.44 (0.51)	0.63 (0.88)	11.50 (1.86)	6.25 (2.65)	5.88 (4.11)	13.33 (1.97)	12.19 (1.17)	7.99 (3.96)
Reading									
NARA - Fluency							21.50 (6.98)	21.88 (7.22)	11.33 (5.26)
NARA - Comprehension							7.06 (2.49)	7.38 (2.78)	3.43 (1.97)

Note. * Age range in (years;months); ** Measures from the computer-based PA screening and monitoring tool described in Chapter 2 and used in Chapters 3, 4, 5 and 6; RO = rhyme oddity, IPI = initial phoneme identity, FPI = final phoneme identity, PB = phoneme blending, PD = phoneme deletion, PS = phoneme segmentation; NARA = Neale Analysis of Reading Ability (Neale, 1999).

Appendix C: Tables of Test Item 'Fit' and Significant 'Misfit' Using Rasch Modal Analysis

The following tables present the mean-square statistic (MNSQ), ZSTD statistic and point-measure correlation for each test item in the computer-based phonological awareness (PA) assessment at the start, middle and end of the first year at school. Test items with a mean-square statistic greater than 2.0, and a ZSTD statistic less than -2 and greater than +2 demonstrated a statistically significant model 'misfit' (Linacre, 2010). Possible explanations as to why certain items demonstrated a significant model 'misfit' are discussed. All practice items demonstrated a significant model 'fit' and thus the focus here is on test items. Grey shading in each table represents the correct multiple-choice response while pink shading represents items that demonstrated a significant modal 'misfit'.

Rhyme Oddity

Rhyme oddity test items by 'fit' or 'misfit' at school-entry

Test Items				Outfit Statistics			Interpretation
Which word does not rhyme?				MNSQ	ZSTD	PT-measure correlation	'Fit' or 'Misfit'
1	cat	mat	bus	4.98	2.81	0.29	Misfit
2	peg	doll	leg	2.72	3.01	0.29	Misfit
3	saw	toe	bow	1.27	0.72	0.66	Fit
4	sand	hand	cup	0.63	-0.61	0.77	Fit
5	hen	car	pen	0.36	-0.81	0.51	Fit
6	dog	book	hook	0.37	-0.92	0.62	Fit
7	bun	sun	kite	0.82	-0.13	0.52	Fit
8	tent	lock	sock	1.66	1.53	0.62	Fit
9	shell	duck	bell	0.33	-1.01	0.58	Fit
10	ring	sing	lamb	0.36	-0.91	0.59	Fit

Rhyme oddity test items by 'fit' or 'misfit' by the middle of the school year

Test Items Which word does not rhyme?				Outfit Statistics MNSQ	ZSTD	PT-measure correlation	Interpretation 'Fit' or 'Misfit'
1	cat	mat	bus	4.90	2.61	0.28	Misfit
2	peg	doll	leg	2.70	2.93	0.29	Misfit
3	saw	toe	bow	1.26	0.84	0.65	Fit
4	sand	hand	cup	0.62	-0.71	0.78	Fit
5	hen	car	pen	0.35	-0.74	0.41	Fit
6	dog	book	hook	0.35	-0.10	0.60	Fit
7	bun	sun	kite	0.83	-0.26	0.55	Fit
8	tent	lock	sock	1.70	1.64	0.60	Fit
9	shell	duck	bell	0.32	-1.03	0.60	Fit
10	ring	sing	lamb	0.38	-0.96	0.61	Fit

Rhyme oddity test items by 'fit' or 'misfit' by the end of the school year

Test Items Which word does not rhyme?				Outfit Statistics MNSQ	ZSTD	PT-measure correlation	Interpretation 'Fit' or 'Misfit'
1	cat	mat	bus	9.90	4.81	0.69	Misfit
2	peg	doll	leg	9.90	7.42	0.29	Misfit
3	saw	toe	bow	0.23	-1.01	0.60	Fit
4	sand	hand	cup	9.90	6.23	0.64	Misfit
5	hen	car	pen	0.50	-0.44	0.78	Fit
6	dog	book	hook	1.01	0	0.69	Fit
7	bun	sun	kite	0.16	-1.51	0.85	Fit
8	tent	lock	sock	0.26	-1.32	0.83	Fit
9	shell	duck	bell	0.36	-1.01	0.75	Fit
10	ring	sing	lamb	0.25	-1.13	0.64	Fit

In the rhyme oddity task, items 1 and 2 demonstrated a significant ‘misfit’ at the start, middle and end of the school year. Item 4 demonstrated a significant ‘misfit’ at the end of the school year. Explanations as to why these items demonstrated a significant ‘misfit’ is offered below:

- **Rhyme Oddity Item 1:** Inspection of responses to item 1 showed that 87 per cent, 100 per cent and 100 per cent of participants responded to this item correctly at the start, middle and end of the school year respectively. This test item asked children, 'Which word does not rhyme? /cat, mat, bus/'. This item was designed to be one of the easiest rhyme oddity test item so as to encourage success at the start of the assessment. Increased ease was achieved by making the correct option as salient as possible by 1) contrasting a high frequency rime unit (e.g., ‘at’) with a low frequency rime unit (e.g., ‘us’) and, 2) contrasting a short ‘stop’ sound (e.g., ‘t’) with a long ‘continuant’ sound (e.g., ‘s’). Although Rasch analysis indicates that this item is easy for the ability of five-year-old children, inclusion of easier items can help to foster initial success as children begin the computer-based assessment. It may also indicate that children who respond incorrectly to such an item may 1) not understand the task and response format, or 2) be likely to exhibit poor PA performance throughout the task and assessment.
- **Rhyme Oddity Item 2:** Inspection of responses to item 2 demonstrated that this item was of average difficulty and thus the ‘misfit’ cannot be attributed to the item being too easy or too difficult. A key difference between item 2 and the remaining rhyme oddity test items is that one of the distractor options and the correct option had one phoneme in common: ‘doll’ (correct option) and ‘leg’ (distractor option) both contain the /l/ phoneme. In no other rhyme oddity test item did the distractor option contain an identical phoneme to the correct option. This distractor option (i.e., 'leg') may have

been too strong and may therefore offer an explanation as to why item 2 showed a 'misfit'. It is also possible that image quality played a role in the item 'misfit'. For example, the image of a 'leg' was of a leg only, as opposed to being in the context of the body of a person. Further, it may have been too visually similar in length and diameter to the image of the 'peg'. Further inspection of this and similar test items in the future is necessary.

- **Rhyme Oddity Item 4:** Item 4 showed a model 'fit' at the start and middle of the school year and a significant model 'misfit' at the end of the school year. Inspection of responses to item 4 at the end of the school year showed that 98 per cent (93/95) of participants scored correctly on this item suggesting that it had become increasingly easier over time. This item asked children, 'Which word does not rhyme? /sand, hand, cup/'. The correct option in this item was made salient by contrasting a CVCC syllable structure among the two distractor options with a CVC syllable structure in the correct option which may have reduced the difficulty of this item by six years of age.

Initial Phoneme Identity

Initial phoneme identity test items by 'fit' or 'misfit' at school-entry

Test Items					Outfit Statistics			Interpretation
What word starts with the ___ sound?					MNSQ	ZSTD	PT-measure correlation	'Fit' or 'Misfit'
1	/m/	doll	bear	milk	1.90	1.11	0.49	Fit
2	sound?	mat	dog	book	0.19	-1.23	0.52	Fit
3	/s/	bee	sun	tent	0.40	-0.82	0.69	Fit
4	sound?	saw	tie	hook	1.88	1.61	0.78	Fit
5	/k/	bus	kite	arm	1.25	0.62	0.69	Fit
6	sound?	comb	dish	soap	0.21	-1.25	0.65	Fit
7	/b/	cat	leg	ball	0.25	-0.96	0.83	Fit
8	sound?	car	boat	shoe	0.59	-0.36	0.81	Fit
9	/f/	foot	hat	pig	0.11	-1.55	0.72	Fit
10	sound?	duck	bell	fire	0.88	0.12	0.76	Fit

Initial phoneme identity test items by 'fit' or 'misfit' by the middle of the school year

Test Items					Outfit Statistics			Interpretation
What word starts with the ___ sound?					MNSQ	ZSTD	PT-measure correlation	'Fit' or 'Misfit'
1	/m/	doll	bear	milk	3.67	1.72	0.17	Fit
2	sound?	mat	dog	book	2.80	1.43	0.30	Fit
3	/s/	bee	sun	tent	0.41	-1.11	0.73	Fit
4	sound?	saw	tie	hook	0.45	-1.62	0.78	Fit
5	/k/	bus	kite	arm	0.90	0.21	0.57	Fit
6	sound?	comb	dish	soap	1.27	0.61	0.62	Fit
7	/b/	cat	leg	ball	0.49	-1.52	0.78	Fit
8	sound?	car	boat	shoe	0.58	-1.97	0.81	Fit
9	/f/	foot	hat	pig	0.50	-0.91	0.80	Fit
10	sound?	duck	bell	fire	0.90	0.01	0.71	Fit

Initial phoneme identity test items by 'fit' or 'misfit' by the end of the school year

Test Items:					Outfit Statistics			Interpretation
What word starts with the ___ sound?					MNSQ	ZSTD	PT-measure correlation	'Fit' or 'Misfit'
1	/m/	doll	bear	milk	0.41	1.01	0.67	Fit
2	sound?	mat	dog	book	0.51	-0.32	0.33	Fit
3	/s/	bee	sun	tent	1.91	1.61	0.43	Fit
4	sound?	saw	tie	hook	1.46	0.91	0.58	Fit
5	/k/	bus	kite	arm	1.05	0.33	0.67	Fit
6	sound?	comb	dish	soap	0.63	-1.12	0.79	Fit
7	/b/	cat	leg	ball	0.97	0.01	0.77	Fit
8	sound?	car	boat	shoe	0.68	-0.92	0.82	Fit
9	/f/	foot	hat	pig	0.65	-0.71	0.81	Fit
10	sound?	duck	bell	fire	0.81	-0.21	0.77	Fit

In the phoneme identity task, all items demonstrated a model 'fit' at the start, middle and end of the first year at school. This is because all test items produced a mean-square statistic less than 2.0 and a ZSTD statistic between -2.0 and +2.0.

Final Phoneme Identity

Final phoneme identity test items by 'fit' or 'misfit' at school-entry

Test Items				Outfit Statistics			Interpretation
What word ends with the ___ sound?				MNSQ	ZSTD	PT-measure correlation	'Fit' or 'Misfit'
1	room	seal	soup	0.77	-0.50	0.82	Fit
2	hat	hole	sun	2.35	-3.20	0.71	Misfit
3	rope	rice	pan	0.58	1.97	0.50	Fit
4	food	fan	hook	1.01	0	0.34	Fit
5	green	grass	fire	1.02	0	0.36	Fit
6	hand	horse	milk	1.01	0	0.39	Fit
7	ball	goat	bank	1.01	0	0.36	Fit
8	cake	camp	bed	1.01	0	0.34	Fit
9	world	walk	face	1.01	0	0.33	Fit
10	tent	toast	map	1.01	0	0.32	Fit

Final phoneme identity test items by 'fit' or 'misfit' by the middle of the school year

Test Items				Outfit Statistics			Interpretation
What word ends with the ___ sound?				MNSQ	ZSTD	PT-measure correlation	'Fit' or 'Misfit'
1	room	seal	soup	1.90	1.01	0.82	Fit
2	hat	hole	sun	1.84	1.01	0.83	Fit
3	rope	rice	pan	0.49	-0.32	0.87	Fit
4	food	fan	hook	1.04	0.20	0.62	Fit
5	green	grass	fire	1.39	0.90	0.34	Fit
6	hand	horse	milk	0.69	-0.51	0.50	Fit
7	ball	goat	bank	1.22	0.61	0.48	Fit
8	cake	camp	bed	0.48	-1.81	0.76	Fit
9	world	walk	face	0.52	-1.10	0.62	Fit
10	tent	toast	map	0.48	-0.80	0.46	Fit

Final phoneme identity test items by 'fit' or 'misfit' by the end of the school year

Test Items				Outfit Statistics			Interpretation
What word ends with the ___ sound?				MNSQ	ZSTD	PT-measure correlation	'Fit' or 'Misfit'
1	room	seal	soup	0.50	-1.30	0.84	Fit
2	hat	hole	sun	0.31	-1.10	0.68	Fit
3	rope	rice	pan	0.20	-1.10	0.87	Fit
4	food	fan	hook	8.63	3.20	0.81	Misfit
5	green	grass	fire	0.58	-0.81	0.74	Fit
6	hand	horse	milk	0.94	0.01	0.71	Fit
7	ball	goat	bank	0.96	0.10	0.76	Fit
8	cake	camp	bed	0.79	-0.22	0.73	Fit
9	world	walk	face	0.65	-0.41	0.53	Fit
10	tent	toast	map	0.55	-0.61	0.55	Fit

In the final phoneme identity task, item 2 demonstrated a significant 'misfit' at the start of the school year and item 4 demonstrated a significant 'misfit' at the end of the school year. Explanations as to why these items demonstrated a significant 'misfit' is offered below:

- **Final Phoneme Deletion Item 2:** Inspection of responses to item 2 at the start of the school year showed that only seven per cent (7/95) of participants responded correctly to this test item. According to the 'Measure' statistic, this item was the most difficult out of the ten final phoneme identity items at school-entry. By the middle and end of the school year this item demonstrated a model 'fit'. Thus, extreme difficulty may provide one possible explanation as to why this item showed a significant model 'misfit' at school-entry.
- **Final Phoneme Deletion Item 4:** Inspection of responses to item 4 at the end of the school year showed that over 90 per cent of participants responded correctly to this test item, suggesting that it had become a very easy item by this stage of schooling.

The 'Measure' statistic for this item, in comparison to other items in this task, confirmed this assumption (see Appendix D).

Phoneme Blending

Phoneme blending test items by 'fit' or 'misfit' at school-entry

Test Items				Outfit Statistics			Interpretation
Click on the picture that you think I am saying?				MNSQ	ZSTD	PT-measure correlation	'Fit' or 'Misfit'
1	dog	dot	man	1.33	1.00	0.68	Fit
2	mouse	mouth	ring	0.73	-1.11	0.77	Fit
3	seal	seat	duck	1.17	0.70	0.58	Fit
4	bug	sun	bun	0.63	-0.42	0.41	Fit
5	cat	cap	lock	0.68	-0.21	0.38	Fit
6	flip	drum	flag	1.00	0	0.41	Fit
7	crab	crane	snake	1.00	0	0.32	Fit
8	bread	spade	space	1.00	0	0.35	Fit
9	tray	clown	train	1.00	0	0.36	Fit
10	stop	star	plane	1.00	0	0.34	Fit
11	point	fast	pond	1.00	0	0.35	Fit
12	bank	band	toast	1.00	0	0.42	Fit
13	desk	lamp	lamb	1.00	0	0.39	Fit
14	wand	mask	world	1.00	0	0.31	Fit
15	cast	cost	jump	0.16	-0.50	0.24	Fit

Phoneme blending test items by 'fit' or 'misfit' by the middle of the school year

Test Items				Outfit Statistics			Interpretation
Click on the picture that you think I am saying?				MNSQ	ZSTD	PT-measure correlation	'Fit' or 'Misfit'
1	dog	dot	man	0.28	-1.51	0.66	Fit
2	mouse	mouth	ring	1.22	0.61	0.70	Fit
3	seal	seat	duck	0.15	-1.32	0.83	Fit
4	bug	sun	bun	0.21	-0.62	0.88	Fit

5	cat	cap	lock	1.96	1.90	0.82	Fit
6	flip	drum	flag	1.70	1.82	0.65	Fit
7	crab	crane	snake	0.54	-0.61	0.78	Fit
8	bread	spade	space	1.13	0.41	0.73	Fit
9	tray	clown	train	1.10	0.43	0.69	Fit
10	stop	star	plane	0.68	-0.14	0.85	Fit
11	point	fast	pond	0.93	0.12	0.61	Fit
12	bank	band	toast	0.49	-1.01	0.69	Fit
13	desk	lamp	lamb	0.37	-1.32	0.73	Fit
14	wand	mask	world	0.46	-1.13	0.64	Fit
15	cast	cost	jump	0.44	-1.12	0.64	Fit

Phoneme blending test items by 'fit' or 'misfit' by the end of the school year

Test Items				Outfit Statistics			Interpretation
Click on the picture that you think I am saying?				MNSQ	ZSTD	PT-measure correlation	'Fit' or 'Misfit'
1	dog	dot	man	0.29	-1.60	0.67	Fit
2	mouse	mouth	ring	1.20	0.55	0.68	Fit
3	seal	seat	duck	0.14	-1.21	0.80	Fit
4	bug	sun	bun	0.22	-0.70	0.89	Fit
5	cat	cap	lock	1.92	1.96	0.79	Fit
6	flip	drum	flag	1.68	1.76	0.67	Fit
7	crab	crane	snake	0.56	-0.68	0.76	Fit
8	bread	spade	space	1.12	0.34	0.72	Fit
9	tray	clown	train	1.09	0.32	0.68	Fit
10	stop	star	plane	0.69	-0.21	0.86	Fit
11	point	fast	pond	0.94	0.32	0.58	Fit
12	bank	band	toast	0.51	-1.21	0.71	Fit
13	desk	lamp	lamb	0.39	-1.50	0.75	Fit
14	wand	mask	world	0.44	-0.90	0.61	Fit
15	cast	cost	jump	0.45	-1.21	0.62	Fit

In the phoneme blending task, all items demonstrated a model 'fit' at the start, middle and end of the first year at school. This is because all test items produced a mean-square statistic less than 2.0 and a ZSTD statistic between -2.0 and +2.0.

Phoneme Deletion

Phoneme deletion test items by 'fit' or 'misfit' at school-entry

Test Items				Outfit Statistics			Interpretation
Say (original word). Say (original word) but don't say ____ sound at the start/end. What word do we get?				MNSQ	ZSTD	PT-measure correlation	'Fit' or 'Misfit'
1 (cup)	pup	egg	up	2.01	-2.21	0.67	Misfit
2 (door)	oar	core	sew	0.84	-1.51	0.66	Fit
3 (farm)	bow	arm	art	0.33	-1.62	0.53	Fit
4 (bake)	ape	sun	ache	0.38	-0.72	0.37	Fit
5 (bear)	ear	tear	bye	1.49	1.61	0.18	Fit
6 (price)	rice	ice	moon	2.49	3.01	0.25	Misfit
7 (store)	ring	tore	oar	1.00	0	0.34	Fit
8 (break)	ache	horse	rake	1.00	0	0.35	Fit
9 (spin)	pin	in	lock	1.00	0	0.41	Fit
10 (train)	sun	rain	aim	1.00	0	0.32	Fit
11 (beach)	bee	beak	cat	1.00	0	0.44	Fit
12 (seed)	mouse	sea	seat	1.00	0	0.33	Fit
13 (tooth)	toot	fan	two	1.00	0	0.31	Fit
14 (plate)	play	plum	ring	1.00	0	0.41	Fit
15 (rose)	dog	row	rope	1.00	0	0.36	Fit

Phoneme deletion test items by 'fit' or 'misfit' by the middle of the school year

Test Items:				Outfit Statistics			Interpretation
Say (original word). Say (original word) but don't say ____ sound at the start/end. What word do we get?				MNSQ	ZSTD	PT-measure correlation	'Fit' or 'Misfit'
1 (cup)	pup	egg	up	2.38	-2.8	0.77	Misfit
2 (door)	oar	core	sew	0.53	-1.4	0.76	Fit
3 (farm)	bow	arm	art	0.55	-0.2	0.84	Fit
4 (bake)	ape	sun	ache	0.58	-0.2	0.83	Fit
5 (bear)	ear	tear	bye	0.60	-0.3	0.82	Fit
6 (price)	rice	ice	moon	0.49	1.3	0.83	Fit
7 (store)	ring	tore	oar	0.68	-0.4	0.75	Fit
8 (break)	ache	horse	rake	0.35	-1.1	0.73	Fit
9 (spin)	pin	in	lock	1.42	0.8	0.40	Fit
10 (train)	sun	rain	aim	0.64	-0.4	0.60	Fit
11 (beach)	bee	beak	cat	0.49	1.6	0.42	Fit
12 (seed)	mouse	sea	seat	0.69	-0.3	0.45	Fit
13 (tooth)	toot	fan	two	0.53	-1.1	0.58	Fit
14 (plate)	play	plum	ring	0.24	-1.6	0.39	Fit
15 (rose)	dog	row	rope	0.50	-1.8	0.35	Fit

Phoneme deletion test items by 'fit' or 'misfit' by the end of the school year

Test Items				Outfit Statistics			Interpretation
Say (original word). Say (original word) but don't say ____ sound at the start/end. What word do we get?				MNSQ	ZSTD	PT-measure correlation	'Fit' or 'Misfit'
1 (cup)	pup	egg	up	2.03	-3.52	0.37	Misfit
2 (door)	oar	core	sew	0.59	-0.81	0.47	Fit
3 (farm)	bow	arm	art	0.26	-1.83	0.52	Fit
4 (bake)	ape	sun	ache	0.24	-1.44	0.69	Fit
5 (bear)	ear	tear	bye	2.09	-2.20	0.72	Misfit

6 (price)	rice	ice	moon	0.80	0.11	0.72	Fit
7 (store)	ring	tore	oar	0.10	-1.41	0.84	Fit
8 (break)	ache	horse	rake	0.10	-1.42	0.85	Fit
9 (spin)	pin	in	lock	0.38	-1.31	0.62	Fit
10 (train)	sun	rain	aim	0.72	-0.31	0.75	Fit
11 (beach)	bee	beak	cat	9.90	9.91	0.20	Misfit
12 (seed)	mouse	sea	seat	1.54	2.00	0.48	Fit
13 (tooth)	toot	fan	two	0.24	-2.00	0.69	Fit
14 (plate)	play	plum	ring	0.21	-2.00	0.60	Fit
15 (rose)	dog	row	rope	0.16	-1.81	0.47	Fit

In the phoneme deletion task, item 1 demonstrated a significant ‘misfit’ at the start of the school year, item 6 demonstrated a significant ‘misfit’ at the start of the school year, and items 5 and 11 demonstrated a significant ‘misfit’ at the end of the school year.

- **Phoneme Deletion Item 1:** Ninety-one per cent, ninety-four per cent and ninety-six percent of participants responded correctly on this item at the start, middle and end of the school year respectively. It is possible that children can draw on their knowledge of rhyme to correctly answer this test item, thus making it an easier item than intended. This test item asked children to, 'Say cup. Say cup again, but don’t say the /k/ sound at the start. What word do we get? Is it /pup, egg, or up/?' Rather than deleting the initial sound in the word ‘cup’, it is possible that children can use their rhyme knowledge to guess the answer as ‘pup’ (phonetically similar distractor option) or ‘up’ (correct option). The likelihood that children will respond with ‘up’(correct option) may be increased because this is the last option presented in the multiple choice sequence. Future investigation into this item and similar items is necessary.
- **Phoneme Deletion Item 6:** This item asks children to, 'Say price. Say price again, but don’t say the /p/ sound at the start. What word do we get? Is it /rice, ice, or

moon/?' This item was considered to be of average difficulty at school-entry and thus the misfit cannot be attributed to the item being too difficult or too easy. A number of factors may interact to make this item more challenging at the start of the school year: 1) this is the first test item in the phoneme deletion task that involves an initial blend (after the practice item with an initial blend) , 2) the word ‘price’ may be less familiar to five-year-old children in comparison to other words used in this task, and 3) the animated image for ‘price’ and the static images for ‘rice’ and ‘ice’ may lack clarity. This item demonstrates a model ‘fit’ by the middle and end of the school year suggesting that it becomes increasingly appropriate as children mature in PA development. Investigation regarding how this item can be modified for children at school-entry is necessary.

- **Phoneme Deletion Item 5:** Ninety-three per cent of participants scored correctly on this item at the end of the school year suggesting it had become too easy for the majority of children. This item asked children to, 'Say bear. Say bear again, but don’t say the /b/ sound at the start. What word do we get? It is /ear, tear, or bye/?' This item is one of two items in the phoneme deletion task that involved a CV word being reduced to a V word (i.e. vowel only) and may account for why this item showed a significant ‘misfit’ by the end of the school year.
- **Phoneme Deletion Item 11:** This item presented with an unusual respond pattern throughout the first year of school. This item demonstrated a modal ‘fit’ at school-entry and the middle of the school year and was classified as a moderately difficult item. However, by the end of the school year this item demonstrated a significant model ‘misfit’ and was classified as very difficult. This item asked children to, 'Say beach. Say beach again but don’t say the /ch/ sound at the end. What word do we get? It is bee, beak or cat?' Providing a reasonable explanation as to why this item

suddenly demonstrated a 'misfit' is difficult. It is possible that with an increased awareness of written orthography children in the sample experienced difficulty deleting a sound that is represented by two letters. Therefore, detailed inspection into this item is warranted in future investigations.

Phoneme Segmentation

Phoneme segmentation test items by 'fit' or 'misfit' at school-entry

Test Items		Outfit Statistics			Interpretation
How many sounds in the word …..?		MNSQ	ZSTD	PT-measure correlation	'Fit' or 'Misfit'
1	Moon (3)	0.63	-0.41	0.79	Fit
2	Tooth (3)	0.56	-0.41	0.87	Fit
3	Cow (2)	2.08	-2.82	0.29	Misfit
4	Cup (3)	0.08	-1.41	0.45	Fit
5	Soap (3)	Winsteps would not analyse items 4 to 18 due to lack of variance in the dataset (i.e., the majority of responses were incorrect)			Misfit
6	Saw (2)				Misfit
7	Flush (4)				Misfit
8	Crab (4)				Misfit
9	Sew (2)				Misfit
10	Step (4)				Misfit
11	Plate (4)				Misfit
12	Star (3)				Misfit
13	Bank (4)				Misfit
14	Lock (3)				Misfit
15	Jump (4)				Misfit
16	Pond (4)				Misfit
17	Bear (2)				Misfit
18	Cast (4)				Misfit

Phoneme segmentation test items by 'fit' or 'misfit' by the middle of the school year

Test Items How many sounds in the word …..?		Outfit Statistics MNSQ	ZSTD	PT-measure correlation	Interpretation 'Fit' or 'Misfit'
1	Moon (3)	2.50	-2.80	0.78	Misfit
2	Tooth (3)	0.52	-1.44	0.77	Fit
3	Cow (2)	0.56	-0.12	0.83	Fit
4	Cup (3)	0.59	-0.21	0.80	Fit
5	Soap (3)	0.61	-0.31	0.81	Fit
6	Saw (2)	1.39	1.32	0.85	Fit
7	Flush (4)	0.67	-0.40	0.76	Fit
8	Crab (4)	0.34	-1.21	0.72	Fit
9	Sew (2)	1.41	0.92	0.39	Fit
10	Step (4)	0.62	-0.42	0.59	Fit
11	Plate (4)	1.10	1.51	0.40	Fit
12	Star (3)	0.58	-0.20	0.43	Fit
13	Bank (4)	0.52	-1.00	0.56	Fit
14	Lock (3)	0.23	-1.71	0.40	Fit
15	Jump (4)	0.48	-1.90	0.37	Fit
16	Pond (4)	Winsteps could not analyse items 16, 17 and 18 due to a lack of variance in the dataset (i.e., the majority of responses were incorrect)			Misfit
17	Bear (2)				Misfit
18	Cast (4)				Misfit

Phoneme segmentation test items by 'fit' or 'misfit' by the end of the school year

Test Items How many sounds in the word …..?		Outfit Statistics MNSQ	ZSTD	PT-measure correlation	Interpretation 'Fit' or 'Misfit'
1	Moon (3)	3.96	2.50	0.82	Misfit
2	Tooth (3)	1.22	0.62	0.70	Fit
3	Cow (2)	0.15	-1.30	0.83	Fit
4	Cup (3)	0.21	-0.61	0.88	Fit

5	Soap (3)	0.28	-1.53	0.66	Fit
6	Saw (2)	1.70	1.82	0.65	Fit
7	Flush (4)	0.54	-0.61	0.78	Fit
8	Crab (4)	1.13	0.41	0.73	Fit
9	Sew (2)	1.10	0.42	0.69	Fit
10	Step (4)	0.68	-0.12	0.85	Fit
11	Plate (4)	0.93	0.11	0.61	Fit
12	Star (3)	0.49	-1.00	0.69	Fit
13	Bank (4)	0.37	-1.31	0.73	Fit
14	Lock (3)	0.46	-1.10	0.64	Fit
15	Jump (4)	0.44	-1.10	0.64	Fit
16	Pond (4)	Winsteps could not analyse items 16, 17 and 18 due to a lack of variance in the dataset (i.e., the majority of responses were incorrect)			Misfit
17	Bear (2)				Misfit
18	Cast (4)				Misfit

Phoneme segmentation is challenging for children at school-entry but becomes increasingly appropriate as children begin to interact with classroom literacy instruction. Thus, it was expected that items within this task would show a significant ‘misfit’, particularly at school-entry. At school-entry, phoneme segmentation items 1, 2 and 4 demonstrated a model ‘fit’, while items 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17 and 18 demonstrated a significant model 'misfit'. Items 1 (‘moon’), 2 (‘tooth’) and 4 (‘cup’) were simple CVC words containing either a long medial vowel, earlier developing sounds (e.g., /k/, /p/, /m/), or both. This may have increased the salience of the phonemes in these items, allowing five-year-old children to segment them correctly. Item 3 (‘cow’) was thought to be easy during the construction of the test. However, over 85 per cent of children indicated that this item had three sounds, as opposed to two. Form video observation it was revealed that children often added a final schwa phoneme when segmenting ‘cow’ verbally. Only nine per cent of children scored correctly on this item. The majority of phoneme segmentation items demonstrating a significant ‘misfit’ had a CCVC or a CVCC syllable structure. Less than 10 per cent of participants provided a correct response to phoneme segmentation items 5 to 18 at school-entry, suggesting that these items demonstrated a ‘misfit’ because they were extremely difficult. Although these items were difficult at school entry, there inclusion in the computer-based assessment allows teachers to present these items to high ability students and provides more challenging test items to be used in the development of a computer-adaptive PA assessment.

By the middle and end of the school year, only four phoneme segmentation items demonstrate a significant ‘misfit’. These were items 1, 16, 17, and 18. Possible explanations as to why these items demonstrate a misfit are offered below:

- **Phoneme Segmentation Item 1:** Ninety-two per cent and ninety-seven per cent of participants responded correctly to this item by the middle and end of the school year respectively. This item was considered very easy by the middle and end of the school year according to the measure statistic. This item asked children, 'How many sounds do you hear in the word /moon/?' The simple CVC syllable structure, use of the long vowel ‘oo’ and early developing sounds ‘m’ and ‘n’ may have contributed to the ease of this item. Retaining such an item for inclusion in the development of a computer-adaptive test is important to allow an instrument to present test items that meet the ability levels of a wide range of children.
- **Phoneme Segmentation Items 16, 17 and 18:** Items 16 and 18 were considered harder phoneme segmentation items because they required children to segment words that contained a word-final blend. Item 17 may have been more challenging because in New Zealand English there is no /r/ sound produced on the end of the word 'bear', however, when produced as an isolated word in segmentation tasks children may be tempted to add a final /r/ sound to this word.

- Three per cent and five per cent of participants responded correctly to item 16 by the middle and end of the school year respectively. This item required children to segment the word ‘pond’.
- Four per cent and six per cent of participants responded correctly to item 17 by the middle and end of the school year respectively. This item required children to segment the word ‘bear’. It is important to note that this word was considered easy by the end of the year when children were required to delete the initial phoneme to create a new word (i.e., phoneme deletion item 5).
- Three per cent and five per cent of participants responded correctly to item 18 by the middle and end of the school year respectively. This item required children to segment the word ‘cast’.

Retaining these items within the computer-based assessment may allow teachers to identify high performing students. Including difficult items to avoiding ceiling effects is also an important part of test development and will contribute to the development of an effective computer-adaptive version of the PA assessment.

Examples of letter-name and letter-sound knowledge test items at school-entry

Letter-name knowledge test items by ‘fit’ or ‘misfit’ at school-entry

Test Items		Outfit Statistics			Interpretation
Show me the letter?		MNSQ	ZSTD	PT-measure correlation	‘Fit’ or ‘Misfit’
1	m	0.77	-0.50	0.82	Fit
2	s	0.35	1.52	0.71	Fit
3	k	0.58	1.67	0.50	Fit
4	b	0.65	1.22	0.34	Fit
5	n	0.77	0.91	0.38	Fit
6	f	0.45	1.34	0.39	Fit

7	d	0.68	1.18	0.77	Fit
8	h	0.48	1.65	0.47	Fit
9	p	0.45	0.98	0.69	Fit
10	t	0.58	1.21	0.35	Fit
11	w	0.57	0.83	0.55	Fit
12	g	0.69	1.51	0.78	Fit
13	c	0.72	1.43	0.77	Fit
14	z	0.43	1.21	0.45	Fit
15	l	0.98	1.32	0.56	Fit
16	q	0.67	0.72	0.65	Fit
17	v	0.78	1.55	0.35	Fit
18	y	0.92	-0.40	0.96	Fit

Letter-sound knowledge test items by 'fit' or 'misfit' at school-entry

Test Items		Outfit Statistics			Interpretation
Show me the letter that makes the (sound)?		MNSQ	ZSTD	PT-measure correlation	'Fit' or 'Misfit'
1	m	0.35	1.45	0.45	Fit
2	s	0.96	1.21	0.81	Fit
3	k	0.60	1.21	0.67	Fit
4	b	0.89	1.54	0.37	Fit
5	n	0.45	1.21	0.48	Fit
6	f	0.46	1.23	0.40	Fit
7	d	0.79	1.10	0.79	Fit
8	h	0.22	1.69	0.33	Fit
9	p	0.46	1.31	0.89	Fit
10	t	0.78	1.34	0.45	Fit
11	w	0.89	0.89	0.67	Fit
12	g	0.48	1.61	0.89	Fit
13	c	0.49	1.89	0.34	Fit
14	z	0.40	1.01	0.56	Fit
15	l	0.34	1.04	0.67	Fit
16	q	0.87	1.23	0.45	Fit

17	v	0.39	1.59	0.37	Fit
18	y	0.89	1.09	0.87	Fit

All letter-name and letter-sound test items demonstrated a model 'fit' at the start, middle and end of the school year. This is because all test items produced a mean-square statistic less than 2.0 and a ZSTD statistic between -2.0 and +2.0.

Appendix D: Tables of Phonological Awareness Test Item Difficulty Using Rasch Model Analysis

The following tables present the 'measure' statistic for each test item in the computer-based phonological awareness (PA) assessment tool at the start, middle, and end of the first year at school. The 'measure' statistic is essentially a logit (log-odds) score assigned to an item to indicate its difficulty (Linacre, 2010). Items with increasingly positive logit scores are more difficult, while items with increasingly negative logit scores are easier. Practice items obtained 'measure' statistics in the easy range, and thus the focus here is on test items. Grey shading in each table indicates the correct multiple-choice answer.

Rhyme Oddity

Rhyme oddity test items by difficulty at school-entry

Test Items				Outfit Statistics		
Which word does not rhyme?				MEASURE	ERROR	Interpretation
10	ring	sing	lamb	3.72	0.33	More Difficult
9	shell	duck	bell	3.67	0.32	↑
7	bun	sun	kite	2.96	0.28	
3	saw	toe	bow	1.77	0.25	
8	tent	lock	sock	0.89	0.25	
2	peg	doll	leg	-0.68	0.29	
4	sand	hand	cup	-1.47	0.29	
6	dog	book	hook	-2.80	0.31	
5	hen	car	pen	-3.96	0.38	↓
1	cat	mat	bus	-4.11	0.39	Less Difficult

Rhyme oddity test items by difficulty by the middle of the school year

Test Items				Outfit Statistics		
Which word does not rhyme?				MEASURE	ERROR	Interpretation
10	ring	sing	lamb	3.65	0.32	More Difficult
9	shell	duck	bell	3.55	0.30	↑
7	bun	sun	kite	2.85	0.25	
3	saw	toe	bow	1.81	0.21	
8	tent	lock	sock	0.93	0.22	
2	peg	doll	leg	-0.70	0.26	
4	sand	hand	cup	-1.43	0.25	
6	dog	book	hook	-2.76	0.28	
5	hen	car	pen	-3.90	0.31	↓
1	cat	mat	bus	-5.10	0.32	Less Difficult

Rhyme oddity test items by difficulty by the end of the school year

Test Items				Outfit Statistics		
Which word does not rhyme?				MEASURE	ERROR	Interpretation
10	ring	sing	lamb	6.12	0.47	More Difficult
9	shell	duck	bell	4.24	0.36	↑
8	tent	lock	sock	1.92	0.36	
7	bun	sun	kite	0.82	0.38	
5	hen	car	pen	-0.99	0.43	
3	saw	toe	bow	-4.11	0.50	
2	peg	doll	leg	-5.71	0.67	
6	dog	book	hook	-8.31	0.40	
1	cat	mat	bus	-9.40	0.45	↓
4	sand	hand	cup	-9.59	0.45	Less Difficult

Initial Phoneme Identity

Initial phoneme identity test items by difficulty at school-entry

Test Items:					**Outfit Statistics**		
Which word starts with the ___ sound?					MEASURE	ERROR	Interpretation
6	/k/ sound?	comb	dish	soap	6.69	0.55	More Difficult
9	/f/ sound?	foot	hat	pig	5.86	0.50	↑
10		duck	bell	fire	3.79	0.48	
8	/b/ sound?	car	boat	shoe	1.15	0.38	
7		cat	leg	ball	0.61	0.36	
5	/k/ sound?	bus	kite	arm	-0.66	0.36	
4	/s/ sound?	saw	tie	Hook	-0.93	0.36	
3		bee	sun	tent	-4.19	0.47	↓
2	/m/ sound?	mat	dog	Book	-6.14	0.44	
1		doll	bear	milk	-6.14	0.44	Less Difficult

Initial phoneme identity test items by difficulty by the middle of the school year

Test Items					Outfit Statistics		
Which word starts with the ___ sound?					MEASURE	ERROR	Interpretation
6	/k/ sound	comb	dish	soap	2.67	0.31	More Difficult
9	/f/ sound	foot	hat	pig	2.47	0.31	↑
10		duck	bell	fire	2.47	0.31	
8	/b/ sound	car	boat	shoe	0.28	0.36	
7		cat	leg	ball	0.28	0.36	
4	/s/ sound	saw	tie	hook	0.15	0.37	
3		bee	sun	tent	-0.73	0.40	
5	/k/ sound	bus	kite	arm	-1.56	0.42	
2	/m/ sound	mat	dog	Book	-2.87	0.53	↓
1		doll	bear	milk	-3.17	0.58	Less Difficult

Initial phoneme identity test items by difficulty by the end of the school year

Test Items					Outfit Statistics		
Which word starts with the ___ sound?					MEASURE	ERROR	Interpretation
10	/f/ sound	duck	bell	fire	2.01	0.37	More Difficult
9		foot	hat	pig	1.87	0.37	↑
7	/b/ sound	cat	leg	ball	0.88	0.39	
6	/k/ sound	comb	dish	soap	0.57	0.40	
8	/b/ sound	car	boat	shoe	0.57	0.40	
5	/k/ sound	bus	kite	arm	0.06	0.42	
4	/s/ sound	saw	tie	hook	-0.70	0.45	↓
3		bee	sun	tent	-1.86	0.53	
2	/m/	mat	dog	book	-3.30	0.78	
1	sound	doll	bear	milk	-4.50	0.37	Less Difficult

Final Phoneme Identity

Final phoneme identity test items by difficulty at school-entry

Test Items				**Outfit Statistics**		
Which word ends with the ___ sound?				MEASURE	ERROR	Interpretation
2	hat	hole	sun	6.50	0.38	More Difficult
4	food	fan	hook	5.05	1.82	↑
5	green	grass	fire	5.05	1.82	
6	hand	horse	milk	5.05	1.82	
7	ball	goat	bank	5.05	1.82	
8	cake	camp	bed	5.05	1.82	
9	world	walk	face	5.05	1.82	
10	tent	toast	map	5.05	1.82	
3	rope	rice	pan	1.06	0.39	↓
1	room	seal	soup	-1.97	0.38	Less Difficult

Final phoneme identity test items by difficulty by the middle of the school year

Test Items				Outfit Statistics		
Which word ends with the ___ sound?				MEASURE	ERROR	Interpretation
10	tent	toast	map	2.82	0.31	More Difficult
5	green	grass	fire	2.21	0.28	↑
6	hand	horse	milk	2.13	0.28	
9	world	walk	face	1.46	0.27	
7	ball	goat	bank	1.32	0.27	
4	food	fan	hook	0.57	0.28	
8	cake	camp	bed	0.32	0.29	
3	rope	rice	pan	-2.78	0.50	
2	hat	hole	sun	-3.63	0.58	↓
1	room	seal	soup	-4.43	0.69	Less Difficult

Final phoneme identity test items by difficulty by the end of the school year

Test Items				Outfit Statistics		
Which word ends with the ___ sound?				MEASURE	ERROR	Interpretation
9	world	walk	face	5.47	0.38	More Difficult
10	tent	toast	map	5.20	0.37	↑
6	hand	horse	milk	2.11	0.33	
5	green	grass	fire	1.90	0.33	
8	cake	camp	bed	1.67	0.34	
7	ball	goat	bank	1.03	0.37	
3	rope	rice	pan	-3.30	0.83	
2	hat	hole	sun	-4.10	1.05	
1	room	seal	soup	-4.71	0.83	↓
4	food	fan	hook	-9.5	0.55	Less Difficult

Phoneme Blending

Phoneme blending test items by difficulty at school-entry

Test Items					Outfit Statistics	
Click on the picture that you think I am saying?				MEASURE	ERROR	Interpretation
6	flip	drum	flag	4.89	1.87	More Difficult
7	crab	crane	snake	4.89	1.87	↑
8	bread	spade	space	4.89	1.87	
9	tray	clown	train	4.89	1.87	
10	stop	star	plane	4.89	1.87	
11	point	fast	pond	4.89	1.87	
12	bank	band	toast	4.89	1.87	
13	desk	lamp	lamb	4.89	1.87	
14	wand	mask	world	4.89	1.87	
15	cast	cost	jump	3.56	1.08	
5	cat	cap	lock	1.53	0.52	
4	bug	sun	bun	1.28	0.48	↓
3	seal	seat	duck	-1.84	0.29	
2	mouse	mouth	ring	-1.92	0.29	
1	dog	dot	man	-2.61	0.30	Less Difficult

Phoneme blending test items by difficulty by the middle of the school year

Test Items					Outfit Statistics	
Click on the picture that you think I am saying?				MEASURE	ERROR	Interpretation
15	cast	cost	jump	3.33	0.32	More Difficult
14	wand	mask	world	3.23	0.32	↑
11	point	fast	pond	2.84	0.31	
12	bank	band	toast	2.64	0.31	
6	flip	drum	flag	2.64	0.31	↓
13	desk	lamp	lamb	2.35	0.32	
9	tray	clown	train	1.94	0.33	

8	bread	spade	space	1.49	0.35	
7	crab	crane	snake	1.36	0.35	
10	stop	star	plane	0.30	0.44	
5	cat	cap	lock	-0.12	0.48	
4	bug	sun	bun	-3.19	0.56	
3	seal	seat	duck	-4.16	0.51	
2	mouse	mouth	ring	-6.46	0.56	
1	dog	dot	man	-7.73	0.78	Less Difficult

Phoneme blending test items by difficulty by the end of the school year

Test Items				Outfit Statistics		
Click on the picture that you think I am saying?				MEASURE	ERROR	Interpretation
15	cast	cost	jump	3.39	0.30	More Difficult
14	wand	mask	world	3.18	0.32	
11	point	fast	pond	2.78	0.29	
12	bank	band	toast	2.59	0.28	
6	flip	drum	flag	2.59	0.34	
13	desk	lamp	lamb	2.30	0.31	
9	tray	clown	train	1.88	0.30	
8	bread	spade	space	1.41	0.31	
7	crab	crane	snake	1.29	0.32	
10	stop	star	plane	0.25	0.42	
5	cat	cap	lock	-0.03	0.49	
4	bug	sun	bun	-3.07	0.56	
3	seal	seat	duck	-4.01	0.49	
2	mouse	mouth	ring	-6.35	0.52	
1	dog	dot	man	-7.70	0.74	Less Difficult

Phoneme Deletion

Phoneme deletion test items by difficulty at school-entry

Test Items Say (original word). Say (original word) again but don't say ____ sound at the start/end. What word do we get?				Outfit Statistics MEASURE	ERROR	Interpretation
7 (store)	ring	tore	oar	3.72	1.82	More Difficult
8 (break)	ache	horse	rake	3.72	1.82	↑
9 (spin)	pin	in	lock	3.72	1.82	
10 (train)	sun	rain	aim	3.72	1.82	
11 (beach)	bee	beak	cat	3.72	1.82	
12 (seed)	mouse	sea	seat	3.72	1.82	
13 (tooth)	toot	fan	two	3.72	1.82	
14 (plate)	play	plum	ring	3.72	1.82	
15 (rose)	dog	row	rope	3.72	1.82	
4 (bake)	ape	sun	ache	1.75	1.75	
3 (farm)	bow	arm	art	0.96	0.82	
5 (bear)	ear	tear	bye	-0.06	-0.03	
6 (price)	rice	ice	moon	-0.91	-0.90	↓
2 (door)	oar	core	sew	-1.40	-1.40	
1 (cup)	pup	egg	up	-2.70	-0.34	Less Difficult

Phoneme deletion test items by difficulty by the middle of the school year

Test Items Say (original word). Say (original word) again but don't say ____ sound at the start/end. What word do we get?				Outfit Statistics MEASURE	ERROR	Interpretation
15 (rose)	dog	row	rope	5.72	0.56	More Difficult
14 (plate)	play	plum	ring	4.98	0.45	
9 (spin)	pin	in	lock	4.42	0.40	
11 (beach)	bee	beak	cat	3.99	0.37	
12 (seed)	mouse	sea	seat	3.72	0.36	

13 (tooth)	toot	fan	two	3.01	0.34	↑
10 (train)	sun	rain	aim	2.34	0.33	
8 (break)	ache	horse	rake	1.29	0.36	
7 (store)	ring	tore	oar	0.74	0.38	
6 (price)	rice	ice	moon	-1.40	0.46	
4 (bake)	ape	sun	ache	-2.03	0.46	
3 (farm)	bow	arm	art	-2.44	0.46	
5 (bear)	ear	tear	bye	-4.72	0.52	
1 (cup)	pup	egg	up	-9.81	1.15	↓
2 (door)	oar	core	sew	-9.81	1.15	Less Difficult

Phoneme deletion test items by difficulty by the end of the school year

Test Items					Outfit Statistics	
Say (original word). Say (original word) again but don't say _____ sound at the start/end. What word do we get?				MEASURE	ERROR	Interpretation
11 (beach)	bee	beak	cat	8.89	0.77	More Difficult
15 (rose)	dog	row	rope	6.95	0.45	
14 (plate)	play	plum	ring	5.35	0.34	↑
9 (spin)	pin	in	lock	4.91	0.33	
13 (tooth)	toot	fan	two	4.18	0.32	
12 (seed)	mouse	sea	seat	3.66	0.32	
10 (train)	sun	rain	aim	2.79	0.34	
8 (break)	ache	horse	rake	-0.92	0.49	
7 (store)	ring	tore	oar	-1.66	0.50	
6 (price)	rice	ice	moon	-2.68	0.51	
4 (bake)	ape	sun	ache	-4.23	0.52	
5 (bear)	ear	tear	bye	-4.23	0.52	
3 (farm)	bow	arm	art	-6.60	0.67	↓
2 (door)	oar	core	sew	-7.08	0.72	
1 (cup)	pup	egg	up	-9.32	1.12	Less Difficult

Phoneme Segmentation

Phoneme segmentation test items by difficulty at school-entry

Test Items		Outfit Statistics		
How many sounds in the word …..?		MEASURE	ERROR	Interpretation
3	Cow (2)	8.12	4.92	More Difficult
4	Cup (3)	6.11	6.11	
2	Tooth (3)	-2.43	-2.43	
1	Moon (3)	-8.60	-8.60	↑
5	Soap (3)	Winsteps would not analyse items five to 18 due to lack of variance in the dataset (i.e., the majority of responses were incorrect)		
6	Saw (2)			
7	Flush (4)			
8	Crab (4)			
9	Sew (2)			
10	Step (4)			
11	Plate (4)			
12	Star (3)			
13	Bank (4)			
14	Lock (3)			↓
15	Jump (4)			
16	Pond (4)			
17	Bear (2)			
18	Cast (4)			Less Difficult

Phoneme segmentation test items by difficulty by the middle of the school year

Test Items		Outfit Statistics		
How many sounds in the word …..?		MEASURE	ERROR	Interpretation
16	Pond (4)	> 9	-	More Difficult
17	Bear (2)	> 9	-	↑
18	Cast (4)	> 9	-	
15	Jump (4)	5.68	0.54	
14	Lock (3)	4.92	0.48	
9	Sew (2)	4.39	0.39	

11	Plate (4)	4.01	0.35	
12	Star (3)	3.70	0.38	
13	Bank (4)	2.98	0.32	
10	Step (4)	2.30	0.36	
8	Crab (4)	1.25	0.33	
7	Flush (4)	0.72	0.39	
6	Saw (2)	-1.37	0.46	
4	Cup (3)	-2.03	0.42	
3	Cow (2)	-2.40	0.43	
5	Soap (3)	-4.70	0.50	
2	Tooth (3)	-9.78	1.12	
1	Moon (3)	-9.79	1.10	Less Difficult

Phoneme segmentation test items by difficulty by the end of the school year

Test Items			Outfit Statistics	
How many sounds in the word …..?		MEASURE	ERROR	Interpretation
16	Pond (4)	> 9	-	More Difficult
17	Bear (2)	> 9	-	
18	Cast (4)	> 9	-	
15	Jump (4)	3.33	0.32	
14	Lock (3)	3.23	0.30	
11	Plate (4)	2.84	0.30	
6	Saw (2)	2.64	0.30	
12	Star (3)	2.64	0.33	
13	Bank (4)	2.35	0.31	
9	Sew (2)	1.94	0.32	
8	Crab (4)	1.49	0.36	
7	Flush (4)	1.36	0.32	
10	Step (4)	0.30	0.45	
4	Cup (3)	-3.19	0.55	
3	Cow (2)	-4.61	0.53	
2	Tooth (3)	-6.46	0.55	
5	Soap (3)	-7.70	0.76	Less Difficult

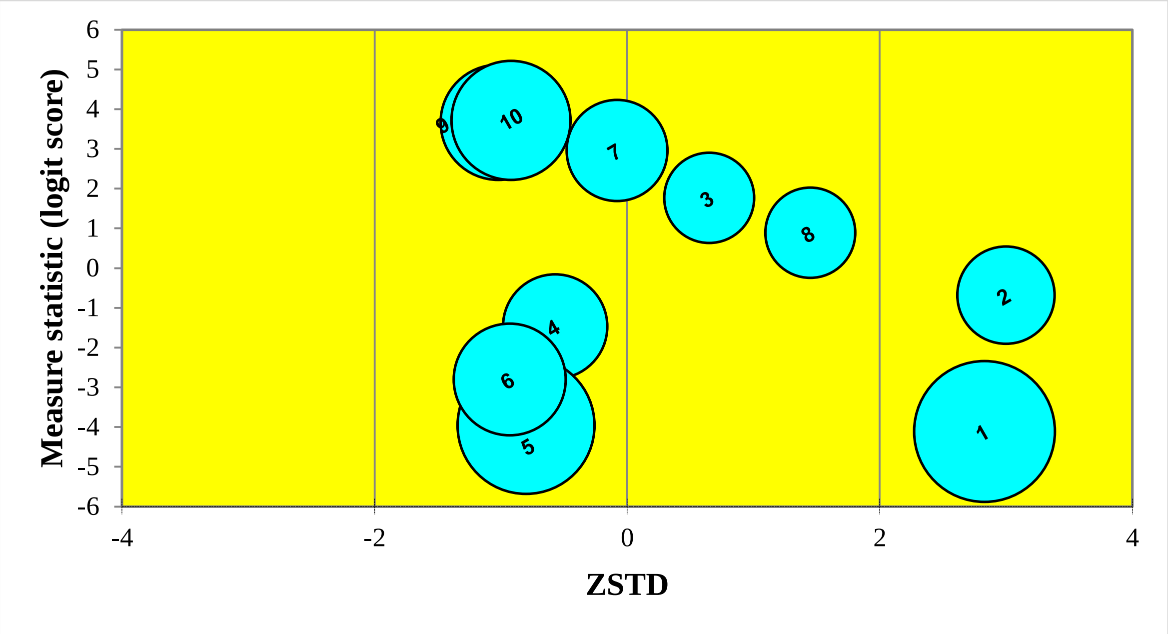
1	Moon (3)	-8.69	0.48

Appendix E: Bubble Charts Profiling Test Item 'Fit' and Difficulty for Phonological Awareness Tasks

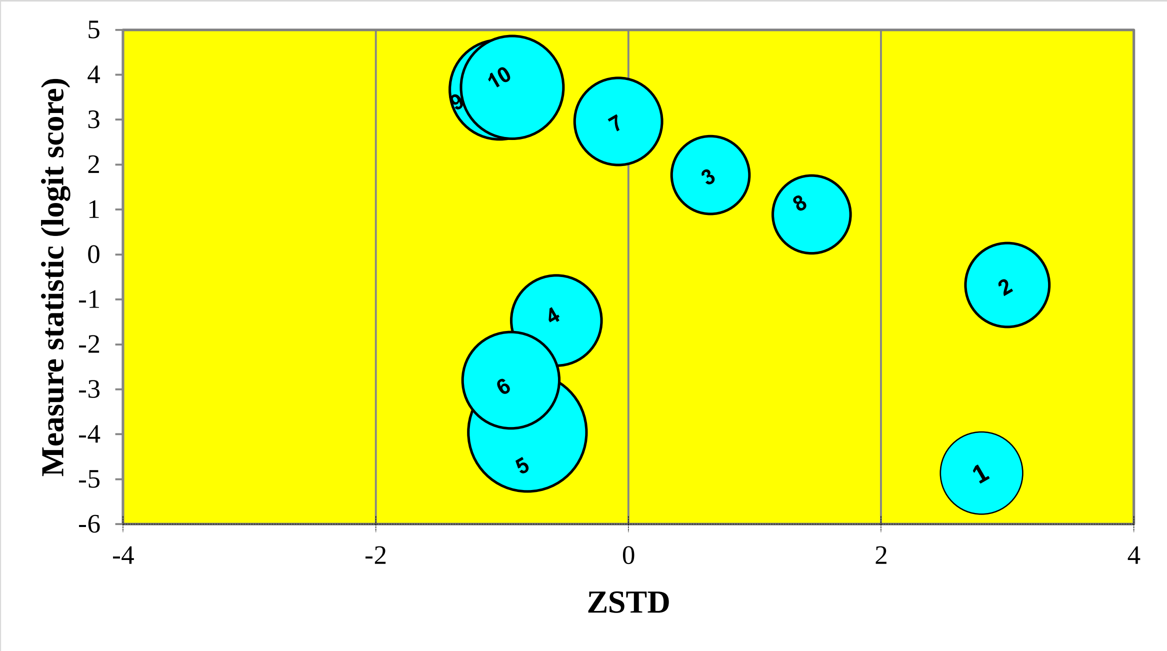
The bubble charts presented in this appendix compare the test items within each PA task by item 'fit' and item difficulty. The horizontal axis represents the ZSTD statistic, where items within the range of -2 to +2 demonstrate a model 'fit' and items outside this range demonstrate a significant model 'misfit'. The vertical axis represents item difficulty in logits, as described in Chapter 6.

Rhyme Oddity

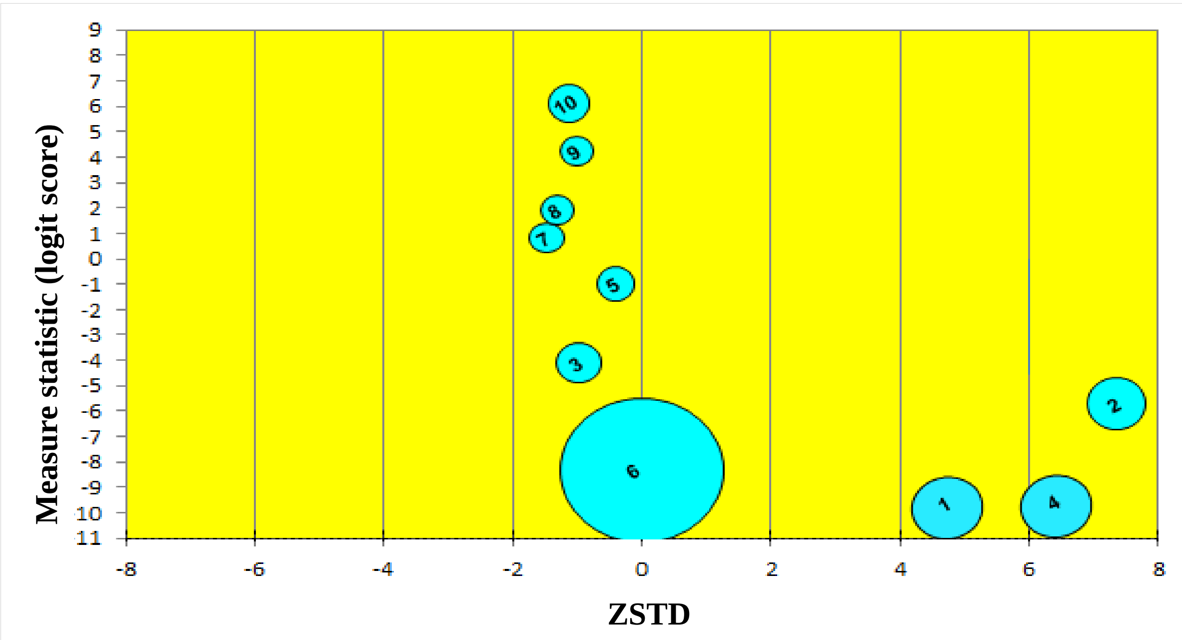
Rhyme oddity test items by 'fit' and hierarchy of difficulty at school-entry



Rhyme oddity test items by 'fit' and hierarchy of difficulty by the middle of the school year

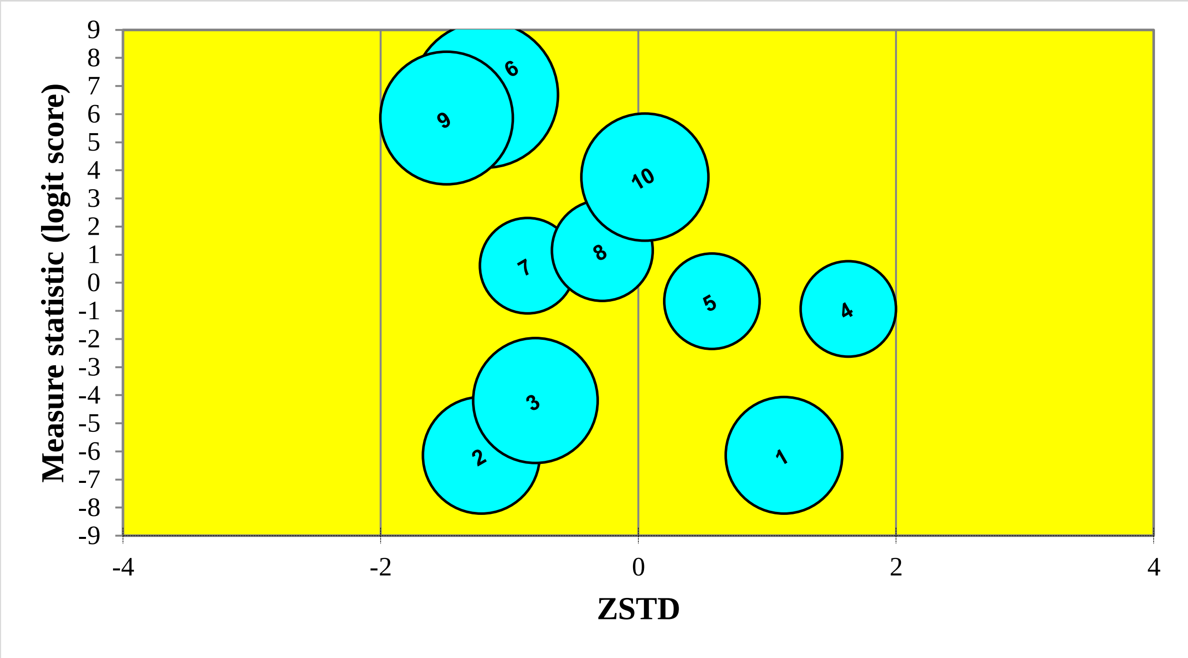


Rhyme oddity test items by 'fit' and hierarchy of difficulty by the end of the school year

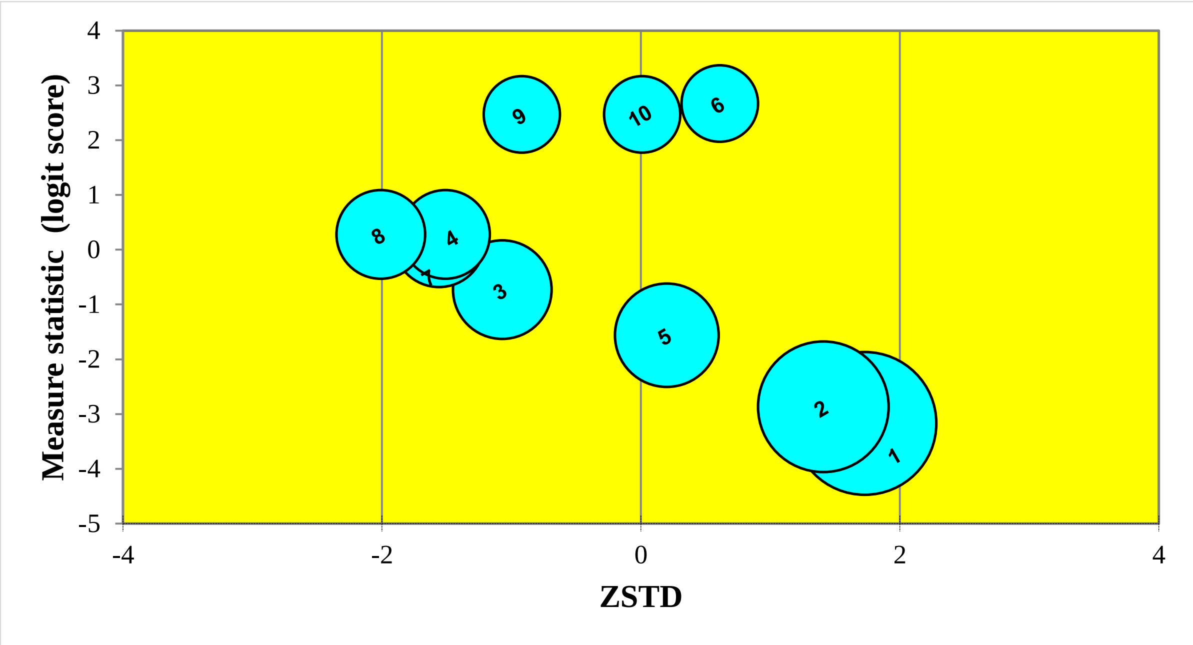


Initial Phoneme Identity

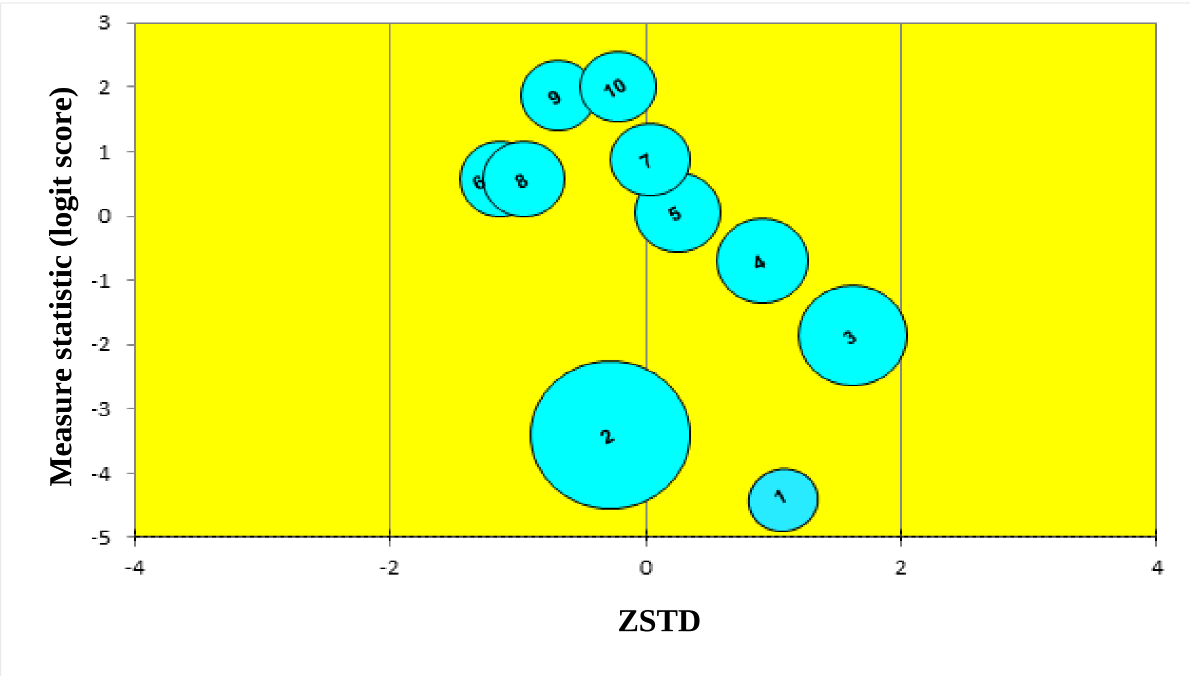
Initial Phoneme Identity test items by 'fit' and hierarchy of difficulty at school-entry



Initial Phoneme Identity test items by 'fit' and hierarchy of difficulty by the middle of the school year

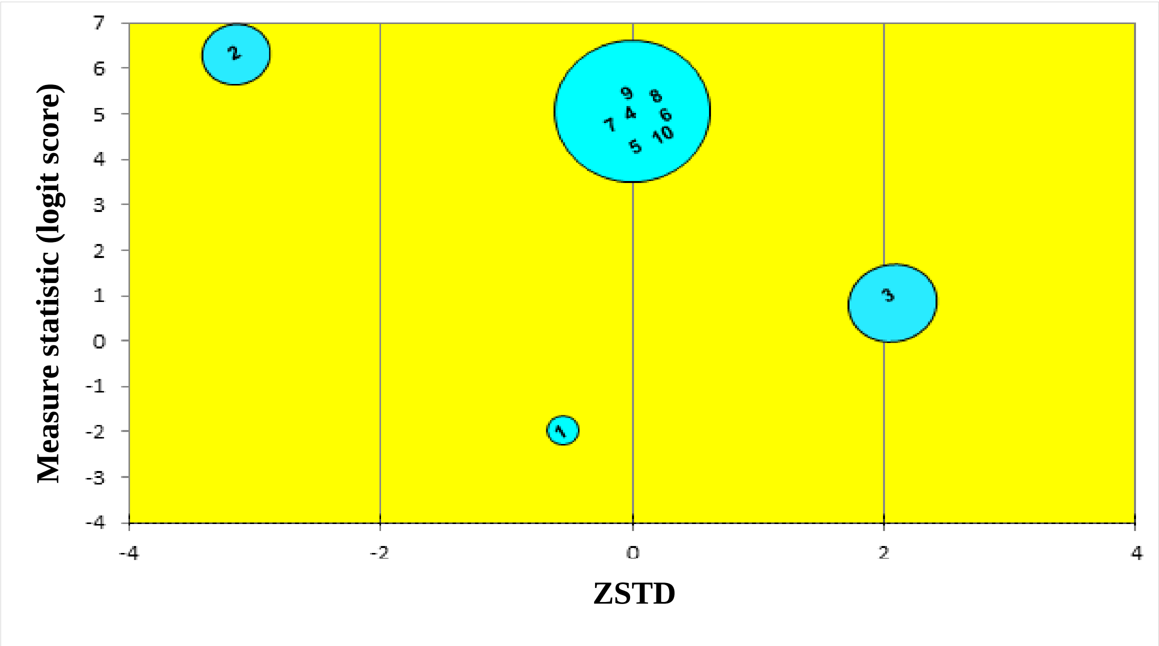


Initial Phoneme Identity test items by 'fit' and hierarchy of difficulty by the end of the school year

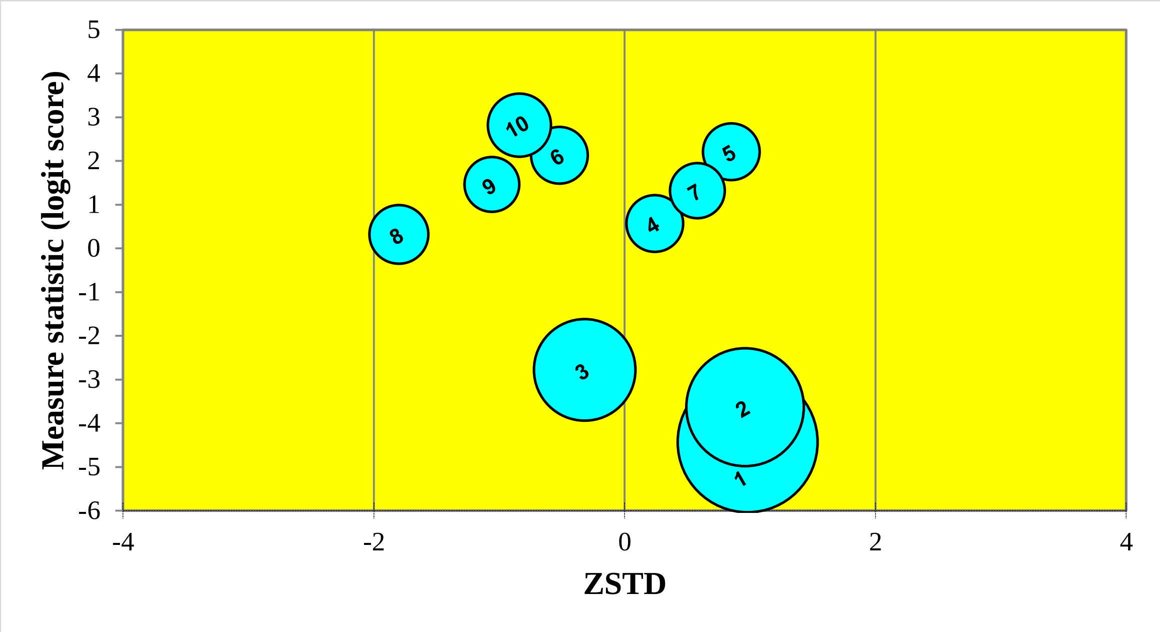


Final Phoneme Identity

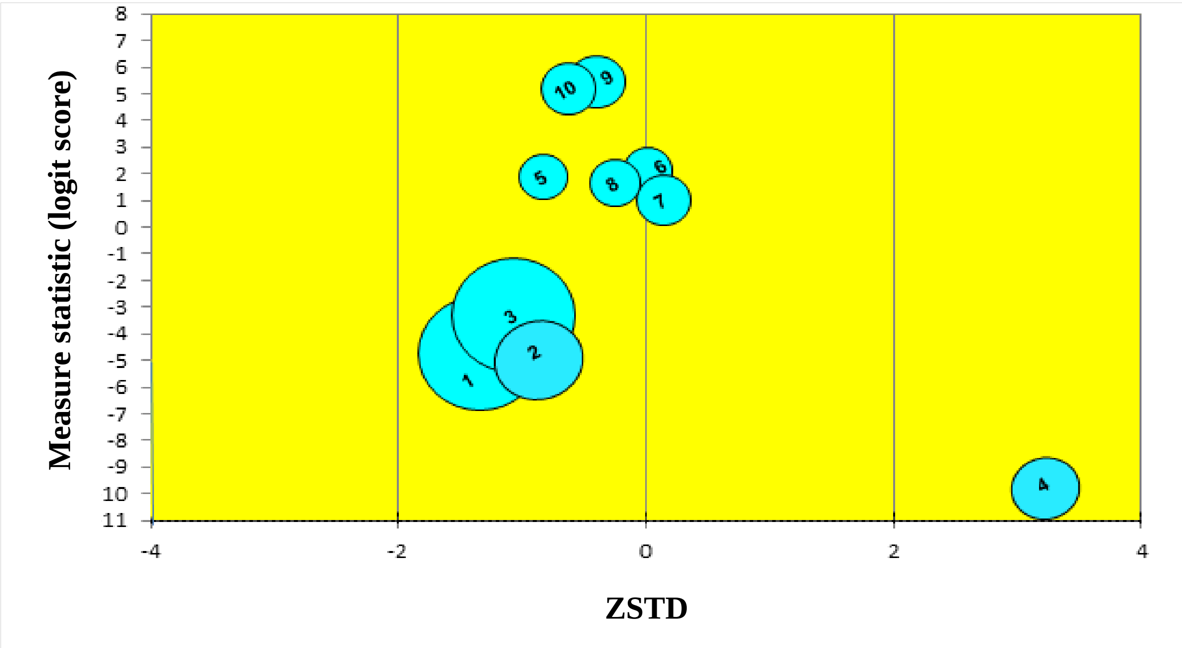
Final Phoneme Identity test items by 'fit' and hierarchy of difficulty at school-entry



Final Phoneme Identity test items by 'fit' and hierarchy of difficulty by the middle of the school year

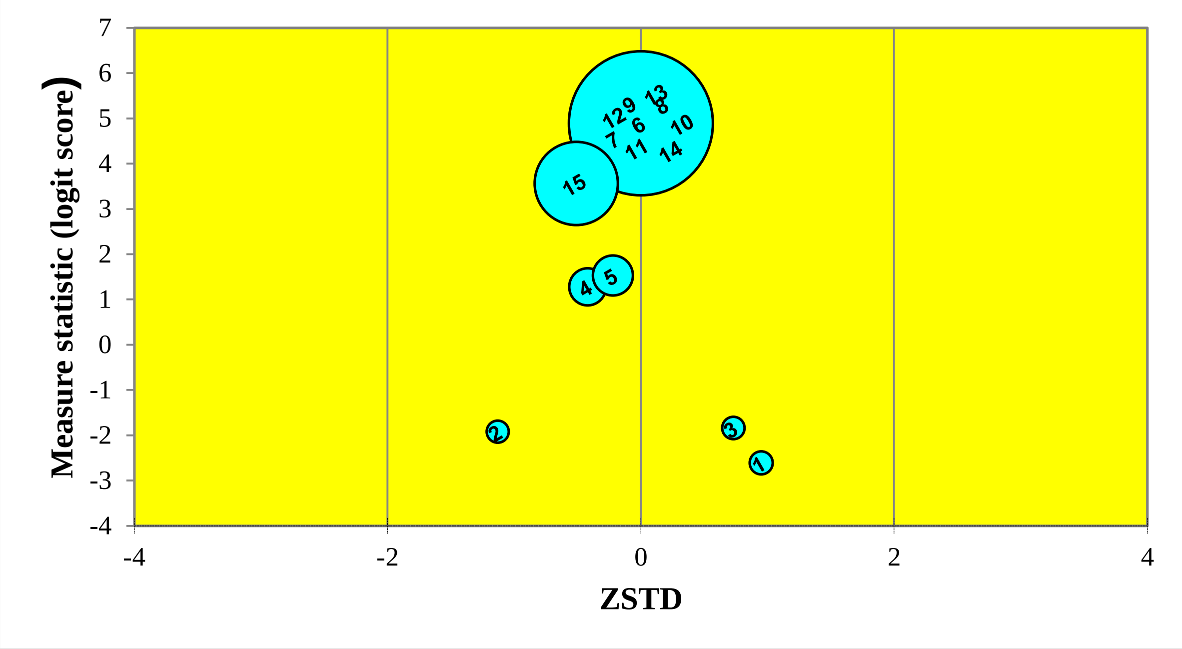


Final Phoneme Identity test items by 'fit' and hierarchy of difficulty by the end of the school year

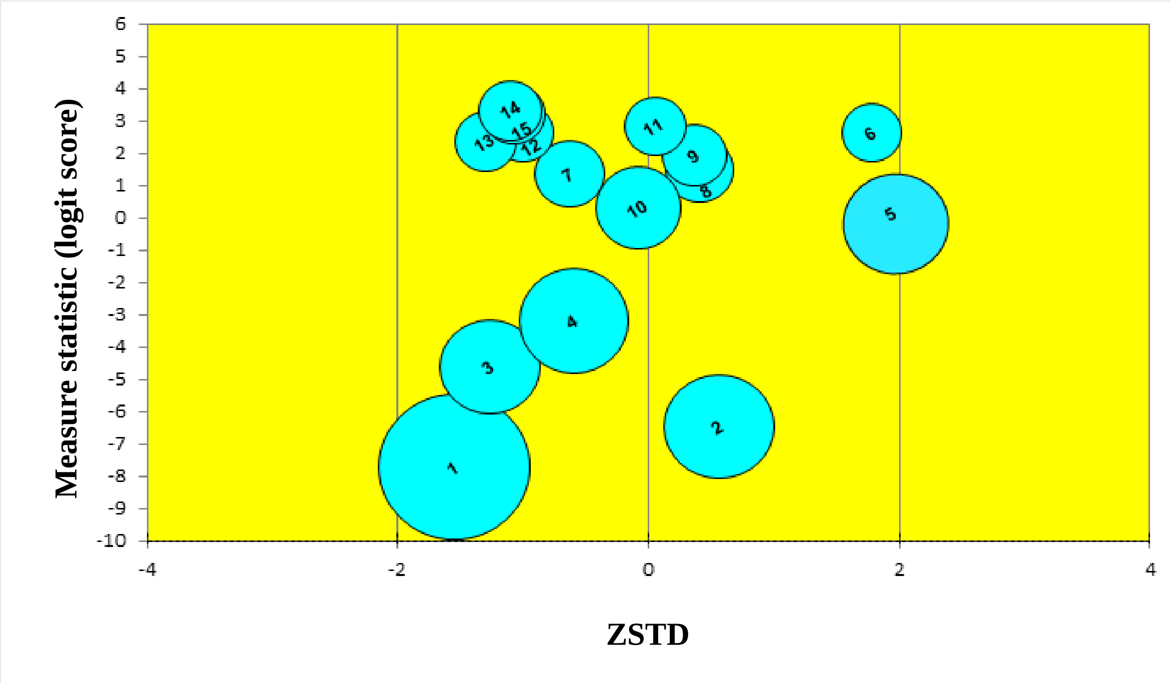


Phoneme Blending

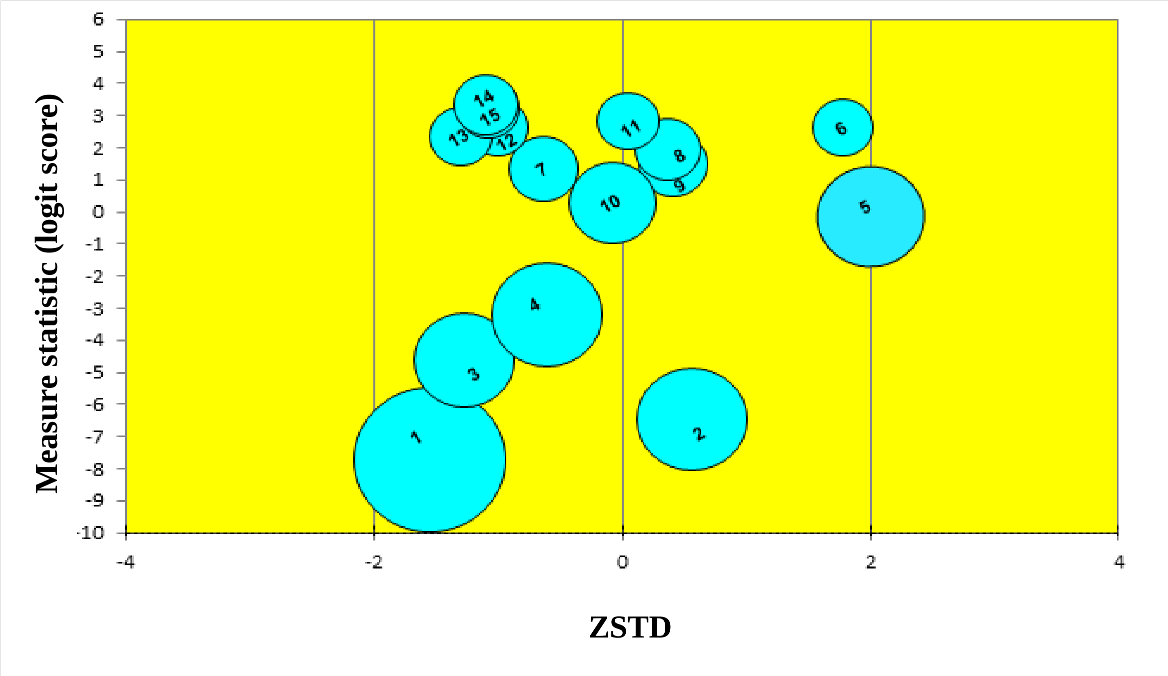
Phoneme Blending test items by 'fit' and hierarchy of difficulty at school-entry



Phoneme Blending test items by 'fit' and hierarchy of difficulty by the middle of the school year

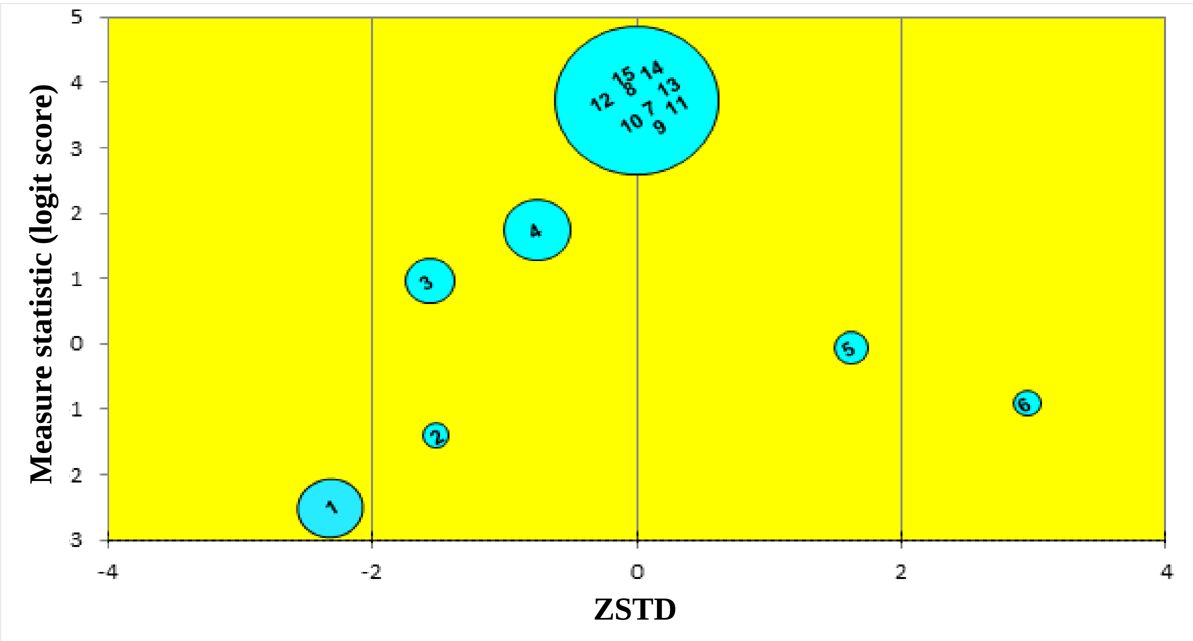


Phoneme Blending test items by 'fit' and hierarchy of difficulty by the end of the school year

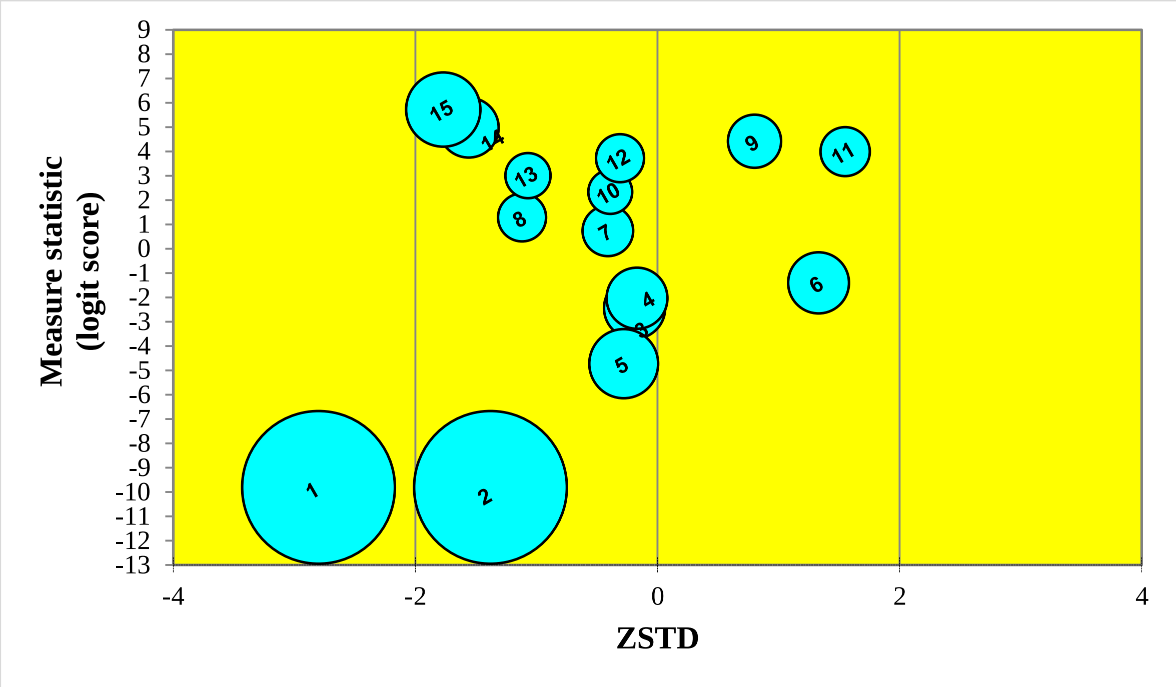


Phoneme Deletion

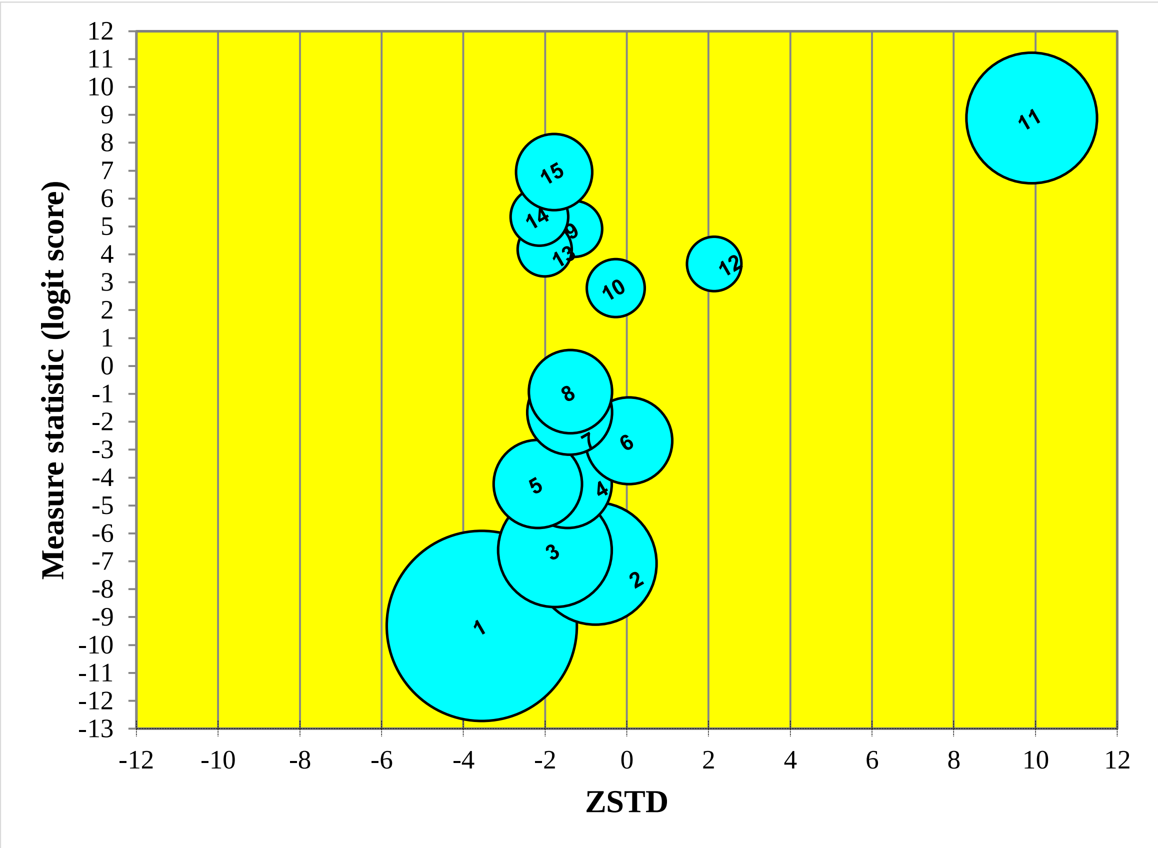
Phoneme Deletion test items by 'fit' and hierarchy of difficulty at school-entry



Phoneme Deletion test items by 'fit' and hierarchy of difficulty by the middle of the school year

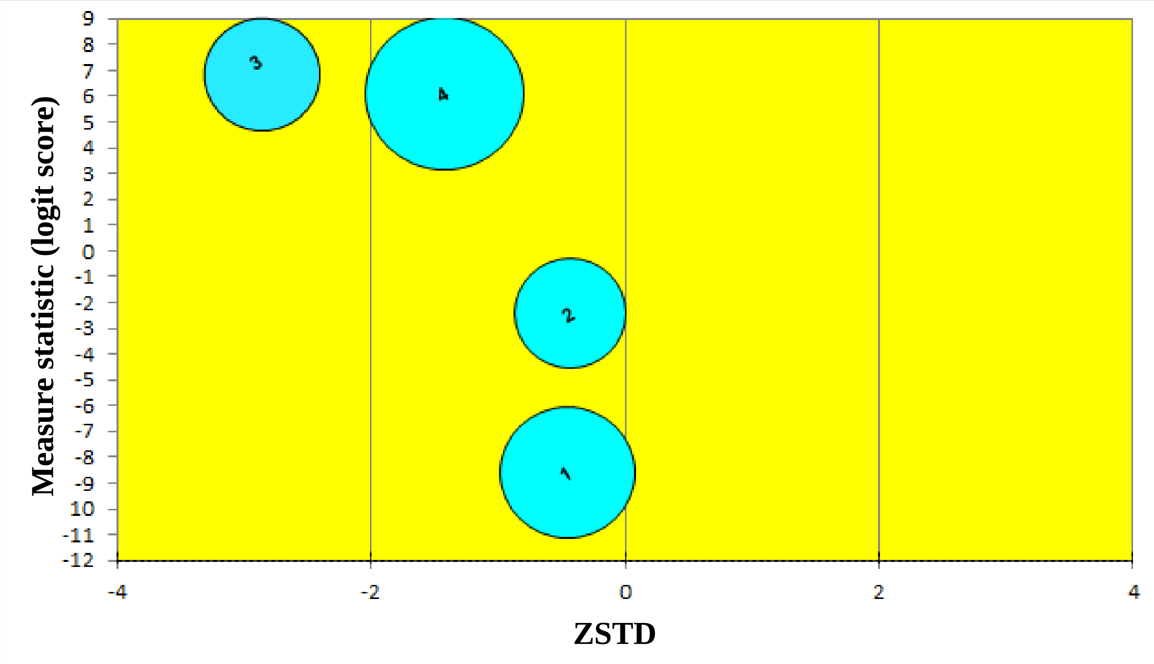


Phoneme Deletion test items by 'fit' and hierarchy of difficulty by the end of the school year

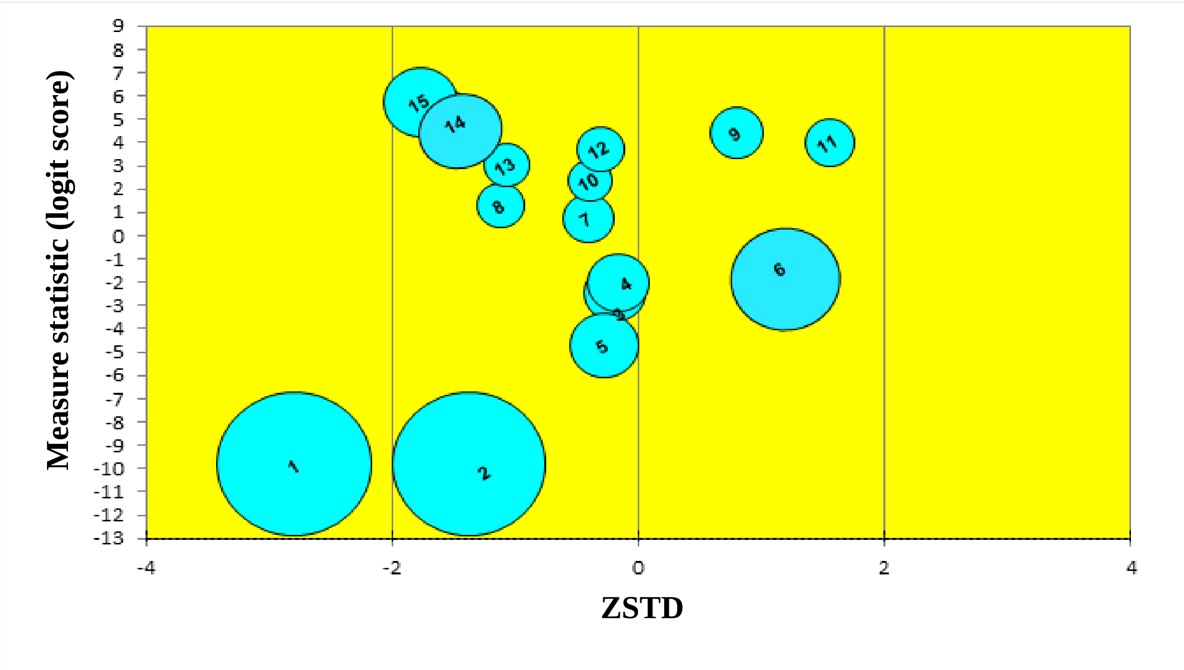


Phoneme Segmentation

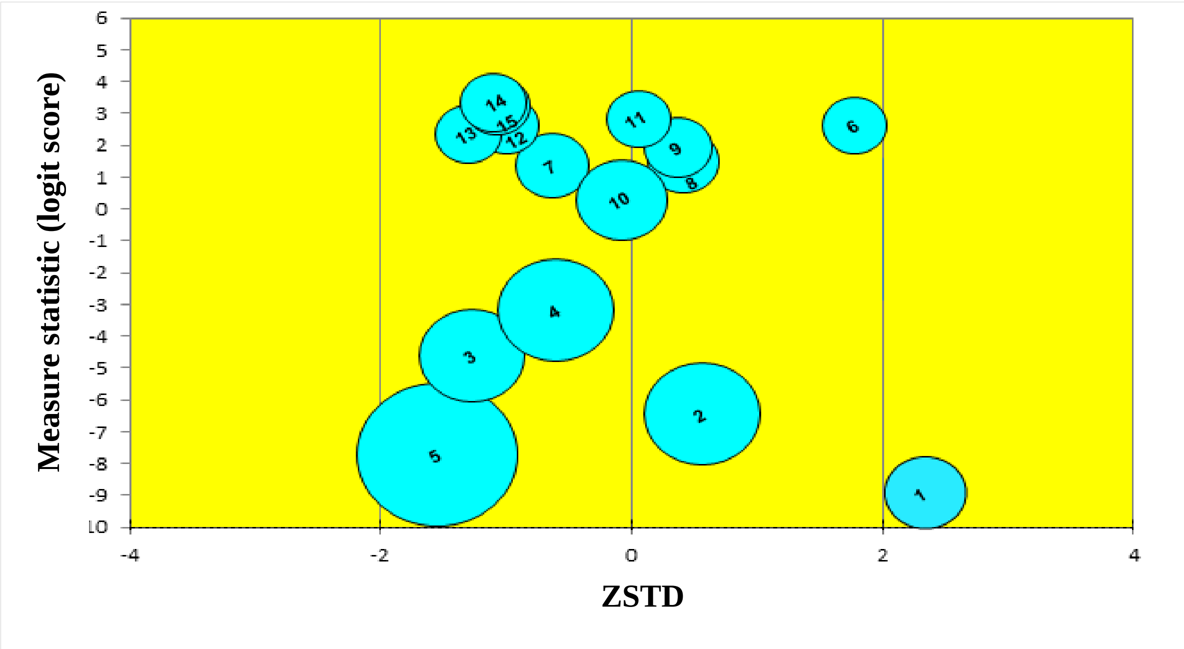
Phoneme Segmentation test items by 'fit' and hierarchy of difficulty at school-entry



Phoneme Segmentation test items by 'fit' and hierarchy of difficulty by the middle of the school year



Phoneme Segmentation test items by 'fit' and hierarchy of difficulty by the end of the school year



Appendix F: Examples of the Computer-Based Phonological Awareness Screening and Monitoring Tool in Use [Compact-Disc]

A compact-disc of test item examples from the computer-based screening and monitoring tool is provided in this appendix. The information presented on the compact-disc is divided into Part A and Part B as follows:

Part A: Computer-Based Phonological Awareness Screening and Monitoring Tool Examples. This video clip profiles examples of test items using 'camtastia' software to capture the information that is presented on the computer screen. In this video clip you will see the cursor moving over and clicking on multiple-choice response options.

Part B: A child completing test items in the Computer-Based Phonological Awareness Screening and Monitoring Tool. In this video clip you will see examples of a child completing a selection of test items from the computer-based screening and monitoring tool.