



Provided by the author(s) and University of Galway in accordance with publisher policies. Please cite the published version when available.

Title	An investigation of a precision teaching intervention programme as a tier 2 intervention targeting foundational reading skills with at risk readers
Author(s)	Brosnan, Julie
Publication Date	2015-02-06
Item record	http://hdl.handle.net/10379/5305

Downloaded 2024-05-18T03:46:22Z

Some rights reserved. For more information, please see the item record link above.



**An Investigation of a Precision Teaching Intervention Programme as a Tier 2
Intervention Targeting Foundational Reading Skills With At Risk Readers**

Thesis submitted for the Degree of Doctor of Philosophy

Julie A. Brosnan, B. A., MSc.



School of Psychology

National University of Ireland, Galway

Galway

February, 2015

Primary Supervisor:

Dr. Olive Healy

Co-Supervisor:

Dr. Kendra Brooks Newsome

TABLE OF CONTENTS

SUMMARY OF CONTENTS	i
ABSTRACT	iii
ACKNOWLEDGMENTS	iv
LIST OF TABLES	vi
LIST OF FIGURES	vii
LIST OF ACRONYMS	ix
Chapter 1: Introduction	12
1.1. At Risk Readers	12
1.2. Learning to Read in English	3
1.3. Critical Foundational Reading Skills	4
1.4. The Importance of Fluency in Foundational Reading Skills	7
1.5. Optimising Instruction: Harnessing Sublexical Fluency and Frequency Effects	8
1.6. Early Intervention: Targeting Foundational Reading Skills in Kindergarten	12
1.7. Response to Intervention	14
1.8. Precision Teaching	27
1.9. Instructional Intensity, Instructional Efficiency and Precision Teaching ...	49
1.10. Single Case Experimental Designs	55
1.11. Standardised Scores	57
1.12. The Current Research	59
Chapter 2: Introduction	63
2.1. Experiment 1: Letter Sounds	65
2.1.1. Method.....	68
2.1.2. Results and Discussion	77
2.1.3. Experiment 1: Conclusions	84
2.2. Experiment 2: Letter Names	85
2.2.1. Method.....	86
2.2.2. Results and Discussion	88
2.2.3. Experiment 2: Conclusions	95
2.3. Experiment 3: Sound Isolation	96
2.3.1. Method.....	97
2.3.2. Results and Discussion	100
2.3.3. Experiment 3: Conclusions	107
2.4. Experiment 4: Sound Deletion	108
2.4.1. Method.....	108
2.4.2. Results and Discussion	111
2.4.3. Experiment 4: Conclusions	119
2.5. Chapter 2: Results and Discussion	120
2.5.1. Generalised Outcomes In Reading Measures for Participants in Experiments 1, 2, 3, and 4	123
2.5.2. Social Validity Results.....	128
2.6. Chapter 2: General Limitations	129

Chapter 3: Introduction	134
3.1. Experiment 5: Letter Sounds	135
3.1.1. Method.....	135
3.1.2. Results and Discussion	137
3.1.3. Experiment 5: Conclusions	142
3.2. Experiment 6: Multi Letter Grapheme Phoneme Conversions	143
3.2.1. Method.....	143
3.2.2. Results.....	145
3.2.3. Experiment 6: Conclusions	148
3.3. Experiment 7: Blending Sounds	152
3.3.1. Method.....	153
3.3.2. Results and Discussion	157
3.3.3. Experiment 7: Conclusions	162
3.4. Chapter 3: Results and Discussion	165
3.4.1. Generalised Outcomes in Reading Measures for Participants in Experiments 5, 6, and 7.....	167
3.4.2. Partial and Complete Decoding: Generalisation Probes Across Experiments 5, 6 and 7.....	172
3.5. Social Validity	178
3.5.1. Parental Interview.....	178
3.5.2. Participant Interview	179
Chapter 4 – Introduction	180
4.1. Experiment 8: Letter Sounds	181
4.1.1. Method.....	183
4.1.2. Results and Discussion	185
4.1.3. Experiment 8: Conclusions	191
4.2. Experiment 9: Decoding (Segmenting and Blending Words)	192
4.2.1. Methods.....	193
4.2.2. Results.....	196
4.2.3. Experiment 9: Conclusions	198
4.3. Experiment 10: Blending Sounds into Words	199
4.3.1. Method.....	200
4.3.2. Results.....	202
4.3.3. Experiment 10: Conclusions	204
4.4. Results and Discussion	205
4.4.1. Generalised Outcomes in Reading Measures for Participants in Experiments 8, 9, and 10.....	206
4.4.2. Partial and Complete Decoding: Generalisation Probes Across Experiments 8, 9, and 10.....	210
4.5. Social Validity	214
Chapter 5: Introduction	216
5.1. Experiment 11	218
5.1.1. Method Condition 1: Decoding (Segmenting and Blending) Words.....	220
5.1.2. Condition 1: Results and Discussion.....	223
5.2.1. Method Condition 2: High Frequency Words	228
5.2.2. Condition 2: Results.....	231
5.2.3. Results and Discussion	236
5.3.1. Experiment 11: Conclusions	245
5.4. Social Validity	247
Chapter 6: General Discussion	249
6.1. Summary of Experiments	251
6.1.1. Experiment 1 (Chapter 2): Letter Sounds	251

6.1.2. Experiment 2 (Chapter 2): Letter Names	251
6.1.3. Experiment 3 (Chapter 2): Sound Isolation.....	252
6.1.4. Experiment 4 (Chapter 2): Sound Deletion	253
6.1.5. Experiment 5 (Chapter 3): Letter Sounds	254
6.1.7. Experiment 7 (Chapter 3): Blending Sounds	256
6.1.8. Experiment 8 (Chapter 4): Letter Sounds	256
6.1.9. Experiment 9 (Chapter 4): Decoding Words.....	257
6.1.10. Experiment 10 (Chapter 4): Blending Sounds.....	257
6.1.11. Experiment 11 (Chapter 5): Decoding Words and High Frequency Words	258
6.2. Summary of Generalised Reading Outcomes Across Experiments	261
6.2.1. Experiments 1, 2, 3, and 4 (Letter Sounds, Letter Names, Sound Isolation, and Sound Deletion)	261
6.2.2. Experiments 5, 6, and 7 (Letter Sounds, MLGPCs, and Blending Sounds)	262
6.2.3. Experiments 8, 9, and 10 (Letter Sounds, Decoding Words, and Blending Sounds).....	263
6.2.4. Experiment 11 (Decoding Words and High Frequency Words).....	264
6.3. Implications of the Generalised Reading Outcomes Resulting from the PT Intervention Programme	264
6.3.1. Pre- Post-Test Outcomes – At Risk Readers “Closing the Gap” with Average Peers	265
6.3.2. Generalisation Checks: Sublexical Fluency and Dysfluency Effects on Decoding Words.....	267
6.4. The PT Intervention Programme as an Effective Means for Establishing Sublexical Fluency (Mastery) in Foundational Reading Skills.....	271
6.5. The PT Intervention Programme as a Tier 2 Screening and Intervention Protocol for At Risk Senior Infant/Kindergarten Readers	274
6.6. The PT Intervention Programme for Progress Monitoring and Decision Making	276
6.6.1. The PT Intervention Programme as a Dynamic Assessment.....	279
6.8. Qualitative Outcomes of the PT Intervention Programme	280
6.9. Limitations and Directions for Future Research.....	281
6.10. Summary of Conclusions.....	289
REFERENCES.....	292
APPENDIX A.....	320
APPENDIX B.....	324
APPENDIX C	326
APPENDIX D.....	328
APPENDIX E	329
APPENDIX F	331
APPENDIX G.....	333
APPENDIX H.....	335
APPENDIX I	336

SUMMARY OF CONTENTS

The current research describes the implementation of a precision teaching (PT) intervention programme as a Tier 2 intervention targeting fluency in foundational reading skills with at risk Senior Infant readers ($n=36$). The current research was the first application of PT as a Tier 2 intervention that combines: (a) fluency intervention; (b) progress monitoring, and (c) decision rules for programme modifications. Chapter 1 provides a literature review in relation to the core topics of the thesis. Chapter 2 provides an account of four experiments (Exp. 1, letter sounds; Exp. 2, letter names; Exp. 3, sound isolation; and Exp. 4, sound deletion) conducted in the first setting, a rural English speaking school. Chapters 3 and 4 describe six experiments (Exp. 5, 6, and 8, letter sounds; Exp. 7 and 9, blending sounds, and Exp. 10 decoding words) conducted in the second setting, an urban English speaking school. Chapter 5 describes one experiment (Exp. 11, decoding and high frequency words) conducted in the third setting, a rural Irish speaking school. The first author implemented Experiments 1-10 under the supervision of the second and third authors, and a research assistant conducted Experiment 11 under the supervision of all authors. The findings of these Chapters indicate that the PT intervention was effective in establishing fluency in the foundational skills targeted, that it was an effective and efficient form of intervention resulting in at risk readers closing the gap with average performing peers.

Chapter 6 provides a general discussion of the overall findings within the individual chapters; it identifies limitations of the current research and makes suggestions for future research. The theoretical significance of findings for the role of sublexical fluency in overall reading development is described. Practical implications

regarding the importance of targeting fluency in foundational reading skills with at risk readers are also provided.

ABSTRACT

This thesis describes the first application of a Tier 2 intervention programme targeting fluency in foundational reading skills with at risk Senior Infant readers ($n=36$). The PT intervention programme combined: (a) fluency intervention; (b) progress monitoring, and (c) decision rules for programme modifications. The effectiveness and efficiency of the PT intervention programme were evaluated in 11 Single Case Experimental Designs (SCED) that incorporated pre- post-test outcomes of reading measures. These experiments investigated the effects of the PT intervention on four foundational reading skills: letter sounds and names (Experiments 1, 2, 5, 6 and 8); phonemic awareness (Experiments 3, 4, 7, and 10); decoding words (Experiments 9 and 11); and high frequency words (Experiment 11). Results of the SCED indicate that the PT intervention was effective in establishing fluency in the foundational skills targeted. Pre- post-test outcomes indicated that the PT intervention was an effective and efficient form of intervention that helped at risk readers to close the gap with average performing peers. The findings have theoretical significance for the role of sublexical fluency in overall reading development, and practical implications regarding the importance of targeting fluency in foundational reading skills with at risk readers.

Keywords: Precision teaching, response to intervention, sublexical fluency, tier 2 intervention, foundational reading skills, at risk readers

ACKNOWLEDGMENTS

This thesis was the most challenging and rewarding experience of my life (after you, Leigh!). Thanks to the 36 Senior Infant students that I worked with over two years: characters that continue to inspire me, and will stay with me always. Thanks to the teachers that welcomed a stranger with charts, timers, and talk of measurement. I am eternally grateful.

Precision Teaching, it simply rocks. This thesis came to fruition through the overwhelming support from the precision teaching community, in particular, my supervisor Dr. Kendra Brooks Newsome. Your expertise, mentorship, and passion for science in education shaped the project and my own path within it. I cannot thank you enough.

I extend equal gratitude and respect to my supervisor Dr. Olive Healy. You have been a great support to me for many years, and I am lucky to have you as a mentor. Thank you for helping me to find my path, and then encouraging me along it.

I am very grateful to Dr. Pat Morgan for providing me with guidance and practical support when I most needed it. I would not be here only for you, and I will never forget it.

To my family: well done for putting up with me. And, thank you! I am finally here. My beautiful daughter Leigh Brosnan, you have experienced me in education for 16 of your 21 years on this earth. Thank you for your never-ending encouragement. You are blazing the academia trail now; I am so proud of you. Thanks to Mary Brosnan for the practical support that being in education this long entails. And to Christina Brosnan for listening to me wax lyrical on the merits of precision teaching. To Barry Foley, Lulu Sinnott, and Clare Sinnott Foley - that

combined are a 3rd level education factory: thanks for all the academic interactions and inspirations.

Thanks to Brendan Reddy for being there at the beginning, and the end; and to Ann Marie Flanagan for the bits in between. Thanks to all my friends, the support has been tremendous. A final thanks to house music, and to my mini-rig, essential cogs in the wheel.

This work was supported by doctoral scholarship from the Irish Research Council's EMBARK Postgraduate Scholarship Scheme [RS/2010/946], and Galway University Foundation Funding by the National University of Ireland.

LIST OF TABLES

Table 1. <i>Pre-Post-Test Outcomes in YARC Letter Sound Knowledge (LSK) Subtest for Participants that Received the PT Intervention in Single Letter Sound (Experiment 1)</i>	81
Table 2. <i>Pre-Post-Test Outcomes in DIBELS Letter Naming Fluency (LNF) Subtest for Participants that Received the PT Intervention in Letter Names (Experiment 2)</i>	92
Table 3. <i>Pre-Post-Test Outcomes in YARC Sound Isolation (SI) Subtest for Participants that Received the PT Intervention in Sound Isolation (Experiment 3)</i>	103
Table 4. <i>Pre-Post-Test Outcomes in YARC Sound Deletion (SD) Subtest for Participants that Received the PT Intervention in Single Letter Sound (Experiment 4)</i>	116
Table 5. <i>Participant Characteristics, and Participant Identification Across Experiments 1, 2, 3 and 4</i>	122
Table 6. <i>Outcomes in the YARC Early Word Reading Subtest Across Experiments 1, 2, 3, and 4</i>	126
Table 7. <i>Outcomes in the Curriculum Based Measure Letter Sound Fluency (LSF) and DIBELS Nonsense Word Fluency (NWF) Subtest Across Experiments 1, 2, and 3, and 4</i>	128
Table 8. <i>Pre-Post-Test Outcomes in YARC Letter Sound Knowledge (LSK) Subtest for Participants That Received the Intervention in Single Letter Sounds, or Single Letter Sounds and Multi Letter Grapheme Phoneme Conversions</i>	152
Table 9. <i>Gains in CTOPP Blending Words (BW) Subtest for Participants That Received the Intervention in Blending Sounds</i>	160
Table 10. <i>Participant Characteristics, and Participant Identification Across Experiments 5, 6, and 7</i>	166
Table 11. <i>Outcomes in YARC Early Word Reading (EWR) Subtest for Experiments 5, 6, and 7</i>	169
Table 12. <i>Outcomes in Words Read Correctly on the Recoded DIBELS Nonsense Word Fluency Subtest (NWF) and Real Word CVC Probe for Experiments 5, 6, and 7</i>	171
Table 13. <i>Outcomes in YARC Letter Sound Knowledge (LSK) and CBM Letter Sound Fluency (LSF) Subtests</i>	188
Table 14. <i>Participant Characteristics, and Participant Identification Across Experiments 8, 9, and 10</i>	206
Table 15. <i>Outcomes in YARC Early Word Reading and WIAT-II Pseudoword Decoding Subtests Across Experiments 8, 9, and 10</i>	208
Table 16. <i>Outcomes in Words Read Correctly on the Recoded DIBELS Nonsense Word Fluency Subtest (NWF) and Real Word CVC Probe Across Experiments 8, 9, and 10</i>	209
Table 17. <i>Outcomes in Letter Sounds, Word Reading, and Pseudoword Decoding</i>	238
Table 18. <i>Outcomes in Words Read Correctly on the Recoded DIBELS Nonsense Word Fluency Subtest (NWF) and Real Word CVC Probe</i>	240

LIST OF FIGURES

<i>Figure 1. Response to Intervention (RTI) Framework with Three Tiers of Increasing Instructional Intensity (Tier 1, Tier 2, and Tier 3).</i>	17
<i>Figure 2. The Instructional Hierarchy.</i>	26
<i>Figure 3. Description of a Typical PT Session Targeting Letter Sounds.</i>	32
<i>Figure 4. Standard celeration chart (as featured in Kubina & Yurich, 2012).</i>	34
<i>Figure 5. Three add-subtract charts displaying the same data, but different vertical axis scales: the same is data displayed on a standard celeration chart.</i>	39
<i>Figure 6. Real world Example of Reporting Effectiveness and Efficiency Data Combined with Qualitative Information.</i>	54
<i>Figure 7. Experiment 1: multiple-baseline across participants design for the PT intervention in letter sounds.</i>	79
<i>Figure 8. Experiment 1, maintenance data displaying performance in letter sounds at median baseline, intervention end and five post-intervention checks: 1 week; 2 weeks; 4-6 weeks; 8-10 weeks, and 34-38 weeks.</i>	83
<i>Figure 9. Experiment 2: multiple-baseline across participants design for the PT intervention in letter names.</i>	90
<i>Figure 10. Experiment 2, maintenance data displaying performance in letter names at median baseline, intervention end and 5 post-intervention checks: 1 week; 2 weeks; 4-6 weeks; 8-10 weeks, and 34-38 weeks.</i>	94
<i>Figure 11. Experiment 3: multiple-baseline across participants with probes across skills design for the PT intervention in sound isolation.</i>	102
<i>Figure 12. Experiment 3, maintenance data displaying performance in sound isolation at median baseline, intervention end and five post-intervention checks: 1 week; 2 weeks; 4-6 weeks; 8-10 weeks, and 34-38 weeks.</i>	106
<i>Figure 13. Experiment 4: multiple-baseline across participants with multiple probes across skills design for the PT intervention in sound deletion.</i>	114
<i>Figure 14. Experiment 4, maintenance data displaying performance in sound deletion at median baseline, intervention end and five post-intervention checks: 1 week; 2 weeks; 4-6 weeks; 8-10 weeks, and 34-38 weeks.</i>	118
<i>Figure 15. Experiment 5: multiple-baseline across participants design for the PT intervention in letter sounds.</i>	139
<i>Figure 16. Experiment 5, maintenance data displaying performance in letter sounds at median baseline, intervention end and five post-intervention checks: 1 week; 2 weeks; 4-6 weeks; 10-12 weeks, and 26-28 weeks.</i>	141
<i>Figure 17. Experiment 6: multiple-baseline across participants design for the PT intervention in multi-letter grapheme phoneme conversions.</i>	146
<i>Figure 18. Experiment 6 maintenance data displaying performance in multi-letter grapheme phoneme conversions at median baseline, intervention end and five post-intervention checks: 1 week; 2 weeks; 4-6 weeks; 10-12 weeks, and 26-28 weeks.</i>	148
<i>Figure 19. Experiment 7: multiple-baseline across participants design for the PT intervention in blending sounds.</i>	159
<i>Figure 20. Experiment 6, maintenance data displaying performance in blending sounds at median baseline, intervention end and five post-intervention checks: 1 week; 2 weeks; 4-6 weeks; 10-12 weeks, and 26-28 weeks.</i>	162

<i>Figure 21. Generalisation outcomes for participants that receive the PT intervention for letter sounds. Generalisation checks for partial and complete decoding administered within and across baseline and intervention conditions.</i>	<i>174</i>
<i>Figure 22. Generalisation outcomes for participants that receive the PT intervention for letter sounds and multi-letter grapheme phoneme conversions (MLGPCs). Generalisation checks for partial and complete decoding administered within and across baseline and intervention conditions.</i>	<i>176</i>
<i>Figure 23. Generalisation outcomes for participants that receive the PT intervention for letter sounds and blending sounds. Generalisation checks for partial and complete decoding administered within and across baseline and intervention conditions.</i>	<i>177</i>
<i>Figure 24. Experiment 8: multiple-baseline across participants design for the PT intervention in letter sounds.</i>	<i>186</i>
<i>Figure 25. Experiment 8, maintenance data displaying performance in letter sounds at median baseline, intervention end and three post-intervention checks: 1 week; 2 weeks, and 14-15 weeks.</i>	<i>190</i>
<i>Figure 26. Experiment 9: multiple-baseline across participants design for the PT intervention in decoding words.</i>	<i>197</i>
<i>Figure 27. Experiment 10: multiple-baseline across participants design for the PT intervention in blending sounds.</i>	<i>203</i>
<i>Figure 28. Generalisation outcomes for participants that receive the PT intervention for letter sounds only, and participants that received the PT intervention in letter sounds and blending sounds. Generalisation checks for partial and complete decoding administered within and across baseline and intervention conditions.</i>	<i>212</i>
<i>Figure 29. Generalisation outcomes for participants that receive the PT intervention for letter sounds and decoding words. Generalisation checks for partial and complete decoding administered within and across baseline and intervention conditions.</i>	<i>213</i>
<i>Figure 30. Experiment 11, Condition 1: multiple-baseline across participants with multiple probes across word sets design for the PT intervention in decoding words.</i>	<i>225</i>
<i>Figure 31. Experiment 11, Condition 1 - maintenance data displaying performance in decoding words at median baseline, intervention end and three post-intervention checks: 1 week; 2 weeks, and 3 weeks.</i>	<i>228</i>
<i>Figure 32. Experiment 11, Condition 2: multiple-baseline across participants with multiple probes across word sets design for the PT intervention in high frequency words.</i>	<i>233</i>
<i>Figure 33. Experiment 11, Condition 2: maintenance data displaying performance in high frequency words at median baseline, intervention end and three post-intervention checks: 1 week; 2 weeks, and 3 weeks.</i>	<i>236</i>
<i>Figure 34. Generalisation outcomes for participants that receive the PT intervention for decoding words. Generalisation checks for partial and complete decoding administered within and across baseline and intervention conditions.</i>	<i>243</i>
<i>Figure 35. Generalisation outcomes for participants that receive the PT intervention for high frequency words. Generalisation checks for partial and complete decoding administered within and across baseline and intervention conditions.</i>	<i>244</i>

LIST OF ACRONYMS

BW: Blending words

CBM: Curriculum based measurement

CEM: Curriculum embedded mastery/measurement

CPM: Correct per minute

CVC: Consonant-vowel-consonant

CTOPP: Comprehensive test of phonological awareness

CW: Compound words

DEIS: Delivering equality of opportunity in schools

DIBELS: Dynamic indicators of basic early literacy skills

DPM SCC: Daily per minute standard celeration chart

EL1: English as a first language

EL2: English as a second language

EWR: Early word reading

FP: Final phoneme

GPCs: Grapheme phoneme conversions

HFW: High frequency words

IP: Initial phoneme

IOA: Inter-observer agreement

IR: Incremental rehearsal

LNF: Letter naming fluency

LSF: Letter sound fluency

LSK: Letter sound knowledge

MBD: Multiple baseline design

MLGPCs: Multi-letter grapheme phoneme conversions

MS: Multi-syllabic words

NCRTI: National center on response to intervention

NW: Nonsense words

NWF: Nonsense word fluency

NWF(R): Nonsense word fluency – recoded

OECD: Organisation for economic co-operation and development

PBC: Personal best component

PM: Programme modification

PI: Procedural integrity

PISA: Programme for international student assessment

PT: Precision teaching

PWD: PseudoWord decoding

RBSP: Random between stimulus prompt

R-CBM: Word reading curriculum based measurement

REAPS, Retention, endurance, stability, performance standards

RG: Rate gain

RRG: Rate ratio gain

RTI: Response to intervention

RW: Real words

SAFMEDS: Say all fast minute everyday shuffled

SCC: Standard celeration chart

SCED: Single case experimental designs

SD: Sound deletion

SI: Sound isolation

SSG: Standard score gain

SSRG: Standard score ratio gain

TBL: True baseline

TOY: The one-year chart (electronic standard celeration chart)

TPM SCC: Timings per minute standard celeration chart

VC: Vowel-consonant

WIAT-II: Wechsler individual attainment test, 2nd edition

WRC: Words read correctly

YARC: York assessment of reading comprehension

Chapter 1: Introduction

1.1. At Risk Readers

Reading failure is a global concern. Some 40 out of the 64 countries in the Programme for International Student Assessment (PISA) exhibited a below average share of low achievers in reading (The Organisation for Economic Co-operation and Development; OECD, 2014); illustrating that reading failure is comparatively concentrated in specific countries. Reading failure is also unequally distributed within countries. It is concentrated in schools serving disadvantaged, minority, and limited English proficient children (Slavin, Lake, Davis, & Madden, 2011). Reading difficulty is also proportionally larger for boys than girls (Brooks, 2007; OECD, 2014). For example, in the U.S. two thirds of special education service is delivered to males identified as learning disabled (Cortilla, 2011). Therefore, students can be considered at risk of reading failure if originating from socio- or economically disadvantaged groups including second language learners, or if performance is below the 50th percentile on standardised assessments (Suggate, 2014). Prevalence rates in the U.S. show that in 2011, 33% of 4th grade students, and 26% of 12th grade students, performed below the basic proficiency level in reading (Hemphill & Vanneman, 2011). Schools in Ireland categorised as “disadvantaged” report that 30% of the students score below the 10th percentile on nationally standardised tests (Eivers, Shiel, & Shortt, 2004).

Reading failure has many implications, some quantifiable in simple economic terms, others that result in longer-term social and socio-economic consequences. Substantial costs are incurred in terms of special education, remediation programmes,

grade repetition, and school dropout (Slavin et al., 2011). In the United Kingdom, the reported estimated annual cost of students' failure to master basic literacy skills in primary school is approximately £2.3 billion (Gross, Hudson, & Price, 2009). In the U.S., the cost of educating students with learning disabilities is 1.6 times that of a general education student (Cortilla, 2011).

These “measureable” costs are but financial, other costs to the individual may be less visible - yet no less important. For example, students who fail to acquire basic literacy skills may also exhibit behavioural problems, which increase negative teacher interactions and may impact academic outcomes (Moore Parten, Robertson, Maggin, Oliver, & Wehby, 2010). Such behaviour problems can present as early as kindergarten (Fergusson & Lynskey, 1997) and may extend into later grades (Gellert & Elbro, 1999). Studies have reported that students who struggle to read report feeling less support from teachers (e.g., Patrick, Mantzicopoulos, Samarapungavan, & French, 2008), and spend the majority of general reading instruction in passive or independent learning tasks without teacher assistance (e.g., Wanzek, Roberts, & Al Otaiba, 2014).

A good start in reading is therefore crucially important, as early success in reading is synonymous with early success in school (Slavin et al., 2011), and in predicting a lifetime of literacy experience (Cunningham & Stanovich, 2001). Nevertheless, prevalence rates of reading failure in English language countries, and its disproportional representation among certain groups, may indicate that learning to read written English text is not a straightforward process for all students.

1.2. Learning to Read in English

Learning to read involves learning how one's writing system works (Perfetti & Marron, 1998). The writing system encodes language, quintessentially: "reading is the cognitive process of understanding speech that is written down" (Mol & Bus, 2011, pp. 267). All contemporary writing systems connect to the spoken language; however systems differ in the speech units that they map on to, and this has implications for how reading is acquired and taught (Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001). For example, English is an alphabetic writing system in which graphic units (graphemes) are related to phonemes. A grapheme is the smallest intact linguistic unit, and a phoneme is the smallest speech sound segment. In other words, graphemes are printed letters and phonemes are their associated sounds; the specific association is known as a grapheme-phoneme correspondence (GPC). Graphemes can be represented as single or multiple letters; multiple letter graphemes have been referred to as multi-letter grapheme phoneme conversions (MLGPCs; Solity & Vousden, 2009). A single grapheme may represent several phonemes (e.g., *ch* as in *chord* and *chat*), and several graphemes may represent a single phoneme (e.g., *ea* pronounced in *easy*, and *ee* pronounced in *bleed* for the phoneme /i/).

Because of these multiple mappings, the alphabet is considered to be both economic and productive i.e., 26 letters can be configured to write an exponential number of words (Rayner et al., 2001). The 26 letters in the alphabet (single or MLGPCs) represent 195 graphemes, and map onto 461 different sounds (Gontijo, Gontijo, & Shilcock, 2003). Written English can therefore be considered to demonstrate economy (only 26 letters), at the cost of complexity (one letter/letter configuration maps to many sounds; Rayner et al., 2001).

Learning to read in English therefore involves acquiring myriad mappings between graphemes and phonemes; this is referred to as the “alphabetic principle” (Adams, 1990; Snowling & Hulme, 2010). Providing students with a productive knowledge of GPCs conveys the basic alphabetic principle (Adams, 1990). Due to the multiple mappings and the apparently meaningless and somewhat arbitrary nature of GPCs however, mastering the alphabetic principle of English can be an arduous task for the beginning reader (Rayner et al., 2001).

The process of reading, or decoding, involves extracting the correct GPCs present in a word, and then recoding them back together to read the word as a whole unit. Therefore, the minimal requisites for decoding behaviour are a command of GPCs, and the ability to blend those sounds together to make words whole. Awareness of and ability to manipulate phonemes is known as “phonemic awareness”. Letter sounds and phonemic awareness are recognised as the two foundational reading skills critical for decoding (Snowling & Hulme, 2010).

1.3. Critical Foundational Reading Skills

Letter sound knowledge is a strong predictor of beginning reading (e.g., Evans, Bell, Shaw, Moretti, & Page, 2006; Leppänen, Aunola, Niemi, & Nurmi, 2008; McBride-Chang, 1999; Pennington & Lefly, 2001). Experimental studies have also demonstrated a causal relationship between letter sound knowledge and reading achievement (e.g., de Graaff, Bosman, Hasselman, & Verhoeven, 2009; Piasta, Purpura, & Wagner, 2010). Predictably, failure to master letter sounds can result in the student becoming at risk for reading difficulties (e.g., Gersten et al., 2009; Snowling & Hulme, 2010).

The importance of letter sound knowledge to reading can be observed in the effectiveness of phonics and subsequent reading outcomes. Converging evidence shows that phonics instruction is highly effective in early reading instruction (e.g., de Graaff et al., 2009; National Reading Panel, 2000). Because letter sound knowledge is necessary for accurately decoding words, it is considered to provide the basis for phonics instruction (Huang, Tortorelli, & Invernizzi, 2013). Phonics refers to a system of teaching that builds on the alphabetic principle however, instructional procedures and materials vary widely. Analytic, analogy, embedded, synthetic, and phonics through spelling are the main phonic instructional approaches (Adams, 1990). Each of these approaches focus on letter sound correspondences, but what distinguishes each one is the instructional focus and unit of analysis. For example, in analytic phonics, students are taught to analyse letter sounds in previously known words, and sounds are not practiced in isolation. Conversely, in synthetic phonics the student is explicitly taught to convert letters into sounds, and then to blend the sounds together to make a word. Blending sounds together to make a word is a phonemic awareness skill that is critical for accurate decoding (Snowling & Hulme, 2010).

Phonemic awareness refers to awareness of the smallest speech segment, the phoneme, whereas phonological awareness refers to awareness of larger speech segments such as syllables. A recent meta-analytic review by Melby-Lervåg, Halaas Lyster, and Hulme (2012) concluded that deeper phonological skills (i.e., phoneme level) appear to be more closely related to reading achievement than shallow phonological skills (i.e., larger units such as syllables). For example, Savage and Carless (2005) demonstrated that explicit phoneme manipulation skills at five years, predicted reading outcomes measured at seven years, while onset-rime manipulation did not. In other research, Hatcher, Hulme, and Snowling (2004) showed that for

young children at risk for reading delay, additional training in phoneme awareness and linking phonemes was beneficial whereas additional training in rhyme was not.

Phonological awareness is believed to emerge in a developmental nature in typically developing children (Adams, 1990; Blaiklock, 2004; Philips, Clancy-Menchetti, & Lonigan, 2008; Pufpaff, 2009). However, phonemic awareness skills are not necessarily a naturally developing ability, but may require explicit teaching and practice opportunities (Philips et al., 2008). This is evidenced in the findings that differential performance in phonological skills persists between children from lower socioeconomic groups and their more affluent peers (Hecht, Burgess, Torgesen, Wagner, & Raschotte, 2000). In addition, children with speech and language difficulties have been shown to exhibit deficits in these skills (Snowling & Hulme, 2012).

Phonemic awareness skills and letter sounds are predictive of reading ability, and an intervention specifically targeted at these skills should be chosen for students who are struggling to master decoding skills (Snowling & Hulme, 2010). Learning to decode print however, is necessary but not sufficient, in learning to read. The development of decoding exists on a continuum ranging from slow and effortful to speedy, accurate, and effortless. Children begin learning to read by acquiring these basic decoding skills, and then applying them with greater accuracy and speed (Adams, 1990; Verhoeven & van Leeuwe, 2009). This shift marks the progression from the beginning reader to the proficient reader. Proficiency in reading is wholly dependent on the accurate and fluent decoding that conveys understanding (Snowling & Hulme, 2010), beginning with the accurate and fast retrieval of GPCs (Verhoeven & van Leeuwe, 2009).

1.4. The Importance of Fluency in Foundational Reading Skills

Phonemes and graphemes are the constituents of words; and awareness of these units and their associations have been described as *sublexical* skills (Ritchey & Speece, 2006). *Lexical* refers simply to word level, while *sub* refers to below the word level. Phonemic awareness and alphabet knowledge are considered precursors to and initially independent of word knowledge and therefore they can be considered sublexical. Sublexical fluency combines phonological and alphabetical knowledge with automaticity (Burke, Hagan-Burke, Zou, & Kwok, 2010; Ritchey & Speece, 2006). Automaticity refers to fast, effortless, and accurate retrieval of information that does not require conscious control or attention (Lagerge & Samuels, 1974).

The importance of sublexical fluency in promoting reading acquisition has gained increasing attention (Burke et al., 2010; Coyne, Kame'enui, & Simmons, 2001; Hudson, Pullen, Lane, & Torgesen, 2009; Hudson, Torgesen, Lane, & Turner, 2010; Kairaluoma, Ahonen, Aro, & Holopainen, 2010; Katzir et al., 2006; Ritchey & Speece, 2006; Wolf & Bowers, 1999; Wolf & Katzir-Cohn, 2001). Just as fluency in word recognition is thought to free up cognitive resources for higher order skills such as comprehension (LaBerge & Samuels, 1974; National Reading Panel, 2000; Verhoeven & van Leeuwe, 2009), fluency in sublexical skills may free up cognitive resources for successful word decoding (Burke et al., 2010; Hudson et al., 2010; Katzir et al., 2006; Ritchey & Speece, 2006). Decoding depends on rapid and accurate access to GPCs, and the ability to blend the sounds together (Hudson et al., 2009). Burke and colleagues (2010) have demonstrated the importance of phonological fluency and alphabetic fluency in relation to emergent reading ability.

Conversely, slow access and production of these skills (referred to as “dysfluency”), will limit word-decoding ability (Ritchey & Speece, 2006). If letters and their corresponding sounds are not identified automatically, decoding will be less efficient (Wolf & Bowers, 1999). Hudson and colleagues (2009) showed that dysfluency in sublexical skills of phonemic blending and letter sounds had the potential to affect the development of proficiency in reading. These authors conclude with an educational implication; students must build fluency and automaticity in each sublexical skill to become skilled readers. Despite the highlighted benefits of fluency, and the potential problems associated with sublexical dysfluency, fluency in core reading skills is infrequently targeted in the early grades (Bursock & Blanks, 2010). Burke and colleagues (2010) assert that alphabetic fluency should be a target for intervention as early as kindergarten.

In addition to an increased focus on fluency in core reading skills in kindergarten, early reading instruction could be maximised through careful consideration of the reading skills that will yield the largest gains. This may be considered the optimal skill set to equip students in learning to read.

1.5. Optimising Instruction: Harnessing Sublexical Fluency and Frequency

Effects

The theory of optimal instruction asserts that there is an optimal amount of information to teach that will result in maximising generalisation (Solity & Vousden, 2009). For example, not all 461 GPCs are equally useful; therefore not all GPCs need to be taught in early reading instruction (Adams, 1990). Frequency analyses of spelling-to-sound units and words have revealed the type of linguistic unit and set of words that will optimise early reading instruction (Solity & Vousden, 2009; Stuart,

Dixon, Masterson, & Gray, 2003; Vousden, 2008; Vousden, Ellefson, Solity, & Chater, 2011).

Vousden (2008) proposes a framework for understanding how humans exploit statistical properties of language, thus creating an optimal solution to reading in English. For example, adult readers are sensitive to the frequency distribution of words, with high frequency words demonstrating an advantage over lower frequency words in many language and memory-based tasks. In addition, low frequency words that comprise low frequency grapheme phoneme mappings take longer to name than words with more frequent mappings. Therefore, Vousden (2008) proposes that adult readers' cognitive processing reflects the statistical structure of the English language at various levels (e.g., GPC, whole word). The author cites evidence from language acquisition with infants to demonstrate that infants are also sensitive to word frequency, and that early language acquisition consists of implicitly extracting the statistical regularities of speech. Vousden uses this converging evidence to highlight the importance of designing curricula that capitalises on these statistical features. For example, targeting GPCs that occur frequently are more likely to be retained due to the regularity with which they will be encountered, and therefore be of greater use in decoding larger amounts of text.

A number of sources (e.g., Carnine, Silbert, & Kame'enui, 1997; Solity & Vousden, 2009; Vousden, 2008; Vousden et al., 2011) have identified the critical high frequency GPCs that enable students to read the majority of phonically regular and irregular words that they encounter in children's literature. Solity and Vousden (2009) showed that only 64 of the possible 461 GPCs enable a student to read a large proportion of monosyllabic words encountered in both adult and children's literature. These authors conclude that teaching additional GPCs is of limited value due to their

relative infrequency, and therefore minimal impact on reading development. Thus, it follows that only these frequently occurring GPCs should be taught in beginning reading instruction. Moreover, automatic recognition of the critical high frequency GPCs would facilitate proficiency in decoding. Fluency in this subset of GPCs therefore seems a logical focus for early reading instruction.

Similar frequency effects demonstrated by GPCs are evident for word reading. Stuart and colleagues (2003) analysed the vocabulary in 685 books drawn from a variety of reading schemes and storybooks read by 5-7 year olds. They found that the 100 most frequently occurring words, which represented just over 1% of all word types, or unique words, accounted for 54.1% of all words encountered. In contrast, 51% of word types appeared only once or twice and accounted for just 2.4% of all words encountered. More recently, Solity and Vousden (2009) conducted frequency analyses on the structure of adult literature, children's reading schemes and children's storybooks. Their findings mirrored those of Stuart et al. (2003); 100 high frequency words (HFW) accounted for over 50% of words in across the three types of literature.

Solity and Vousden (2009) concluded that such findings provide a theoretically motivated rationale for teaching these HFW to automaticity. If these high frequency words can be recognised automatically, then attention and cognitive resources are freed up, to comprehend greater than 50% of the text that such high frequency words account for. Furthermore, if these high frequency words are recognised automatically, attention can be diverted to recognising and comprehending the surrounding words in the sentence. In other words, automaticity in high frequency words speeds up the process of decoding a large proportion of text.

Clarke (1994; 2013) has espoused the importance of teaching the most HFW during the last 20 years. Despite their critical importance, "Dolch" and other word

lists are normally employed. A cross reference between Solity and Vousden's (2009) 100 HFW, and the Dolch lists (11 word lists in total containing 220 words), revealed that the first six Dolch word lists contain 82 of the 100 HFW and 38 less frequent words. This is a 2:1 ratio; for every two HFW learned from the Dolch lists, one less frequent word must also be learned. The subsequent five Dolch lists contain a further 12 HFW, but to acquire these, 88 less frequent words word must also be learned.

Learning less frequent words before HFW may be considered an inefficient use of valuable instructional time. Initially, there is the time taken to learn the words, which by virtue of their infrequency are unlikely to be encountered in reading material and therefore less practiced (Stuart et al., 2003). Subsequently, it is improbable that such words are retained, and must therefore be taught again at a later date. In word learning experiments with five year olds, Stuart, Masterson, and Dixon (2000) demonstrated that for words to be retained, the child must encounter them at least 36 times in text. Additional analyses (Stuart et al., 2003) revealed that 97.7% of word types appear fewer than 36 times in text. Such a finding highlights the critical value of teaching the 100 high frequency words that facilitates access to 54% of reading material encountered. In Solity and Vousden's (2009) words: "An optimal system retains what occurs frequently, because it is seen to be useful, and forgets what occurs infrequently, because it is seen to be less important" (pp. 475).

As letter sounds and phonemic awareness skills are taught in kindergarten it seems logical to intervene at this early stage. Kindergarten can be the first opportunity that schools have to provide preventative reading instruction for students to achieve sufficient reading growth (Wanzek et al., 2014). Moreover, research clearly shows that to achieve long-term and meaningful outcomes, it is optimal for children to succeed in reading by the 1st grade (Cunningham & Stanovich, 1997; Sparks, Patton,

& Murdoch, 2014). To achieve proficiency in word and connected text reading in 1st grade, the component skills of reading need to be automatic, as do high frequency words. It follows that targeting fluency in foundational skills of letter sounds, phonemic awareness, and high frequency words in kindergarten optimises instruction, and lays the groundwork for subsequent success in reading.

1.6. Early Intervention: Targeting Foundational Reading Skills in Kindergarten

A multitude of evidence shows that early intervention (prevention) is more effective than later remediation (e.g., Torgesen, 2000; 2004; Torgesen & Hudson, 2006; Vaughn & Wanzek, 2014), and this is especially true for reading fluency, as it has direct and long-term implications. Moreover, early success in reading creates a positive cycle that persists throughout education and has significant implications for cognitive development (Cunningham & Stanovich, 2001).

1.6.1. Prevention is More Effective Than Remediation. Explicit and systematic instruction that focuses on core reading skills (e.g., phonemic awareness and the alphabetic principle) can prevent reading difficulties for many students (Torgesen, 2000; Vellutino, Scanlon, & Lyon, 2000). Conversely, Torgesen (2004; 2005) reported that although remediation studies have demonstrated improvements in word reading and comprehension, outcomes for fluency across all studies reviewed have been limited. Students who demonstrate persistent reading difficulties beyond the 1st Grade have a large practice deficit relative to their peers; struggling readers read less and therefore have limited opportunity to develop automaticity in word reading (Torgesen, 2004; 2005). Moreover, deficient decoding skills and lack of practice render many reading materials as difficult, this results in unrewarding reading

experiences that consequently lead to avoidance of practice and less involvement in reading activities (Cunningham & Stanovich, 2001).

Long-term deficits in reading fluency have been explained in part by the longitudinal research that investigated reading success, reading behaviours and their reciprocal relation to other academic outcomes. The following section will describe these longitudinal studies.

1.6.2. Early Success in Reading Positively Impacts Future Reading and Cognitive Development. In their seminal research, Cunningham and Stanovich (1997) tracked 27 students over 10 years on the long-term effects of early reading acquisition and subsequent reading development. The students were administered a variety of standardised measures (e.g., reading, cognitive development, vocabulary, and general knowledge) during 1st, 3rd, 5th and 11th grades. Additional measures were administered in 11th grade (print exposure, vocabulary, reading comprehension and declarative knowledge). Hierarchical regression analyses revealed a significant finding; 1st grade reading comprehension was a significant predictor of 11th grade reading achievement. In addition, measures of 1st grade reading skills predicted significant variance in 11th grade print exposure, even when 11th grade reading comprehension had been controlled for. This indicates that a student who acquires core reading skills in first grade is more likely to engage in reading over time.

More recently, Sparks and colleagues (2014) replicated Cunningham and Stanovich's (1997) investigation with a larger sample and additional measures of literacy and language skills. These researchers tracked students over nine years (1st-10th grade) and differentially administered a number of measures (spelling, vocabulary, reading, IQ, and listening comprehension) across these grades. Their findings revealed that 1st grade reading was a strong predictor of 10th grade reading

outcomes. Specifically, early success in decoding skills, predicted growth in decoding throughout the grades, and decoding in 5th grade predicted both reading and language ability in 10th grade. Sparks et al. (2014) also showed that print exposure predicted individual differences in 5th grade vocabulary when 1st grade vocabulary was controlled for, and listening comprehension from 3rd to 5th grade. Moreover, print exposure accounted for significant variance (61.4%) in declarative knowledge when 10th grade IQ was controlled for. This finding suggests that reading quantity may be considered more important than cognitive ability in development of declarative knowledge.

Taken together these findings demonstrate the importance of early acquisition in reading skills. Early success in reading strengthens core reading skills, as well as other literacy skills such as spelling, vocabulary, and listening comprehension. Early intervention and prevention of reading difficulties is a core tenet of Response to Intervention (RTI) Frameworks (Burssock & Blanks, 2010; Hughes & Dexter, 2011) which is described in the following section.

1.7. Response to Intervention

Traditionally, Response to Instruction models were developed and intended for identifying learning disabilities, however, they have evolved into a more general education practice known as “Response to Intervention” or RTI (Kavale & Spalding, 2008). RTI has been proposed as an alternative model to the traditional dual discrepancy models (Hughes & Dexter, 2011); as an all-inclusive early intervention model designed to detect and provide timely support for struggling readers (Gersten et al., 2009). In the U.S., 71% of elementary schools use a form of RTI in reading/language arts (Bramlett, Cates, Savina, & Lauinger, 2010). Its current

popularity stems from problems associated with traditional remedial programmes and special education delivery.

Conventional remedial programmes for literacy difficulties were largely based on dual discrepancy models; the student had to wait until there was a two-year discrepancy between his or her reading level, and their age or grade level. For this reason such models have become known as “wait to fail” models. In addition to reading failure as a pre-requisite for resource allocation, conventional special education delivery was characterised by a number of significant complications. Principal among these difficulties was the static nature of assessment practices and the demonstrated technical adequacy that guide classification decisions (Barnett, Daly, Jones, & Lentz, 2004; MacMann, Barnett, Lombard, Belton-Kocher, & Sharpe, 1989; MacMillan, Gresham, & Bocian, 1998).

The report on the Presidents Commission on Excellence in Special Education (2002) found that although movements such as classification, Individualised Education Programmes (IEPs), and progress monitoring were used widely, their use was not consistent, and decision-making was not based on measurable student data. The report recommended that the traditional classification process be discarded, and a decision-making process based on response to instruction approach be selected. This idea has been consistently referred to as the preferred method in the literature (e.g., Barnett, et al., 2004; Fuchs et al., 2007; Bursock & Blanks, 2010; Vaughn & Linan-Thompson, 2003). In addition, the report recommended that scientifically validated progress monitoring be used to make instructional decisions concerning the provision of special education services, and that dynamic progress monitoring methods be used for continuation of services. In essence, RTI models have been recommended as a solution to address the concerns raised in this and other reports (Barnett et al., 2004).

RTI frameworks aim to prevent reading difficulties and to reduce the gap between struggling and skilled readers by providing for students' educational needs through tiers of increasing instructional intensity (Hughes & Dexter, 2011). There are a number of successful implementations of RTI in school settings, and this initial evidence indicates that these types of frameworks can lessen the gap between struggling and skilled readers by supporting struggling readers *before* they fail (e.g., Denton et al., 2013; Little et al., 2012; Mathes & Denton, 2002; McMaster, Fuchs, Fuchs, & Compton, 2005).

RTI frameworks operate a multi-tiered approach to support students at risk for reading failure by matching students' instructional needs with the appropriate level of instructional intensity. Instructional intensity is organised within tiers, as the students need increases, so too does their access to more intensive instruction. In general, RTI involves three tiers of increasing intensity (Bursock & Blanks, 2010; see Figure 1). Within these tiers a school must incorporate the core components of RTI: (a) an evidence-based core curriculum and targeted interventions; (b) universal screening; (c) progress monitoring and, (d) making decisions about "adequate" progress in successive tiers (Hughes & Dexter, 2011).

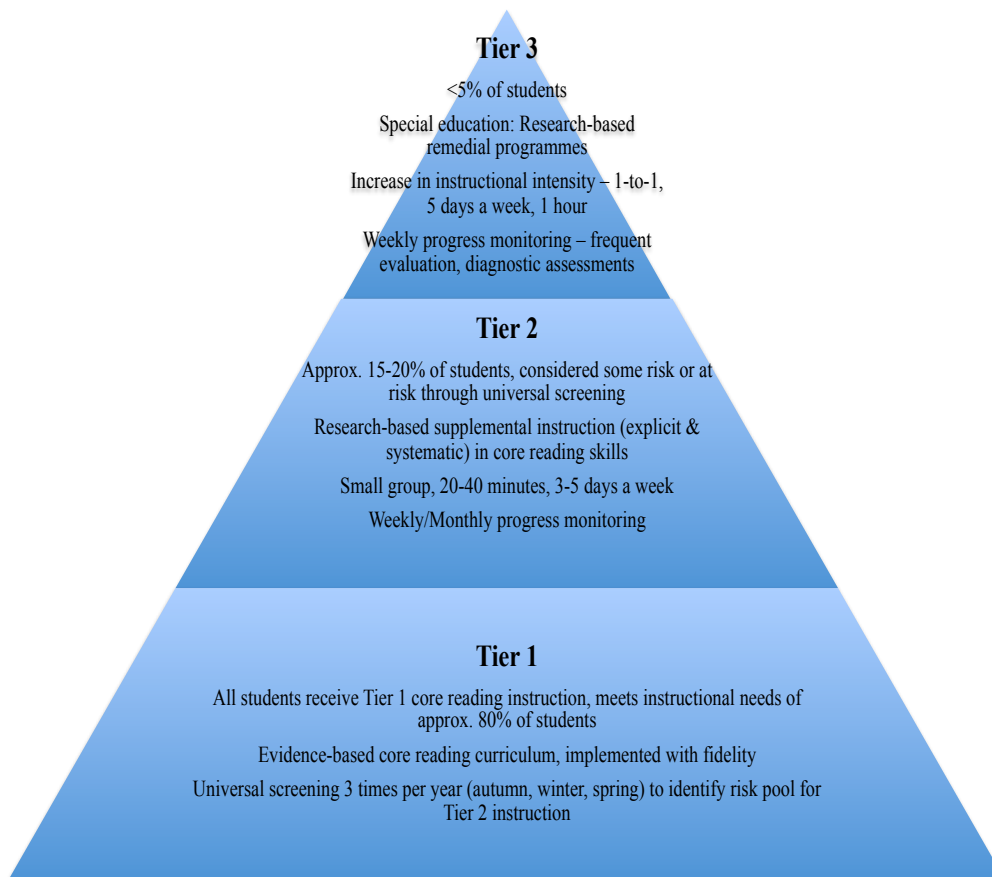


Figure 1. Response to Intervention (RTI) Framework with Three Tiers of Increasing Instructional Intensity (Tier 1, Tier 2, and Tier 3).

Tier 1 (see Figure 1) is the core reading instruction provided in the classroom. It should consist of an evidenced-based curriculum as the foundation of the RTI framework. An effective Tier 1 helps to eliminate inappropriate instruction as a reason for students' inadequate progress (Hughes & Dexter, 2011). To achieve this aim, it is assumed that Tier 1 instruction is implemented with a high degree of fidelity. A recent empirical review of RTI frameworks however, demonstrated that between 2005 and 2011, no study reported fidelity of implementation for general Tier 1 reading instruction (Hill, King, Lemons, & Partanen, 2012). Moreover, the content

of Tier 1 instruction is considered to be somewhat unclear in the literature. Most studies do not state what the Tier 1 instructional programme is, or what exactly is occurring in Tier 1 (Fien et al., in press). Therefore, at this time it remains unclear whether effective instruction is consistently delivered in Tier 1.

The RTI practice guide (Gersten et al., 2009) recommends that in Tier 1 all students be screened for potential reading difficulties at the beginning and middle of the school year. This is known as universal screening to identify a *risk pool* (Fuchs et al., 2007) of students who are not responding sufficiently to Tier 1 instruction and would benefit from additional instruction in core reading skills. Universal screening is the basis for detecting students “at risk” for reading difficulties (Hughes & Dexter, 2011). According to the model, a universal screener must be quick and easy to administer and score, and have adequate scope to identify core reading skill deficits. At present however, there is disagreement within the field on what benchmarks should be used to detect students “at risk” (Gilbert, Compton, Fuchs, & Fuchs, 2012; Hughes & Dexter, 2011; Scholin & Burns, 2012). Lack of consensus on appropriate benchmark criteria is concerned with the phenomenon of false positives and false negatives yielded in universal screening. False positives are students identified as “at risk”, when they are in fact not, and false negatives are students identified as “not at risk”, but who at a later date are identified as “at risk” for reading failure (Jenkins, Hudson, & Johnson, 2007). Over-identification of false positives wastes valuable instructional time; under-identification of false negatives negates timely intervention.

The students identified by the universal screener(s) are provided with Tier 2 intervention (see Figure 1), which is supplemental instruction to Tier 1 that focuses on core reading skills. The proportion of students requiring Tier 2 instruction is estimated at 15-20% of the student population (Yell, 2004). The practice guide for RTI (Gersten

et al., 2009) recommends explicit and systematic instruction for Tier 2. Essential components of explicit instruction are modeling, multiple practice opportunities, and systematic error correction (Carnine, Silbert, Kame'enui, Tarver, & Jungjohann, 2006). The format typically involves small groups (2-4 students), and is delivered for 20-40 minutes 3-5 days per week. It is assumed that the instruction is delivered with high fidelity and is closely aligned to the core curriculum. An aim of Tier 2 is to close the gap between students' current and expected performance (Denton, 2012),

Tier 2 also aims to differentiate between students in need of brief supplementary instruction and those with more specialised instructional needs (Gersten et al., 2009). This aim is achieved in part through individualised progress monitoring. Performance measurements are collected on a weekly/monthly basis and progress toward benchmark standards is evaluated. The properties of the progress monitoring probes are of critical importance when providing tiered support. Highly sensitive probes will not only quickly identify false positives, but also determine subsequent resource allocation. Students who make adequate progress (i.e. respond to the intervention) are returned to Tier 1 (typical classroom instruction). Students who do not make sufficient progress are moved to the third tier, for more intensive intervention.

Tier 3 (see Figure 1) involves the use of intensive interventions for struggling learners with the most significant needs, including those with disabilities (Hoover, 2011), and is provided to an estimated 1-5% of the student population (Yell, 2004). Tier 3 interventions are characterised by an increase in instructional intensity in relation to the previous tiers. This intensification can relate to an increase in instruction time (e.g., from 30 minutes to one hour), a decrease in group size (typically employing one-to-one instruction), extending the duration of the

intervention, and/or providing daily intervention sessions. Instruction is individualised to meet specific student needs, however, how instruction is intensified in Tier 3 varies widely from school to school. Differential outcomes for Tier 3 intervention are also evident, with some students demonstrating significant gains, and others showing little or no progress despite receiving highly intensive interventions (Denton et al., 2013). In Tier 3, progress-monitoring data are collected on a weekly basis and frequently evaluated, and diagnostic assessments are also used in this tier.

For all schools implementing RTI frameworks, the task is to select an evidence-based core curriculum and targeted interventions, a universal screener(s), a progress-monitoring tool, and a framework for making decisions about adequate progress or “response to intervention”. However, RTI has yet to become a recognised and rigorously evaluated, evidenced-based practice (Fuchs & Deschler, 2007). This is due in part to the availability of myriad curriculums, targeted interventions and screeners, and conversely, limited progress monitoring tools and availability of specialised models for making decisions about response to intervention. In sum, there are many techniques and approaches that schools can adopt in their organisation and implementation of a multi-tier support model.

1.7.1. Approaches to Response to Intervention. Typical approaches to RTI to date have included the standard protocol and the problem-solving models. Standard protocols are structured and explicit, and involve the delivery of evidence-based multi-component programmes focused on skill specific areas (Shapiro, 2009). Problem-solving models involve individualised interventions that focus on the student, the learning environment and the curriculum, and use performance data to modify intervention to meet instructional needs (Little et al., 2012). Disadvantages of

both models have been highlighted; the advantages of the problem solving models are viewed as disadvantages of the standard protocol, and vice-versa (Shapiro, 2009).

A significant disadvantage of standardised protocols is that performance data are not systematically used to adjust instruction based on insufficient response and curriculum mastery (Simmons et al., 2013). This results in a trade-off between efficiency and effectiveness for individual students (Shapiro, 2009). All students proceeding at the same pace potentially impede some students' progression i.e., some students may benefit from repeated practice/lessons, while others from a more accelerated pace (Simmons et al., 2013). A main disadvantage of problem solving models is that they can be difficult to implement and require substantial resources and expertise (Shapiro, 2009). Therefore, the likelihood of schools' initial implementation and feasibility of continuation are limited; indicated by scant documentation of such models in the RTI literature.

Extant evidence appears to support the more standardised approaches (Fuchs & Vaughn, 2012), which dominate the literature on RTI (e.g., Coyne et al., 2013; Kamps et al., 2008; McMaster et al., 2005; Simmons et al., 2011). These studies generally report effects favouring treatment groups and a reduction in the percentages of students who remain at risk for reading difficulties (Denton, 2012). The evidentiary favour of standard protocol approaches however, may be due in part to the paucity of research pertaining specifically to problem solving models. In a review of extensive reading interventions, Wanzek and Vaughn (2007) failed to identify any studies implementing a problem solving approach. Some larger scale field studies such as the St. Croix River Education District Model (Bollman, Silbergliitt, & Gibbons, 2007), and the Minneapolis Public Schools (Marston, Muyskens, Lau, & Canter, 2007) have

reported increases in percentages of students passing Curriculum Based Measure (CBM) benchmark scores, and reductions in special education placements.

Recently, elements of both the standard protocol and the problem-solving models have been blended to create hybrid models (e.g., Little et al., 2012; Simmons et al., 2013). Hybrid models use standard protocols that incorporate manualised modifications based on student performance data. Effectively, this involves employing a standardised intervention that systematically uses data to adjust instruction to meet students' instructional needs (Little et al., 2012; Simmons et al., 2013). Combining the standard protocol and problem solving approaches into one model is likely to lead to the greatest responsiveness of students (Shapiro, 2009). Initial evidence from hybrid models however, is mixed. Little and colleagues (2012) found no statistically significant group differences on any outcome measure between those students who received a standard intervention modified in response to student performance, and those receiving the schools typical supplemental reading instruction. In contrast, other research has reported that adjusting curriculum progression based on student performance data yielded significant between group effect sizes (Coyne et al., 2013; Simmons et al., 2013).

Conflicting evidence from hybrid models may reflect the inherent challenges of their implementation. Such challenges include: the effective use of assessment data to intensify interventions (Gersten et al., 2009); which instructional variables to modify and whether certain modifications are more efficacious than others (Little et al., 2012), and the appropriate use of procedures and decision rules relating to curriculum mastery and responsiveness (Simmons et al., 2013). While hybrid models have the potential to address the disadvantages characteristic in both the standardised

and problem solving models, hybrid models may not address the potential challenges encountered in their implementation.

Challenges faced in implementing RTI models however, can be addressed by incorporating techniques based on the science of Applied Behavior Analysis, which offers effective intervention strategies, precise systems of measurement, and standardised problem-solving protocols based on student performance data (Daly, Martens, Barnett, Witt, & Olson, 2007). In fact, RTI emerged from practices within the field of Behavior Analysis, curriculum-based measurement, and functional academic assessments (VanDerHeyden, Witt, & Gilbertson, 2007).

Daly and colleagues (2007) outline a behavior-analytic model of effective instruction for use in RTI. The authors propose that most academic difficulties result from deficiencies in basic academic skills. Because foundational skills are generalisable repertoires that need to be applied comprehensively throughout the curriculum, mastery of these skills is critical to the academic success of the student (Gersten, 2009; Snowling & Hulme, 2010). Daly et al. (2007) describe reading as a generalised repertoire of foundational reading skills, and frame skill proficiency in terms of stimulus control and generalisation. Within the stimulus control paradigm, the student's response comes under the control of the relevant academic stimuli. For example, if a student can produce the phoneme /c/ when presented with the grapheme *c*, the student's response has come under the control of the relevant stimulus. Conversely, if the student produces an incorrect phoneme the stimulus does not control the student's response.

Academic stimuli come to control student responding through differential reinforcement i.e., reinforcing correct responding and removing reinforcement for incorrect responding. Students can acquire skills quickly using this paradigm

effectively. Instructionally, this is achieved through providing the academic stimuli, modeling of the correct response, providing opportunity for the student to respond, and then (differentially) reinforcing a correct response, or error correcting an incorrect response. These complete or discrete learning trials produce effective skill acquisition (Daly, Witt, Martens, & Dool, 1997). In foundational skills such as this there is only one correct response, however, observation and variation in responding can inform the educator on how to proceed in instruction.

This can be understood in terms of the instructional hierarchy (IH) (Daly et al., 1997; Haring, Lovitt, Eaton, & Hanson, 1978), which describes how academic responding changes as a function of instruction. Haring and colleagues first described the IH in 1978; they recognised that academic responding (e.g., reading, writing, mathematics) could be more effectively strengthened if educators were systematic in how they structured antecedent prompts and managed consequences. The IH is based on the principles of behavior that have extensive explanatory power (Ardoin & Daly, 2007), and according to the model different principles of learning are applicable at each level of responding in the learning hierarchy (Daly et al., 1996).

The IH proposes that skill development progresses through four levels (acquisition, fluency, generalisation, and adaptation), which are accompanied by specific instructional procedures that will most efficiently produce mastery at that level (see Figure 2). Extrapolating the stimulus control paradigm, Haring and colleagues (1978) described specific procedural strategies to implement with students, as academic responding progresses. As a student gains a skill, he/she first acquires the skill, and then he/she becomes fluent in skill use. Next the student learns to generalise its use in novel contexts, and finally adapts the skill response as necessary according to novel demands (Daly, Lentz, & Boyer, 1996).

The IH describes how each level of academic responding can be viewed as a potential intervention target itself, and refers to the specific instructional procedures that are appropriate for the level of academic responding (i.e., acquisition, fluency, generalisation, adaptation). This means that each level of academic responding has different corresponding procedures that most efficiently produce mastery at that level (e.g., modeling at acquisition, practice at fluency, etc.) (Daly et al., 1996). According to the IH, instruction should be provided that promotes accurate responding through modeling and error correction, and the ratio of feedback to responses should be 1:1. If errors occur these should be systematically and immediately corrected. Then stimulus control is strengthened beyond a level of accuracy to fluency through practice and performance feedback. The IH emphasises that generalisation should not be expected as a result of accuracy and fluency based instruction, but rather must be programmed for by providing sufficient practice opportunities to promote generalisation (see Figure 2).

The Instructional Hierarchy

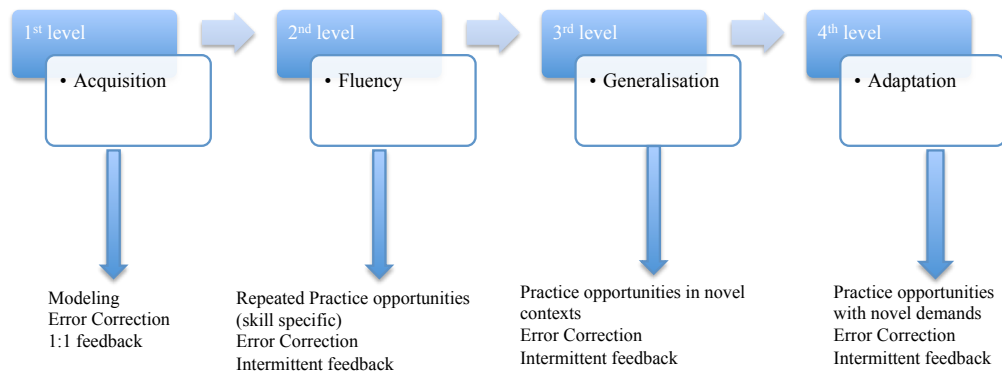


Figure 2. *The Instructional Hierarchy.*

The IH has informed educators to attend to student responding and the way responding changes as it is strengthened; and, how to react to those changes in responding to strengthen response repertoires that are more broadly generalisable (Ardoin & Daly, 2007). Therefore, observing academic responding indicates where the student is on the IH. For example, responding may be slow and effortful, indicating that the relationship between the stimulus and the response is tenuous. Stimulus control can be significantly strengthened through repeated practice and reinforcement for rate of responding using fluency performance criteria (Daly et al., 2007).

Fluency is best targeted with an instructional methodology designed specifically to achieve this outcome, which incorporates a measurement system that accurately measures this construct, thus providing a mechanism for precise

observation of academic responding. Precision Teaching is a behaviour analytic instructional method and system of measurement that fits all of these criteria, and has been described as a “match made in heaven” with RTI frameworks (Johnson & Street, 2013).

1.8. Precision Teaching

Precision Teaching (PT) is a fluency-based instructional method and assessment tool that uses repeated practice, error correction, charting of performance, and reinforcement to target fluency in academic and other skills (Johnson & Street, 2013; Kubina & Yurich, 2012). In PT, fluency is conceptualised as accurate, effortless, flexible, and masterful performance that is maintained over time and can be applied in novel contexts and combined in novel ways (Johnson & Street, 2013; Kubina & Yurich, 2012). The definition of fluency is an important distinction in the PT literature; it refers to a behavioural outcome that is the result of specific instructional procedures.

Ogden Lindsley (1964; 1972) was the founder of PT. Lindsley emphasised evaluation and revision in systematic instruction; specifically, by pinpointing behaviours and counting them within timing units and charting that performance daily. Lindsley (1972) stated that the “child always knows best”, and he encouraged educators to “try, try again” when original procedures did not yield the desired results. This means that if the initial instruction is not producing anticipated outcomes, it is not the fault of the learner; rather it is the responsibility of the educator to identify (and if necessary modify) the environmental variables that will produce the desired results. Since its inception, the principles and practices of PT have resulted in many discoveries in learning and performance (Binder & Watkins, 1990).

Research has revealed that fluent behaviour is retained/maintained over significant periods of time void of practice (e.g., Bullara, Kimball, & Cooper, 1993; Ivarie, 1986; Shirley & Pennypacker, 1994). Fluent behaviour also demonstrates endurance i.e., it can be performed with speed and accuracy for extended periods of time (e.g., Miller, Hall, & Heward, 1995; Whalen, Willis, & Sweeney, 1993). Fluent behaviours also exhibit application (e.g., Berens, Boyce, Berens, Doney & Kenzer, 2003; Lin & Kubina, 2005). Application is where one or more element behaviours reach a specific frequency and then are applied to a compound behaviour (Kubina & Yurich, 2012). For example, increasing frequencies in letter sounds and blending sounds (element behaviours) may apply to decoding words (compound behaviour). The specific frequency levels for any given skill that predict these outcomes are known as “performance standards”. Much research has documented the relationship between achievement of performance standards and the behavioural outcomes of fluency (e.g., Berens et al., 2003; Lin & Kubina, 2005). These fluency outcomes (Retention-R, Endurance-E, Application-A, Performance Standards-PS) can be abbreviated to REAPS (Lindsley, 1992).

The behavioural outcomes of fluency are the defining characteristics of fluent behaviour; therefore, assessment of REAPS is the true test for mastery and fluency (Lindsley, 1992). That is, if the behaviour reaches the required frequency aim and can endure over extended periods of time, after a significant period of no practice, and can be applied to more complex behaviours - then the behaviour can be described as fluent.

The fluency outcomes described are achieved through specific instructional practices that incorporate a precise system of measurement. PT is the only educational method that blends assessment and instruction into the same activity (Kubina &

Yurich, 2012). As both elements contain multiple components, PT as an instructional method and as an assessment tool are described separately below.

1.8.1. PT as an instructional method for building fluency. In a review of 16 reading interventions, Bramlett and colleagues (2010) identified four commonalities for intervention success: practice/drills, modeling with error correction, intensive intervention, and reward or reinforcement. PT combines these strategies into an instructional method. Specifically, PT increases academic rate of responding towards a terminal goal by incorporating feedback, reinforcement, and error correction. PT increases students' academic rate of responding through directly targeting and increasing frequencies to performance standards. The principal agent for cultivating fluency is frequency building, which involves timed repetition of performance and performance feedback (Johnson & Street, 2013; Kubina & Yurich, 2012).

1.8.1.1. Increasing rate of academic responding through frequency building. By its nature, frequency building directly increases learning rates. Skinner (2008) argued that learning disabilities should not be considered a failure to learn but rather reflect failure to learn at the expected pace; therefore learning disabilities are more accurately described as "learning rate difficulties" (pp. 310). By increasing learning rates, students with disabilities/reading difficulties are more likely to approximate performance that is comparable to average performing peers (Konrad, Helf, & Joseph, 2011). In PT frequency building trials are conducted in timed intervals. The student responds to the academic stimuli in a free operant paradigm i.e., the instructor does not provide feedback during the timed trial. This arrangement does not place a ceiling on the student's performance, and increases opportunities for productive practice. Increased opportunities to respond are associated with increased academic outcomes (Mellard, McNight, & Jordan, 2010; Bramlett et al., 2010).

Goal Setting. Research shows that outcomes for students with more intensive instructional needs are improved with high goal specificity (Vaughn, Denton, & Fletcher, 2010). Clear learning goals allow both the educator and the learner to evaluate the learner's performance as they move towards mastery. Learning goals also define mastery (the end point of instruction for that specific skill). This is an important characteristic of the process as it allows the student to self-monitor progress, and promotes student investment and ownership with the learning activity (Konrad et al., 2014).

In PT practice opportunities are goal oriented. Performance is guided by daily goals and weekly goals, and is ultimately aimed towards a mastery criterion (performance standard) at which point training in the skill is terminated. Goals are not set arbitrarily nor to a standard protocol, rather they are determined by the student's previous performance. To achieve a high rate of responding the weekly goal is typically set as a doubling of response frequencies over a week. Daily goals are determined by how that doubling of performance is spread across the week. For example, if a student demonstrates five correct responses and the weekly goal is to double that to 10, then the daily goals are incremental (i.e., six on Monday, seven on Tuesday, and eight on Wednesday etc.). Relating current performance to previous or expected performance in regard to the overall learning goal provides high goal specificity (Chan et al., 2014; Hattie & Timperley, 2007).

Feedback. For feedback to be effective it must be related to clear learning goals and be based on accurate student performance data yielded from goal-focused learning activities (Hattie & Timperley, 2007). To enhance the effectiveness and efficiency of instruction, feedback must concentrate on success, and be immediate, specific, and task oriented rather than person-focused (Chan et al., 2014; Hattie, 2009;

Hattie & Timperley, 2007). This is the case in a typical PT session. Each timed trial is followed up with immediate and quantified feedback on correct responding, and its relation to goal attainment. Immediate feedback allows the student to use the feedback to make adjustments to his or her performance, and is especially important for those students struggling to achieve academic demands (Chan et al., 2014).

Reinforcement. Feedback provided on a successful performance is most likely reinforcing, but this cannot be assumed (Cooper, Heron, & Heward, 2007). Reinforcement is any stimulus that increases the likelihood that the behaviour will increase in future; reinforcement ensures that academic responding comes under the control of the correct stimuli. In PT correct responding and goal attainment are reinforced using task and goal specific praise, and/or tangible items.

Systematic Error Correction. By virtue of their current status, students with academic difficulties have a history of academic underachievement, and may be resistant to corrective feedback. A feedback-focused environment that treats errors as learning opportunities and provides for direct and immediate error correction can create a learning atmosphere that is risk free and student-focused (Chan et al., 2014). In PT errors are called “learning opportunities”. All errors are directly and systematically corrected. The educator models the correct response and the student is given the opportunity to emit the correct response (a learning opportunity). Immediate and direct error correction is crucial (Carnine et al., 2006). While there are many types of error correction, efforts to prompt the student, or make the student guess, can lead to confusion and may inadvertently consolidate the error (Chan et al., 2014; Konrad et al., 2011). Subsequent opportunity to emit the correct response is integral to error correction, it ensures that the feedback was effective and provides the learner with further feedback on a correct response (Chan et al., 2014).

Figure 3 provides a description of a typical PT session targeting letter sounds, illustrating how the various strategies combine as an instructional method.

Billy is a kindergarten student who is struggling with letter sounds. Using timed trials the teacher measures the frequency of his correct and incorrect responding over a number of days to obtain a baseline of his current level of performance. Billy's baseline performance is used to set his daily and weekly performance goals for letter sounds. Billy's PT session is highly structured, he is provided with a goal for each day, and a number of timed trials to practice letter sounds to achieve that goal. Billy is given a sheet of randomised letters, and when the timer begins he says each letter sound as fast as he can. The educator counts Billy's correct and incorrect responding during the timing trial. At the end of the timing, the educator provides immediate feedback (number of corrects) and reinforcement (performance specific praise) for correct responding. This performance is charted on a specialised chart on which Billy can graphically see his performance. The educator then orients Billy to his learning opportunities (i.e., any incorrect responding), and provides systematic error correction for each incorrect response. Therefore, Billy is given the opportunity to emit the correct response in each instance. This process is repeated a number of times (usually 3-5 timed trials with all elements included).

Figure 3. Description of a Typical PT Session Targeting Letter Sounds.

PT is not only a powerful targeted intervention that focuses on fluency; it is also a sensitive progress-monitoring tool that is particularly suited for use within multi-tier support models (White, 2009; Johnson & Street, 2013; Kubina & Yurich, 2012). This is because PT affords a precise system of measurement and progress

monitoring that facilitates the consistent use of decision rules, and exhibits formative and dynamic assessment characteristics.

1.8.2. Precision Teaching as a Precise System of Measurement. PT is not only an effective targeted intervention; its focus on frequency provides invaluable progress data. Frequency is a standard unit of measurement that is highly sensitive to even the smallest changes in performance (Kubina & Lin, 2007). Each timed trial yields a frequency count, therefore each trial measures if performance is increasing, decreasing, or remaining the same. While frequency is the best measure of performance, progress in rate of performance over time is the best measure of learning (Johnson & Street, 2013). By measuring and graphing changes in rates of performance over time, learning can be quantitatively described and evaluated on a specialised chart called the Standard Celeration Chart (SCC) (Johnson & Street, 2013; Kubina & Yurich, 2012).

Figure 4 illustrates the SCC, which is a standardised non-linear chart. The horizontal axis is equal interval, each vertical line represents a day of the week, beginning with Sunday and ending with Saturday (see Figure 4). The chart accommodates 140 consecutive days; this means that performance data spanning a school year can be represented on the chart. The vertical axis is a ratio scale, each horizontal line represents a performance frequency, and all performance data are charted as counts per minute (see Figure 4). The vertical ratio scale is a multiply/divide scale that repeats in six cycles of ten units; this means that changes in frequencies on the chart are multiplicative and not additive. The distance between 1 and 10 is the same as the distance between 10 and 100 (see Figure 4); this reflects the

fact that it is as easy to go from 1 to 10 responses per minute, as it is to go from 10 to 100 responses per minute (Johnson & Street, 2013).

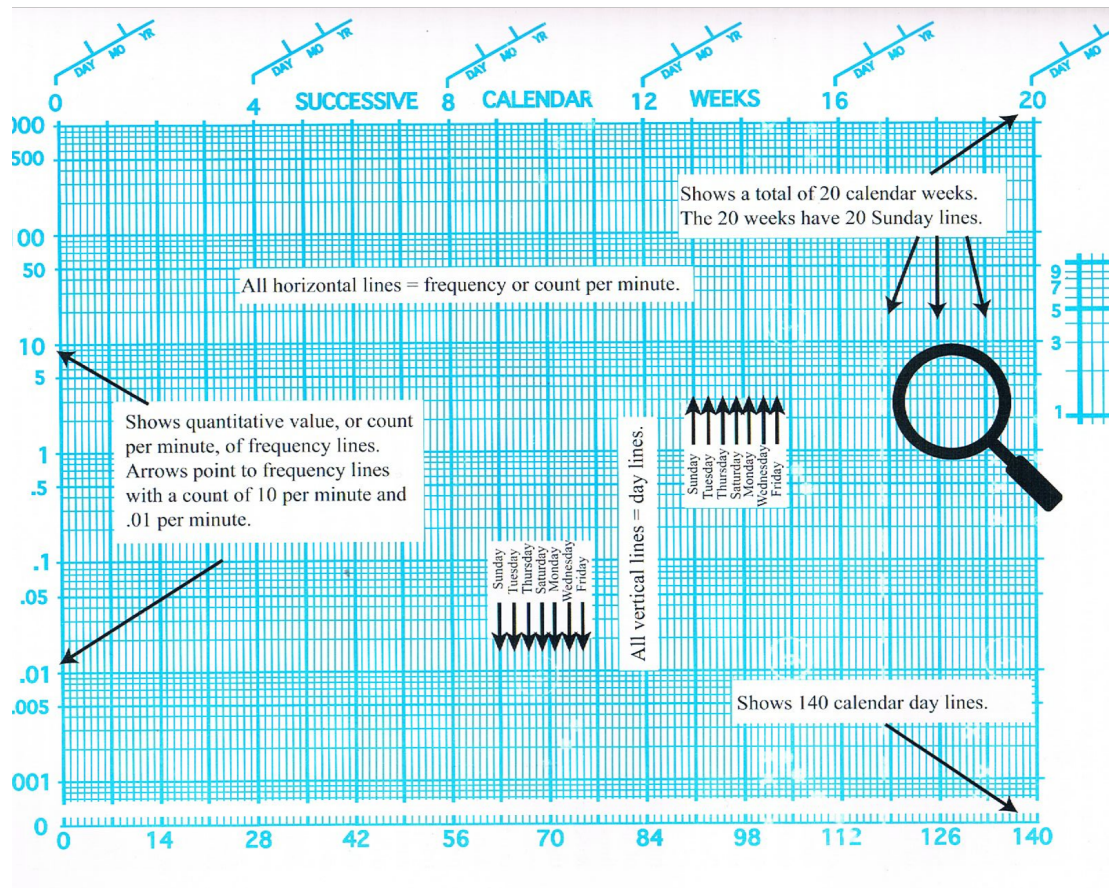


Figure 4. Standard celeration chart (as featured in Kubina & Yurich, 2012).

The multiply/divide count per minute scale of the SCC permits the calculation of celerations i.e., changes in frequency over time. Lindsley used the term celeration to signify either increase in rate (acceleration) or decrease in rate (deceleration) over time. A multiply or divide symbol (\times and \div , respectively) is used to represent changes in behaviour; for instance, a $\times 2$ celeration means learning has doubled in frequency, and a $\div 2$ celeration means that it has halved in frequency. For example, a learner who

read 40 words per minute on the first Tuesday and reads 80 words on the next Tuesday has doubled his performance, or has a celeration of x2 (over seven days).

Because celeration refers to, and quantifies changes in frequencies across time, it provides a standard measure of learning (Calkin, 2005). This expression of learning has been described as providing: “The first simple quantification of learning in the history of the behavioural sciences” (Binder, 1990, p. 2). Celerations permit educators to quantitatively describe, compare, and predict student performance (Kubina & Lin, 2007). These are desirable practices in an RTI framework. A precise and quantitative description of academic learning permits multiple performance comparisons. For example, a student’s baseline celeration (pre-intervention level of performance) can be compared to their celeration during the intervention. Also, the SCC can be used to illustrate magnitude of change occasioned by the introduction of an intervention, quantifying immediacy of intervention effect. Therefore, the SCC can quantitatively index two aspects of a student’s “response” to intervention (immediacy of intervention effect and rate of learning).

Similarly, because celeration is quantifiable, two or more students’ “response” to the same intervention can be compared. This permits the educator to evaluate if a student is progressing at a rate commensurate with peers. In addition, the timeframe in which the performance standard will be reached (mastery criterion) can be calculated using a celeration; allowing the educator to predict the student’s performance and estimate when instruction will be completed. This is an important feature for skill sequencing, longer-term intervention planning, and curriculum integration (Johnson & Street, 2013).

The SCC also permits the educator to systematically and quantitatively set daily and weekly goals as the student progresses towards the terminal learning goal.

The terminal learning goal is generally the performance standard associated with the skill, and this provides a visual representation of a mastery criterion that both the educator and student can refer to while the student progresses through the programme. Daily and weekly goals are based on the student's previous performance, and are also visually represented on the chart. Therefore the SCC allows the student to self-monitor their progress towards the terminal goal, and provides immediate visual performance feedback that can be used to make adjustments to his or her performance.

1.8.2.1. Precision teaching as an assessment and progress-monitoring tool.

Due to its sensitivity, standardisation, and statistical properties SCCs are considered potent progress monitoring tools (Calkin, 2005; Datchuk & Kubina, 2011; Johnson & Street, 2013, Kubina & Yurich, 2012). As data points on the SCC increase in value over time, they signify increasing rates of change, and conversely data points that decrease in value over time signify decreasing rates of change. Because rate of change, and not absolute frequency, is used as the critical property of progress, the chart makes it easy for students and teachers to make quick, daily and timely decisions about whether a student is progressing (Johnson & Layng, 1992).

Progress-monitoring is a crucial element of Tier 2 and Tier 3 intervention in guiding subsequent instruction, and in identifying those students who have made sufficient progress to return to regular classroom instruction (Fuchs & Fuchs, 2006). In the former instance this will permit programme modification when required to suit the instructional needs of the student, and in the second instance it will not waste valuable resources where they are no longer needed. Therefore, the selection of the progress-monitoring procedure is critical.

In many instances the progress-monitoring procedure chosen is Curriculum Based Measurement (CBM). CBM is a procedure that educators use to measure how students are progressing in basic academic areas such as math, reading, writing and spelling. CBMs are administered on a weekly basis, and these data are charted to facilitate evaluation of the student's progress toward meeting academic goals. Binder (1990) outlines a number of obvious conceptual and practical similarities to CBM and PT. There are three main similarities between PT and CBM. First, both procedures utilise timed measures of a student's performance to evaluate performance. Second, both CBM and PT use graphic displays of performance over successive calendar days for recording performance and as progress monitoring tools to facilitate data-based decision making and individualised instructional programming. Third, they both use the term "fluency" to define the objective of mastery learning at all steps in the curriculum sequence (Binder, 1990).

The differences between these approaches however, highlight the advantages to using PT. Binder (1990) elaborates on the important distinctions that give PT advantages over CBM: the choice of graphic display; how performance criteria are established, and the definition of fluency. The progress-monitoring tool that CBM uses are equal interval or "add/subtract" charts that are not consistently standardised on a count per minute scale. Two main pitfalls of these charts is that add-subtract scales do not accurately convey magnitude of behaviour change (Binder, 1990), and scale manipulations can create "stretch to fill" data patterns (Kubina & Yurich, 2012, p. 154).

Figure 5 (adapted from Kubina & Yurich, 2012) illustrates how non-standardised add-subtract charts can warp data through vertical axis scale manipulations. All three of the add/subtract charts are plotted with the same data, yet

all three convey ostensibly different magnitudes of effect. The uppermost add/subtract chart appears to demonstrate a strong effect, whereas the lowermost one conveys almost no effect. In addition, compare the performance differences between the first two data points and the last two data points: both represent a performance increment of two, and the magnitude of change appears to be the same in both instances (see Figure 5). The magnitude of change, however, is not the same.

A change in response frequency from one to three correct responses per minute represents a 200% increase in performance (first two data points); a change from four to six represents a 50% increase (last two data points), and a change from 10 to 11 represents just a 10% increase. Clearly these are very different magnitudes of performance change, yet all appear the same on an add/subtract chart (see Figure 5). As a result large performance changes are not discernable and small changes are magnified, making it difficult to accurately visually inspect the data.

In contrast, the SCC's ratio scale preserves this magnitude of change, and accordingly performance changes are accurately represented - and visible to the naked eye. Figure 5 also shows the same data plotted on a standard celeration chart. The magnitude of change between the first two data points (200% increase) and the last two data points (50% increase) is accurately conveyed and can be easily interpreted through visual analysis. In addition, this magnitude of performance change is quantifiable through celeration, representing a $\times 11.9$ celeration from the first to the last data point. Because the chart is standardised, the ratio scale of the vertical axis is never manipulated, and therefore accurately and consistently represents performance changes that can be quantified with celeration values.

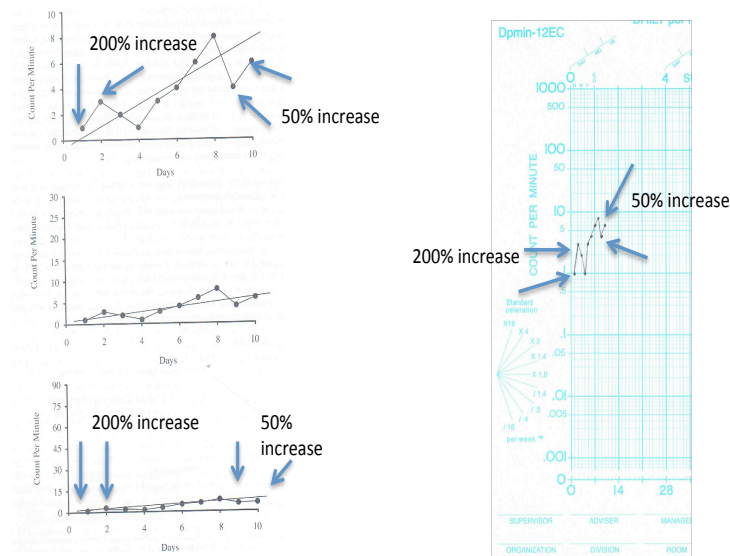


Figure 5. Three add-subtract charts displaying the same data, but different vertical axis scales: the same is data displayed on a standard celeration chart.

These features illustrate the superiority of the SCC as a graphic display (Binder, 1990; Datchuk & Kubina, 2011; Kubina & Yurich, 2012). Due to its standardisation and the specific features described (e.g., celeration calculation) the SCC has significant analytic power in contrast to nonstandard add/subtract charts. Because of these qualities, the SCC is the only chart appropriate for use in fluency building instructional programmes (Fabrizio, 2003). PTs use of the SCC as a graphic display therefore offers distinct advantages for progress monitoring.

The second difference outlined by Binder (1990) is how CBM and PT establish performance criteria. CBM uses local norms as performance criteria; this causes difficulties in measurement and potentially in special education provision. In terms of measurement, if the whole class performs below the mastery level, then the class norm is not a fair representation of the mastery criterion. Consequently, a student exhibiting skill deficits may not be considered in need of special education services, as the overall performance in the class is low (Slate & Jones, 2000).

Conversely, PT uses absolute standards as performance criteria for fluency training. High performance standards are set for pre-requisite or “tool” skills to ensure the student progresses smoothly through the curriculum, and to achieve the benefits of fluency (e.g., maintenance). Maintenance refers to the magnitude of continued performance across time after an intervention has ended (Cooper et al., 2007). Maintenance of skills is achieved through behavioural fluency; Kubina and Yurich (2012) cite 21 published studies demonstrating that frequency building to performance standards result in retention of the skill over long periods of time (i.e., maintenance).

This difference in establishment of performance criteria stems from opposing conceptualisations of fluency by CBM and PT. Tindal (1989, as cited in Binder, 1990, p. 2) states that in CBM: “There is no objective standard of fluency. We have to know the normative information”. Contrastingly, in PT fluency is described as a behavioural outcome that is primarily achieved through frequency building to performance standards. The end goal of frequency building is to achieve performance criteria that are associated with fluent behavior (e.g., effortless performance; Johnson & Street, 2013; Kubina & Yurich, 2012). Accordingly, in PT there is an objective standard of fluency that is determined through evaluation of the level of speed and accuracy

necessary to ensure retention/maintenance, endurance, and application of skills and knowledge (Binder, 1988).

The differences between PT and CBM for progress monitoring illustrate that PT has advantages over CBM. Both PT and CBM are progress-monitoring procedures; PT uses the SCC as a tool, and CBM uses add/subtract charts as a tool. The progress monitoring tool used in RTI frameworks needs to be sensitive to performance changes, to quantify students' rates of improvement and to evaluate instructional effectiveness (National Center on Response to Intervention; NCRTI, 2013). The SCC is a progress-monitoring tool with statistical properties that fulfil these criteria, and its superior graphical display accurately communicates magnitude of change to the naked eye. Ratio scales are commonly used in most areas of science (Johnson & Street, 2013), and a celeration line on the SCC directly communicates direction and speed of learning (Kubina & Yurich, 2012). As a procedure, PTs use of performance standards means that there is a mastery criterion against which a student's performance can be gauged, and the effectiveness of instruction quantitatively evaluated.

Contrastingly, the add/subtract charts used as progress-monitoring tools in CBM do not accurately convey magnitude of behavior change, and can warp data patterns that stretch to fill the scale adopted in the non-standardised graph, obscuring data patterns. Specifically, add/subtract charts tend to underestimate the power of an intervention when performance is occurring at low frequencies, and overestimate the power of an intervention when performance is occurring at high frequencies (Johnson & Street, 2013). This artifact of data display can lead educators to continue with an intervention that is not providing optimal outcomes, and conversely to terminate interventions that are in fact effective. As a procedure, CBM uses local norms to

define fluency; this means that low achieving classrooms will produce low academic standards against which instructional effectiveness is evaluated.

The insufficiencies outlined indicate both measurement issues and some conceptual difficulties with the use of add/subtract charts used in CBM as progress monitoring tools. Recent research has corroborated that the use of the word reading CBM (R-CBM) as a progress monitor is lacking (Ardoin, Christ, Morena, Cormier, and Klingbell; 2013). The R-CBM is the most commonly used and researched form of CBM within progress monitoring contexts. Ardoin and colleagues (2013) conducted a systematic review that included 102 articles pertaining to the R-CBM as a progress monitor. These authors stated that although there is robust evidence for the use of R-CBM as a summative assessment (screening and benchmarking), this does not confer support for the use of R-CBM as a progress monitor. They concluded that the research on R-CBM lacks sufficient evidence for accuracy and sensitivity to evaluate growth for the individual student, and should not be used for making special education provision decisions.

Therefore, the paucity of research on the reliability and validity of the decision rules that accompany these assessments is considered a problem with the use of CBMs (Reed, Cummings, Schaper, & Biancarosa, 2014).

1.8.2.2. Precision teaching and decision rules. Using performance data to determine the instructional variables that need to be manipulated to improve student performance is central to RTI (e.g., Gersten et al., 2009; Hoover, 2011; Hughes & Dexter, 2011; Simmons et al., 2013). To reach defensible conclusions, RTI necessitates the implementation of an integrated and sequenced set of procedures with the correct application of decision rules (Barnett et al., 2004). As CBM is frequently selected as the progress monitoring procedure in RTI frameworks, the decision rules

that accompany these measures are also frequently employed. Despite the widespread use of CBMs as a progress monitoring procedure, there is scant research on the reliability and validity of the components of progress monitoring to make individual student decisions (e.g., goals and decision rules; Reed et al., 2014).

In their systematic review of R-CBM as a progress monitor, Ardoin and colleagues (2013) described two rules commonly used to evaluate progress - the data point decision rule, and the trend line decision rule. Both necessitate the comparison between a straight line connecting a student's initial performance level to the preferred performance level by intervention end (referred to as the goal line). For the data point decision rule, decisions are based on where data points fall in relation to the goal line. The NCRTI (2013b) provides a general guideline whereby four consecutive data points below the goal line indicates an ineffective intervention and that intensifying the intervention will better address the student's instructional needs. Conversely, four data points above the goal line indicates greater gains than expected and an indication that the goal should be increased. Finally, an intervention is maintained when four data points fall above and below the goal line, as the data suggest sufficient progress.

The trend line decision rule necessitates the calculation of a trend line, which can be computed by a number of different procedures. For the trend line decision rule the slope of the line represents a student's estimated growth, which is compared to the goal line to make instructional decisions in a similar fashion to the data point decision rule. Specifically, where the slope of the trend line is greater than that of the goal line, performance is greater than expected (increase the goal/decrease intervention intensity); where it is less the intervention is ineffective (intensify the intervention);

and where it is similar the intervention is working and should be continued (NCRTI; 2012).

While such rules are clearly defined, Ardoin and colleagues (2103) highlighted a number of concomitant problems with their accurate application. In terms of the data point decision rule, the number of data points recommended for decision making varies widely, and more data are required to guide educational decisions than is commonly recommended. The various procedures for calculating trend lines yield differing estimates; even the best procedure has poor predictive accuracy, and at least 20 data points are required to make decisions. Importantly, Ardoin et al. (2013) could not identify any evidence that provided empirical data of the accuracy of R-CBM decision rules for progress monitoring. In light of the combined findings, the authors concluded that: “typical CBM-R progress monitoring practices seem questionable at best” (p. 233).

In addition to the debateable use of R-CBM as a progress monitoring and decision making tool, the timeframe necessary to collect performance data using any CBM measure places a ceiling on its sensitivity to capture intervention ineffectiveness in a timely fashion. In Tier 2, CBM data are generally collected at weekly intervals. This means that a 2-month period must pass before a student’s response can be evaluated; clearly this is neither an optimal nor efficient use of instructional time. In addition, the student may require a slight modification to the intervention that will provide a better fit to his or her instructional need, the introduction of which may occasion response to the intervention (Daly et al., 2007). In contrast, PT yields data points on a daily basis; this micro-level performance data can be used to make instructional decisions within a very short time frame (Malmquist, 2004).

The SCCs standardisation and statistical properties permit exact quantification of learning, and thus decision-making can be applied systematically. For example, in PT the rate of growth typically aimed for, is a doubling of performance over one week of instruction, or a x2 celeration (Lindsley, 1992). Steady progress at a x2 celeration translates to a 100% rate of progress. A x2 celeration can therefore be set as a performance criterion for the intervention which can then be used as a parameter for decision making.

For example, if a student maintains or exceeds a x2 celeration, then no changes are made to the intervention, because the student is considered to be demonstrating progress. If a student's performance falls below a x2 celeration over at least two consecutive days it indicates that a modification may be necessary, in the form of an antecedent or consequence change. Modifications can be systematically applied that do not necessarily change the intervention; rather they slightly alter the intervention to better meet a student's instructional needs. For example, by reducing the task complexity of word training from 10 to five words targeted. Conversely, if the student exceeds the x2 celeration, this implies that he or she will reach the performance standard at a quicker pace, and they can move onto the next target skill – no time is wasted overlearning a skill that is mastered.

Because a x2 celeration is based on a student's previous performance, it is both idiosyncratic to the individual, yet can consistently be applied across a group of students as a performance criterion for an intervention. This is because celeration is quantifiable and easily measured. Consequently, decision rules can be applied consistently and in a timely fashion. A student's celeration, or rate of progress, can be calculated with exactness using just five data points. In effect, within a week the educator can quantify a student's progress towards a mastery criterion. Furthermore, a

x2 celeration line is applied to the SCC as a goal line on a weekly basis; if performance falls below the celeration line over two consecutive days then the educator can make an antecedent change to the intervention to better suit the students instructional needs. Daily performance data represent an opportunity for the timeliest programme modifications; no other intervention yields such data.

A central tenet of RTI is adjustment of instruction in response to content mastery (Gersten et al., 2009), and yet limited research has investigated the impact of adjusting curriculum progression based on student mastery data (e.g., Denton et al., 2010; Simmons et al., 2013; Vaughn et al., 2010). The RTI practice guide (Gersten et al., 2009) recommends Curriculum Embedded Measures (CEM) to monitor student progress. CEMs assess taught content and measure mastery of specific skills. Therefore, CEMS provide information for subsequent instruction, and adjustments based on student performance data (Simmons et al., 2013). The performance standards for academic skills in PT instructional programmes can be considered a CEM as they specify the mastery criteria for that skill. In contrast, CBM assessments monitor student growth on curriculum-independent general outcomes; however, they complement each other in progress monitoring protocols (Simmons et al., 2013).

In summation, student performance data should be frequently and systematically collected and documented as this facilitates timely decision making and instructional adjustment based on student need (Joseph et al., 2014). PT yields performance data on a daily basis and using celeration lines, decisions can be made with just two data points, thereby adjusting instruction to better suit a student's instructional needs. This represents efficient use of instructional time. Mastered content is moved on from, and time re-directed to presenting instructional needs. This is the ultimate goal of collecting student performance data, to identify areas in need of

focus and maximally match future instruction to student need (Chan et al., 2014). Using assessment data in such a way is considered “formative” (William, 2010) and this will be outlined in the next section

1.8.2.3. PT as a formative assessment. Traditional assessment methods have failed to be directly linked to effective on-going intervention planning, and consequently, to positive outcomes for students (Gresham & Witt, 1997). Assessment practices that use results to guide subsequent instruction in terms of content and modifications have become known as “formative assessment” (Black and William, 1998). Indeed, a formative assessment is not an inherent quality of the assessment; rather, it is the use of assessment outcomes to inform subsequent instruction on a regular basis, and to make instructional changes (William & Leahy, 2006; Hattie, 2009). For that reason it has been described as assessment *for* learning, rather than assessment *of* learning.

Formative instructional practices are a product of blending formative assessment outcomes into the planning and delivery of instruction (Konrad et al., 2014). The formative use of assessment findings is thus an iterative approach; it becomes part of a cycle (assessment → intervention → evaluation). PT provides the framework for such a cycle. PT can be considered formative through its repeated use of assessment (data from timed trials) to inform data-based decision making and subsequent instruction (Roberts & Norwich, 2010). Formative instructional practices allow for feedback to be delivered during instruction (Chan et al., 2014). This may be considered similar to “dynamic assessment”, which blends assessment and instruction into the same activity (Grigorinko, 2009; Lidz, 1991).

1.8.2.4. Precision teaching as a dynamic assessment. Because PT combines corrective feedback and intervention with assessment it can be considered a dynamic

assessment. Dynamic assessment is not static; it follows a test-intervene-retest format in which the retest change is interpreted as “response to instruction” (Lidz, 1991). Grigorenko (2009) compared and contrasted the critical components of RTI and dynamic assessment and concluded that the two approaches hail from one “family” in education and psychology whose core feature is blending assessment and intervention into one activity. Blending these components is the hallmark of PT: a cycle of assessment, feedback, and intervention. Each frequency trial in PT is not only a means of fluency training for the student, but also yields assessment data in terms of correct and incorrect responding. By observing performance in subsequent timed trials following corrective feedback, the educator can determine how the student is “responding” to instruction. In this way PT functions as a dynamic assessment. Dynamic assessment has been highlighted as a potentially innovative line of investigation in RTI frameworks as it may be used to index students’ “responsiveness” before allocating interventions, which may be considered long and costly (Connor, Alberto, Compton, & O’Connor, 2014).

In summary, PT fulfills many functions within an RTI framework. It can be used to target fluency in foundational reading skills, and demonstrates characteristics shown to optimise instructional effectiveness (e.g., high goal specificity). It functions as a precise system of measurement to observe academic responding and can be used to index a student’s response to intervention. PT quantitatively measures both level and rate of learning, and provides a systematic mechanism for data-based decision making. PT also demonstrates formative and dynamic assessment properties. These are important characteristics for multi-tier support model implementation and evaluation. Two other considerations highlight the suitability of PT to RTI frameworks, that of instructional intensity and instructional efficiency.

1.9. Instructional Intensity, Instructional Efficiency and Precision Teaching.

School system and teacher accountability has driven intervention effectiveness to the forefront of research and practice. While such a focus is laudable, it has obfuscated practical considerations that ultimately drive educational practices including feasibility and efficiency (Bramlett et al., 2010). In addition, instructional intensity has emerged as an area that requires attention and investigation in RTI frameworks (Coddling & Lane, 2015). An applied behaviour analytic model of RTI that uses PT may hold the potential to address the challenges in dealing with these constructs.

1.9.1. Instructional Efficiency. Instructional effectiveness and instructional efficiency may be considered two sides of the same coin. Both elements must be evaluated in regard to interventions and instructional practices. Konrad and colleagues (2011) describe instructional efficiency as a way of teaching that maximises desired outcomes using the minimal amount of time and resources necessary. Skinner, Belfiore, and Watson (1995) define instructional efficiency within the construct of learning rate i.e., how much a student learns within a unit of time. Time is arguably the most important consideration in schools, and is crucial in multi-tier support models (Barnett et al., 2004; Burns & Sterling-Turner, 2010). When deciding between methods that are equally effective, consideration must be given to the time to implement the strategy and other resources such as effort, teacher availability, and cost. The strategy that takes the least exertion, expense, and time is preferential (Konrad et al., 2014).

Considering the availability of myriad evidence-based educational interventions, and the time constraints in a school day, academic gains expressed as a

function of time is possibly the most important metric for the educator to consider (Bramlett et al., 2010). Reporting learning rates provides educators with a practical standard that can inform intervention selection based on student instructional need within a given timeframe (Burns & Sterling-Turner, 2010). Put simply, it is not enough to know that a strategy works; rather, it is which strategy will accelerate learning the most (Skinner, 2010). Describing gains in this way also creates a more standard unit that can be understood by all stakeholders. Indeed, many researchers (e.g., Nist & Joseph, 2008) have demonstrated that omitting efficiency data can lead to misleading results that ultimately promote interventions that impede rather than accelerate student growth (Skinner, 2010).

Despite the importance of this variable when designing interventions in limited timeframes and evaluating outcomes (Skinner, 2008), instructional efficiency is frequently overlooked (Power, 2008). Bramlett et al. (2010) reviewed research pertaining to academic interventions in four school psychology journals published between 1995-2005. Despite the importance of the time required to implement interventions, the authors found that the majority of researchers investigating learning procedures did not accurately measure the amount of time spent implementing those procedures. The authors assert that instructional efficiency data plays an integral in the educator's evaluation of instructional effectiveness, and that researchers should report both types of data.

Cates, Burns, and Joseph (2010) reiterate this argument; research that examines effectiveness should examine efficiency, but also in consideration of the entire IH. A number of research studies (e.g., Cates et al., 2006; Joseph, Eveleigh, Konrad, Neef, & Volpe 2012) have demonstrated that two or more interventions with similar effectiveness and efficiency may not equally impact levels of the IH other than

acquisition (e.g., maintenance). Burns and Sterling-Turner (2010) assert that maintenance is a critical concern when considering the potential time taken to reteach skills that are not retained. The authors provide the example of a student who is taught five words in five minutes (one word per minute of instruction), but only retains two words, resulting in the other three being retaught. This would mean that including time taken to reteach words, the total instruction time is eight minutes and the efficiency rate would be 0.63 words per minute.

1.9.2. Instructional Intensity. Intrinsically related to the measurement and reporting of intervention efficiency is the construct of intervention intensity. Interventions need to be implemented with sufficient integrity and intensity to achieve desired outcomes for students (Coddling & Lane, 2015). Similar to intervention effectiveness, intervention integrity has received much attention in regard to evidence-based academic interventions. However, similar to intervention efficiency, focus on intervention intensity is lacking to date despite its importance in planning and interpreting intervention research (Coddling & Lane, 2015).

RTI models are based on tiers of increasing instructional intensity (e.g., Daly et al., 2007; Gersten et al., 2009), however, little is known to date about treatment intensity in RTI frameworks (Coddling & Lane, 2015), and there is a dearth of evidence defining or describing treatment intensity (Daly et al., 2007). This is a concern as the higher tiers (usually Tier 2 and Tier 3) increase in intensity, when the lower tier has not met the instructional needs of the student (Mellard et al., 2010). Understanding the nuances of intervention intensity is therefore crucial for implementation of RTI frameworks (Barnett et al., 2004; Coddling & Lane, 2015). Barnett and colleagues (2004) argue that to determine eligibility decisions, single case research designs that focus on an analysis of the intensity of the intervention should

be employed. This permits identification of the most effective, but least intrusive intervention to be delivered. Due to its centrality in the provision of multi-tier support, intervention intensity requires a robust definition and consistency in application within the literature (Yoder & Woynaroski, 2015).

Codding and Lane (2015) reviewed three theoretical articles on intervention intensity (Barnett et al. 2004; Mellard et al., 2010; Warren, Fey, & Yoder, 2007), identifying five overlapping aspects of intensity: treatment session length (measured in minutes); treatment session frequency (measured per day/week); total treatment duration; number of practice opportunities, and interventionist characteristics. Treatment dose was described as session length and frequency, and across two of the articles papers as total treatment duration. According to Codding and Lane (2015), lack of consensus in the literature as to what constitutes intervention intensity makes it difficult to evaluate research outcomes, and may misleadingly be a source of focus for intervention change when a simpler solution exists.

Harn, Stoolmiller, and Chard (2008) reported that students receiving a more intensive intervention package (defined as double the intervention time provided daily) made significantly more progress across a range of early reading measures. This study frames intensity in terms of instructional time, however, more time may not necessarily always be better. Truch (2003; as cited in Torgesen, 2005) reported that rate of gain decelerates rapidly for intensive interventions after the first 12 hours of intervention. Indeed, Daly and colleagues (2007) caution against intensifying an intervention based on non-responsiveness. These authors assert the more parsimonious solution may yield better results; by creating a better fit between the instruction and the student's level of proficiency. They also assert that intensifying an

intervention may not address the problem if the intervention is not matched to the student's instructional needs.

1.9.3. Precision Teaching: Intensive, Effective, and Efficient. PT can be considered an efficient instructional methodology as it increases rates of learning. In addition, its measurement system affords a precise mechanism to record and evaluate rates of learning and level of instructional intensity. By recording the amount of time spent in PT sessions, the educator can frame learning outcomes in terms of total minutes of intervention time (error correction, charting etc.), the number of frequency building trials, and minutes spent frequency building. This provides a standardised system that is interpretable by any educator and permits performance comparisons, which can be combined with qualitative information to provide context. Figure 6 describes effectiveness and efficiency data in an applied example of a seven year-old male student (Abe) in February of the Senior Infants grade¹.

¹ Abe was a participant in the current research (see Table 5, Chapter 2).

At pre-test Abe could correctly produce one letter sound from 26, and his frequency level was 0 correct per minute (CPM) and 44 errors per minute (EPM). Abe received the PT intervention programme targeting letter sounds for a total of 2.8 hours of intervention provided over 20 intervention sessions, within which he completed approximately 80 frequency building trials. At the end of the intervention, Abe could correctly produce 24 of the 26 letter sounds, and his frequency levels were 52 CPM/4 EPM. Importantly, when tested nine months later his frequency levels were 56 CPM/8 EPM. In addition, his rate of learning could be expressed as a celeration and directly compared to his baseline performance (regular classroom instruction). Abe's baseline celeration was (x1.9), this increased to (x8.1) after five days of intervention. Prior to the PT intervention Abe had not learned letter sounds in 1.5 years of typical classroom instruction (i.e., the *Jolly Phonics* curriculum). Abe did not have a learning disability; he required explicit instruction, practice, systematic error correction, and reinforcement in a motivating environment.

Figure 6. Real world Example of Reporting Effectiveness and Efficiency Data Combined with Qualitative Information.

PT practices permit quantification of intervention effectiveness in terms of level and rate. In addition, due to its measurement system it permits intervention intensity to be reported with exactness. This provides invaluable information in determining the overall effectiveness of the intervention, by framing a student's "response" in terms of efficiency and intensity. Students' responding can be further demonstrated using single case experimental designs (SCED). Because PT practices yield daily performance data, it can easily be integrated into SCED, and its focus on

frequency as a metric permits aggregation of such data for advanced statistical analyses such as multi-level modeling.

1.10. Single Case Experimental Designs

Single case experimental designs (SCED) play a critical role in data-based decision making within RTI frameworks (White, 2009; Riley-Tilman & Burns, 2010). A core tenet of RTI is collecting student performance data before, during, and after an intervention, and subsequently comparing performance levels and patterns across these situations. This is the logic of the single case experimental design, repeated measurements of performance across time and conditions to make meaningful performance comparisons, to monitor progress, and ultimately to evaluate response to the intervention (Riley-Tilman & Burns, 2010).

Students' response to intervention can be demonstrated using a variety of SCED (Barnett et al., 2004; Riley-Tilman & Burns, 2010; White, 2009). SCED are a valid methodology for establishing empirical interventions (Stoiber & Kratochwill, 2000), as they can provide a strong basis for causal inference (Kratochwill et al., 2010). By creating data sets of a student's performance, SCED assists in organising the construct of "response to intervention" (Barnett et al., 2004). Riley-Tilman and Burns (2009) state that: "The effective application of intervention, assessment, and SCED allows for fully defensible statements as to a child's response to the intervention" (p. 137). This highlights the main advantage of SCED: school-based teams can use them as scientifically supported methods for making special education decisions (Barnett et al., 2004). In addition, SCED provide the mechanism for school psychologists to merge research with practice (Ross & Begney, 2014). A review of school psychology journals revealed that SCED represented 55% of all "causal

experimental” designs (Bliss, Skinner, Hautau, & Carroll, 2008).

The growing popularity of SCED may be due to the number of other advantages of these designs in RTI models. Specifically, SCED are a highly feasible method of conducting applied research, and no control group is required, therefore no student is denied the intervention (Riley-Tilman & Burns, 2010). In addition, SCED are highly flexible; their logic can be used to build a number of different designs that fit the research question and the applied setting (Swanson & Sachse-Lee, 2000; Riley-Tilman & Burns, 2010).

Several single-subject experimental designs have been devised for both evaluating individual interventions and comparing interventions (Gast, 2010). SCED have been used to establish the effectiveness of a number of educational interventions including: oral reading fluency (Weinstien & Cook, 1992); phonics (Reason & Morfidi, 2010); reading comprehension (Treptow, Burns, & McComas, 2007); geography facts (Barbetta, Heron, & Heward, 1993), and, mathematic problem solving (Case, Harris, & Graham, 1992). In particular, multiple-baseline designs are highlighted in the literature as very suited to implementing literacy research in applied settings (Riley-Tilman & Burns, 2009).

The multiple-baseline across participants design yields data on several students’ response to an intervention, permitting the researcher to assess the effects of the instructional approach, and to detect individual variation in response to the intervention (Axelrod; 1983). Moreover, replications of intervention effect across participants may corroborate the reliability of the results, and by replicating the design with a different group of participants some degree of generalisability may be achieved (Kucera & Axelrod, 1995). Multiple-baseline designs have to date provided evidence

for intervention effectiveness in literacy (e.g., Barger-Anderson, Domaracki, Kearney-Vakulick, & Kubina Jr. 2004; Reason & Morfidi, 2001).

1.11. Standardised Scores

In addition to using SCED to evaluate a student's individual response to intervention, it is important to evaluate the student's progress in relation to average readers. Torgesen (2005) states that: "standardised scores are an excellent metric for determining the 'success' or 'failure' of interventions for children with reading disabilities" (p.7). This is because standardised scores describe the student's performance in relation to a large standardisation sample. If an increase is observed in the standardised score post-intervention, this indicates that a student has "closed the gap" with average readers (ibid). Many UK studies do not report results in standardised scores, instead using reading ages, which are sometimes calculated into ratio gains to evaluate the effectiveness of the intervention (Rose, 2006). Ratio gains generally refer to the gain in reading age made during a chronological time span, expressed as a ratio of that time span. Ratio gains are reported widely as oftentimes they are the only impact measures that can be calculated (Brooks, 2007).

Though apparently easier to interpret, reading ages are considered "statistically unsatisfactory" (Brooks, 2007, pp. 178). Establishing whether gains in test scores are statistically significant is more challenging for reading ages than it is for standardised scores. Even though RG can be calculated from a reading age, this statistic is not considered to be useful with low attaining groups. This is because students in such groups may not be expected to make a one month gain in reading age, for example, in one calendar month (Brooks, 2007). However, a different type of ratio gain can be calculated using the standardised score and is referred to as the

standardised score ratio gain (SSRG). The SSRG statistic permits a crucial analysis of the effectiveness of any reading intervention: the time effectiveness of such interventions (Torgesen, 2005).

Importantly, the SSRG is an efficiency metric and represents the rate of growth in reading calculated by pre- to post-test changes per hour of intervention (Torgesen, 2005). This allows outcomes from interventions of differing length to be directly compared, and frames effectiveness in terms of efficiency. There are significant benefits in using such a metric, as the SSRG is likely to be a more sensitive measure where the total numbers of hours of intervention per student is comparably short (Savage, Abrami, Hipps, & Deault, 2009). Across 14 studies of dyslexic readers, Torgesen (2005) reported effect sizes at immediate post-test ranging from 0.19 - 0.30 across measures and intervention conditions. A summary of the SSRG for typically developing children across a range of high quality studies is currently not available (Savage et al., 2009). However, a small number of studies have been identified: Hatcher, Hulme, and Ellis (1994) demonstrated SSRG in the 0.31 - 0.39 range; McGuinness, McGuinness, and McGuinness (1996) reported SSRG in the 1.7-2.6 range; Hatcher et al., (2004) in the 0.13 - 0.23 range, and Savage et al., (2009) in the -1.7 - 0.70 range.

To conclude, success in early reading is synonymous with academic achievement and critical for future reading and cognitive development. Fluency in sublexical skills and high frequency words lays the groundwork for this success and should be targeted in kindergarten. Learning to read in English is a complex process; therefore to optimise early reading instruction the most generalisable linguistic units and words should be targeted. Students at risk for reading difficulty need to be identified early and provided with skill specific intervention that accelerates learning

rates with sublexical skills. This can be accomplished through an RTI framework that incorporates PT. PT is well matched to RTI frameworks as it provides for an effective fluency-based instructional method, and a precise system of measurement that facilitates progress monitoring and the consistent application of decision rules. Response to intervention can be demonstrated using PT practices and procedures, and further verified by incorporating single case experimental designs and standardised scores.

1.12. The Current Research

The current research involves the application of a Tier 2 intervention programme targeting fluency in foundational reading skills with at risk Senior Infant readers (kindergarten equivalent). Although PT has been highlighted as particularly suitable to RTI frameworks (Johnson & Street, 2013, Kubina & Yurich, 2012; White, 2009), the current research involves the first application of PT as a Tier 2 intervention. The Tier 2 foundational reading skill programme includes a PT intervention programme that combines: (a) fluency intervention; (b) progress monitoring, and (c) decision rules for programme modifications.

The PT intervention programme uses a training protocol to target fluency in foundational reading skills. Snowling and Hulme (2010) propose that for an intervention to be considered “well founded” (p. 3), it must be based on a causal model of how a skill develops, and how to promote that skill in students who are struggling to reach mastery. The current research adopts a causal model of reading development based on the stimulus control paradigm and IH, and uses behaviour analytic models of effective instruction to promote core reading skills.

For an intervention to assist struggling readers to catch up, it must accelerate their rate of learning (Skinner, 2010). PT is selected as the intervention as it directly targets fluency and accelerates rates of learning. In addition, the strategies employed in the PT intervention (e.g., goal specificity, feedback) are effective instructional ingredients for targeting at risk and struggling readers (Bramlett et al., 2010; Konrad et al., 2011; Mellard et al., 2010). The foundational reading skills targeted for fluency include letter sounds, phonemic awareness, decoding, and high frequency words. Letter sounds and phonemic awareness are two critical foundational reading skills for learning to decode print (Adams, 1990; Snowling & Hulme, 2010). Mastery of these skills is critical for at risk kindergarten readers (Gersten et al., 2009).

Proficiency in reading, however, is entirely dependent on fluent decoding (Snowling & Hulme, 2010). In the current research, decoding is targeted for fluency using a set of the most frequently occurring GPCs (Carnine et al., 1997; Solity & Vousden, 2009) embedded in real words. In addition, high frequency words must be automatised for optimization of early success in reading (Solity & Vousden, 2009; Stuart et al., 2003). Therefore, sets of the 100 high frequency words (Solity & Vousden, 2009) are targeted with the PT intervention programme.

The effectiveness of the intervention is investigated using SCED and pre-post-test standard score changes. SCED can demonstrate students' response to intervention (Barnett et al., 2004), and are considered an effective methodology for establishing educational interventions (Riley-Tilman & Burns, 2010). In reporting the SCED, performance data are presented in terms of baseline performance, and increase as a result of the intervention (at the 3rd intervention session, and at intervention end). In addition, because the goal of Tier 2 is to help students close the gap with average performing peers, norm-referenced standardised assessments were administered as

pre- and post-tests. Gains in standard scores indicate that students have closed the gap (Torgesen, 2005). Moreover, SSRG (ratio gains) were computed to provide a measure of instructional efficiency, and to permit comparisons with interventions of differing durations.

Instructional intensity of the intervention is reported in terms of total intervention duration, number of intervention sessions, and average duration of intervention sessions. Reporting instructional intensity in this way permits exact quantification of performance data, by providing a measure of instructional efficiency. In addition, because intervention effectiveness and efficiency should be considered in terms of its effect on the IH (Cates et al., 2006; Daly et al., 2007), maintenance data ranging from three weeks to nine months are reported. These data reveal the long-term effects of building frequencies in foundational reading skills.

Progress monitoring is a core component of Tier 2 interventions (Gersten et al., 2009). The PT treatment package uses the SCC as the progress-monitoring tool for the intervention. Due to the important measurement and data analysis characteristics described, the SCC is the only appropriate chart for use in fluency building instructional programmes (Binder, 1990; Fabrizio, 2003). The SCC was used to graphically display performance and to set goals and to illustrate the mastery criterion for the student. The SCC permitted level and rate of learning to be recorded and quantified for each participant. This permitted measureable performance data to form the basis for making decisions about adequate response to the intervention.

The focus of the intervention is to accelerate – and measure – rate change over time (i.e., learning). A doubling of rate was selected as the target rate of acceleration (i.e., a x2 celeration). Celeration is a quantitative measurement of learning over time. Using a x2 celeration as a rate criterion to determine adequate “response” to

intervention provided clear parameters for data based decision making, and facilitated the consistent use of decision rules to be applied the intervention. Where a student maintained or exceeded a $\times 2$ the intervention was not modified, conversely, when a student's rate of learning fell below a $\times 2$ the intervention was slightly modified in an attempt to accelerate rate of learning. Modifications consisted of antecedent and consequence changes, and were systematically applied across students when performance data demonstrated their need (i.e., two days below $\times 2$ celeration).

Chapter 2: Introduction

The concepts and practices in both RTI and PT are unfamiliar to educators in Ireland. The current research was the first application of PT targeting fluency in foundational reading skills in the Irish primary school system, and the first Tier 2 intervention programme provided to Senior Infant students. Accordingly, its implementation and evaluation were not within the context of an overall RTI framework. Core components of Tier 2, however, were adhered to: universal screening, targeted intervention, progress monitoring, and consistent use of decision rules and data-based decision making (Gersten et al., 2009).

As this research was the first to investigate PT as a Tier 2 intervention programme, and provided a unique focus in Irish classrooms, it was imperative that the universal screener accurately identified students considered at risk in areas of letter knowledge and phonemic awareness. In addition, it was important to demonstrate that the intervention was effective in helping at risk readers catch up with average performing peers, demonstrated through the analysis of standard score gains (Torgesen, 2005). For this reason, it was essential that the screening instruments employed were norm-referenced.

To meet the above requirements, a specific screening instrument was employed, the York Assessment for Reading Comprehension (YARC, Hulme et al., 2009). This measure was chosen for the following four reasons: it focused on foundational reading skills (letter sounds, phonemic awareness, and word reading); it was standardised in the United Kingdom and therefore most closely aligned with Irish educational norms; it identified students on a continuum of skill difficulty according to a large standardisation sample, and accordingly yielded standardised scores.

There is much debate about how to identify students who are considered at risk for reading difficulties (Reynolds & Shaywitz, 2009), and while static and standardised assessments have been criticised (Hosp, 2008), they were utilised within the current research. The use of other assessments, standardised in countries less aligned to the Irish primary school curriculum, and therefore less resembling educational norms, may not have demonstrated adequate sensitivity during screening. Although the YARC measure did not remove the possibility of false positives being identified, it did however reduce the likelihood of false negatives.

An important consideration of the research was to provide skill specific interventions that matched the universal screening outcomes of participants identified as at risk. In this way, intervention provision was focused on the foundational reading skills presenting as most deficient. Accordingly, each participant received differing configurations of interventions, and for different durations. For example, in the universal screening phase Abe² was identified as at risk for reading difficulties in the skills of letter sounds and a phonemic awareness skill (sound isolation). To match the intervention focus to the universal screening outcomes, Abe received the intervention for both of these skills. Other participants for example, may have required intervention for three foundational reading skills. For every participant and each skill targeted, the PT intervention programme combined: (a) fluency intervention; (b) progress monitoring, and (c) decision rules for programme modifications.

The attempt to provide authentic intervention provision (i.e., matched to participant's skill needs) brought with it inherent complications for teasing out intervention effects. The effectiveness of the intervention programme was experimentally investigated within the context of the foundational skill targeted. This

² Abe was a participant in the current research; see Table 5 for participant details.

meant that a participant necessarily took part in more than one experiment if he or she demonstrated difficulties in more than one skill. In Abe's case, he participated in two experiments, as he demonstrated difficulties in both letter sounds and in phonemic awareness. Each experiment evaluated the effect of the PT intervention on the specific foundational skill targeted.

Chapter 2 presents four experiments: each investigates the effectiveness of the proposed PT intervention programme as it applies to a specific foundational reading skill. The foundational skills targeted with the PT intervention were: letter sounds (Experiment 1); letter names (Experiment 2); the phonemic awareness skill of sound isolation (Experiment 3); and the phonological/phonemic awareness skill of sound deletion (Experiment 4). The four experiments directly investigate the short and long-term skill specific outcomes of the PT intervention on the foundational skills targeted. Finally, in order to draw conclusions on the outcomes of the PT intervention programme across participants generalised reading outcomes pre- and post-intervention are presented.

2.1. Experiment 1: Letter Sounds.

Converging evidence highlights the role of letter sounds to reading development (Gersten et al., 2009; Snowling & Hulme, 2010), moreover, the importance of fluency in this sublexical skill to accurate and fast decoding is undisputed (Burke et al., 2010; Verhoeven & van Leeuwe, 2009). Despite this, fluency is infrequently targeted in early grades (Bursock & Blanks, 2010), and fluency specifically targeting letter sounds even less frequent. A small number of studies, however, have been identified. Duhon, House, Poncy, Hastings, and McClurg (2010) targeted letter sound fluency with three 1st grade students identified as

struggling readers by teachers. At pre-test all participants demonstrated 100% accuracy for the 26 letter sounds. These authors used a multiple-baseline across participants design to investigate the effects of the instruction on letter sound fluency, and to evaluate two generalisation procedures (cueing and providing sufficient response exemplars) on decoding words. Their results demonstrated an average gain in letter sound frequencies of 26.7 CPM.

Other researchers have used a procedure known as incremental rehearsal (IR) to teach letter sounds to kindergarten students where accuracy was not at 100%. IR is a drill and practice method used to teach discrete unknown concepts by intermixing them among known concepts at a ratio of 90% known items to 10% unknown items (Burns, Dean, & Foley, 2004; Peterson et al., 2014). IR has previously been used to teach letter sounds via a computer (Volpe, Burns, DuBois, & Zaslofsky, 2011), and more recently targeting students that speak English as a second language (EL2) without the aid of a computer (Peterson et al., 2014). Using a multiple-baseline across sets of letter sounds Volpe and colleagues (2011) demonstrated gains ranging from 13-23 letter sounds CPM after 25 intervention sessions. Similarly, Peterson and Colleagues (2014) used three multiple-baseline across sets of letter sounds to evaluate the effect of IR on letter sound acquisition/fluency. Results showed increases in level and trend in the intervention phases. From a baseline of 12-15 sounds per minute three EL2 participants gained on average 36.3 letter sounds per minute, which is a rate of 3.8 letter sounds per week. Exact levels of accuracy were not provided for this study, however 12 unknown letters were targeted for each participant.

The first experiment described in Chapter 2 extends these findings in a number of ways. First, this experiment is the only Tier 2 intervention to date that specifically targets letter sound fluency. Second, participants in the experiment were selected

through universal screening; in previous research participants were teacher nominated. Third, the Tier 2 intervention programme demonstrates the use of data-based decision making using decision rules to strengthen participants' response to the intervention. Finally, unlike Duhon et al.'s study, 100% accuracy in letter sounds was not a pre-requisite for participant exclusion, and the current study targeted younger participants at Senior Infant level or kindergarten equivalent. Furthermore, earlier studies (Peterson et al., 2014) employed IR, which requires that a level of accuracy prior to intervention because known items are interspersed with unknown items. The current Experiment did not require a specific level of accuracy by participants. Similarly, the current research targeted at risk readers, rather than any specific group e.g., English language learners. Experiment 1 also extends previous findings by reporting treatment outcomes after prolonged periods post-intervention, contributing to an evaluation of the efficiency of the intervention.

The purpose of Experiment 1 was to implement a PT intervention programme as a Tier 2 intervention with Senior Infant readers identified as at risk. Through an evaluation of maintenance data, Experiment 1 aimed to investigate the effectiveness of the PT intervention in establishing letter sound fluency. This Experiment also aimed to evaluate the effect of the intervention in helping participants close the gap with average performing peers in the target skill of letter sounds. The intervention goal involved achievement of performance standards in letter sounds (70-100 correct per minute; K. Brooks Newsome, personal communication, February 5, 2012) at a learning rate of x2 celeration.

2.1.1. Method

2.1.1.1. Participants and Setting

A Senior Infant class (kindergarten equivalent, $n=14$, age range 5.5-6.8 years; $M = 6.1$) within a DEIS (“Delivering Equality of Opportunity in Schools”) school participated in the study. DEIS schools are situated in educationally disadvantaged areas in Ireland, with a high percentage of the school population experiencing socio-economic disadvantage. Parents received participant information letters (see Appendix A and B) through the school with an option to consent to participation. This resulted in a sample of 12 participants (age range 5.6-6.8 years; $M = 6.5$) who were screened to identify students in need of additional instruction in letter sounds.

2.1.1.2. Universal Screening and Pre- Post-Test Measures

The screening instrument utilised was the Letter Sound Knowledge (LSK) subtest of the York Assessment of Early Reading Comprehension (YARC; Hulme, et al. 2009). The LSK screener also served as a pre- post-test measure of outcomes in letter sound accuracy. The LSK subtest consists of 32 items that measure letter sound knowledge. The reported reliability for this subtest is .98 (Cronbach’s alpha) and its predictive validity coefficient with the Single Word Reading Test (Foster, 2007) is .55 (Hulme, et al. 2009). A student is considered to have a severe difficulty in letter sounds if he or she achieves a standard score between 70-79, and to have a moderate difficulty if he or she achieves a standard score between 80-84.

Five participants were identified as having a severe difficulty (P1, P2, P4, P5, P6; age range 5.7-6.18; $M = 6$ years), and two were shown to have a moderate difficulty (P3, P7; age range 5.5-5.9, $M = 5.7$) in letter sound knowledge. These seven participants (three female; four male) were selected for intervention in letter sounds. All participants were typically developing and one participant (P5) attended monthly

speech and language therapy sessions to address speech difficulties.

2.1.1.3. Materials

Intervention materials were stimulus sheets that consisted of letters generated using the letter naming fluency generator (available from <http://www.interventioncentral.org>). Each A4 stimulus sheet consisted of printed letters (11 across and 10 down) in size 14 font. Twenty multiple exemplars were created, and three stimulus workbooks were compiled and bound using all 20 exemplars (two experimenter copies and one participant copy). This resulted in randomized exemplars for each measurement occasion, controlling for practice effects.

Paper versions of the timings per minute standard celeration chart (TPM SCC), the daily per minute standard celeration chart (DPM SCC), and the electronic version of the one-year (TOY) Standard Celeration Chart (Graf, Auman, & Lindsley, 2007) were also used. Data collection materials included mechanical pencils, daily data collection sheets, and two electronic timers.

2.1.1.4. Design

A multiple-baseline across participants design was employed to assess the effectiveness of a PT intervention programme in building frequencies in letter sounds to performance standards.

2.1.1.5. Dependent Variables

Participant outcomes in pre- and post-test assessments and performance data of letter sound rates were used as dependent variables in the experiment. Letter sound rates were measured repeatedly across baseline, intervention, and post-intervention conditions. The intervention performance standard was 70-90 correct letter sounds per minute. Post-intervention letter sound rates were measured up to nine months post-

intervention to investigate maintenance of treatment effects. Pre- post-test gains in a standardised norm-referenced assessment provide further evidence of intervention effectiveness in targeting letter sound accuracy and fluency.

Celeration and performance standards were used to set the criteria for baseline conditions and the PT intervention. Because celeration is a standard measurement, magnitude of effect within and across conditions can be quantified, and celeration values offer enhanced quantitative and visual analysis power (Datchuk & Kubina, 2011). A multiply or divide symbol (\times for an increase; \div for a decrease) is used to communicate celeration in writing, for instance, a $\times 2$ celeration means performance has doubled in frequency, and a $\div 2$ celeration means that it has halved in frequency. For example, a learner who reads 40 words per minute on Week 1 and reads 80 words per minute on Week 2 has doubled his performance, or has a celeration of $\times 2$ (over 7 days). The intervention focus was for participants to achieve performance standards in letter sounds (70-90 CPM) at a learning rate of $\times 2$ celeration.

2.1.1.6. Procedure

Baseline. Baseline probes demonstrated current rates of responding in the absence of the PT intervention to compare to rates of responding following intervention. Stimulus sheets designed for intervention were also used during baseline. A novel stimulus worksheet was used for each baseline measurement to control for practice effects. Letter sounds were probed one or two times per week until a minimum of six probes were collected. Probes were timed during a 15 second duration. Rates of responding were recorded on a daily data collection sheet, on the DPM SCC and on the TOY.

Prior to beginning the timings, participants were told to do their best, to skip a response if it was too difficult, move quickly to the next item and to continue

responding until the timer sounded. Skipped responses were counted as incorrect responses, and if the timer sounded during a response it was recorded as a correct or incorrect response accordingly. Participants did not receive programmed feedback or reinforcement but were each praised equally for their effort and cooperation. No reference was made to speed during this phase. Participant responses were tracked on an experimenter version of the academic stimulus sheet.

Baselines were completed when one, or all, of the following conditions were satisfied: (1) celeration $\leq X1.4$ for correct responding; (2) data were highly variable (i.e., lots of fluctuation around the mean), and/or, (3) errors were consistently observed in the participant's performance. These baseline criteria are based on student performance data, permit quantification between the baseline and intervention condition, and have been used in previous research (Brooks Rickard, 2010; Berens, 2010). When baseline criteria were met, participants were entered into the intervention phase at staggered start points in a multiple-baseline across participants design. Each participant entered intervention when the previous intervention participant demonstrated "response" to the intervention for three consecutive days (i.e., maintaining a $\times 2$ celeration).

2.1.1.7. Intervention

The independent variable involved a Tier 2 PT intervention programme that included PT as an intervention to target fluency in letter sounds, and a decision-making framework to monitor progress. Components of the PT intervention included discrete trials, frequency building to performance standards, progress monitoring and, using decision rules to make data-based instructional changes (each described in detail below). Students received the one-to-one intervention five days per week, daily sessions lasted approximately nine minutes, and the average total intervention

duration was 2.4 hours. Each intervention session consisted of discrete trials in letter sound discrimination and production, and frequency building to performance standards.

Discrete Trials. Discrete trials were used to increase accuracy immediately prior to fluency trials for six of the seven participants. This consisted of presenting the participant with the stimulus sheet on which previous errors were circled in the top two lines. The experimenter pointed to a circled letter (e.g., “p”) and modeled the appropriate sound (/p/). The participant was then asked to repeat the sound, and to quickly discriminate instances of the target letter on the sheet producing the sound of the letter as he/she pointed to instances. Verbal praise was provided for each correct discrimination and production of the letter sound.

Frequency Building to Performance Standards. The timing trials in intervention were identical to that in baseline with the introduction of the following elements:

- 1) Goal Setting. A goal line for celeration was calculated by doubling (x2) the median response frequency obtained in baseline. Subsequently, weekly celeration goals were set as a doubling (x2) in response frequencies per week. Each student used a star chart to display the performance goal for the session;

- 2) Timed practice trials. Fluency trials lasted 15 seconds. The learner practiced decoding single letter sounds within the timed trial, and the experimenter counted and marked correct and incorrect responses on an experimenter copy;

- 3) Feedback. At the end of the trial the experimenter provided feedback on performance to the learner. The experimenter provided verbal praise for an increase in performance and encouragement when a decrease in performance was observed. At this point, performance data were transferred to a timings SCC and the learner could determine his or her performance;

4) Error Correction. The experimenter then systematically corrected any errors made during the timing trial (using errorless learning). Steps 2-4 were repeated five times (i.e., five timings were completed within the session);

5) Charting of Performance. Each timing trial was charted on the timings SCC. The median score from the five trials was plotted on the daily SCC;

6) Differential Reinforcement. The learner was given differential tangible reinforcement for performance described as follows:

- No goal achievement - super effort sticker
- Goal achievement - goal sticker
- Performance Standard achievement - aim star sticker
- Performance Standard maintained for two days – selection of a small tangible item from the treasure chest
- At the end of each intervention session, the experimenter would play a game of ‘X’s and O’s’ with the participant, in each instance the participant would invariably win.

Progress Monitoring and Decision Rules. Decision rules to guide PT intervention were based on participants’ celeration and goal attainment. When a x2 celeration was maintained, the participant was deemed to be responding and no change was made to the intervention. When a x2 celeration was not maintained for two consecutive days the participant was considered non-responsive and antecedent changes were made. Antecedent changes included modelling speed and using flashcards of previously made errors. If a participant was unable to maintain a celeration of x2 for one day after antecedent manipulation, a consequence change known as the personal best component (PBC) was made (3rd consecutive day of goal failure). The PBC is a procedure commonly used in Precision Teaching programmes as a method for

increasing response frequencies (Binder, Haughton, & Bateman, 2002; Brooks Rickard, 2010). Using the PBC the daily goal was set for one response higher than the previous highest frequency. For example, if a student's highest frequency obtained was 10 correct responses, a goal was set on the subsequent session for 11 correct responses.

Termination of the intervention at a reduced criterion was determined as a function of time constraints, or where the frequency level attained was considered to approximate a functional threshold and/or was considered an acceptable level of performance in relation to the celerations observed as a result of the intervention. For example, the participant Abe described in Chapter 1 demonstrated near zero levels of correct letter sounds in baseline and a celeration of $\times 1.9$. The introductions of the intervention resulted in a celeration of $\times 8.1$ after 5 days, after 2.8 hours of the PT intervention Abe demonstrated frequency levels of 52 CPM. Training was terminated due to time constraints; at nine months post intervention his performance maintained at 56 letter sounds CPM.

2.1.1.8. Interobserver agreement (IOA).

The Exact Count-Per-Interval method was used, as it is the most stringent description of IOA for data sets obtained by event recording (Cooper, et al., 2007). Exact Count-Per-Interval is the percentage of total intervals in which two observers recorded the same count. Each timing trial was considered an interval; the experimenter and an observer observed and recorded participants' responding. If there was disagreement in the number of corrects or incorrects between the experimenter and observer the trial was identified as "disagreed". Dividing the number of intervals where there is agreement by the total number of intervals and multiplying by 100 calculates this index. IOA data were collected for 29% of all baseline sessions

demonstrating 100% agreement, and 28.5% of all intervention sessions achieving 98.8% agreement. Overall that is 99.4% agreement across 28.8% across baseline and intervention sessions.

2.1.1.9. Procedural integrity (PI)

The PT intervention programme involved multiple components, and there were important procedural distinctions between the baseline and intervention sessions; therefore measurement of procedural integrity (PI) was crucial. A procedural integrity checklist for baseline (see Appendix C) and for intervention sessions (see Appendix D) were constructed to measure the degree to which steps were carried out as originally intended (as described in the procedure section). An independent observer collected PI data on the experimenter's organization of instructional materials, instructional delivery, data collection and analysis, error correction procedures, and reinforcement. Levels of PI data were calculated by counting the number of steps where protocol was adhered to, divided by the number of steps that were observed, multiplied by 100. PI data were collected for 26.5% of baseline sessions demonstrating 98% adherence, and for 23% of intervention sessions demonstrating 96% adherence.

2.3.1.10. Data analysis

The PT intervention programme was assessed using a multiple-baseline design, in conjunction with pre- post-test changes in standardised and criterion-referenced assessment outcomes. In addition, long-term intervention effects were evaluated using retention data collected at multiple measurement occasions spanning up to nine months post-intervention.

Comparing performance rates between conditions in a multiple-baseline design can convey magnitude of intervention effect. Data were evaluated with respect

to median baseline performance, intervention effect by the 3rd training session and at the end of training. Gains from median baseline to the 3rd intervention session are described as gains in correct letter sounds per minute. Such a gain demonstrated the immediacy and degree of intervention effect. Reporting performance at intervention end permitted the overall evaluation of intervention effect. A student's response to the intervention was indexed using celeration - a participant was deemed to be "responding" to the intervention when maintaining or exceeding a x2 celeration. Conversely, "not responding" to the intervention was defined as falling below a x2 celeration and experiencing goal failure over two consecutive days.

Standard scores are an exceptional metric for evaluating the success or failure of an intervention for children with reading disabilities, as an increase in standard scores indicates a closing of the gap with average readers (Torgesen, 2005). Results are presented in terms of pre-test-post-test changes in standardized scores for letter sounds. In addition, standard score changes are described in relation to hours of intervention received, known as a standard score ratio gain (SSRG; Torgesen, 2005). The SSRG describes the rate of growth in reading as calculated by pre- post-test changes per hour of intervention. Summaries of intervention features, such as the average length of time to reach performance standards are also provided.

Effective interventions should demonstrate immediate-, short- and long-term gains (Suggate, 2014). Maintenance is one of the core indicants of behavioural fluency (Kubina & Yurich, 2012); and high maintenance of academic skills over prolonged periods of time would indicate that the intervention was effective in establishing fluency in letter sounds. Maintenance of letter sounds was assessed over five maintenance checks that span up to nine months post-intervention. Increases and decreases in performance across conditions and maintenance checks are described as

a *fold* change - the ratio between the final and initial value. A fold change is useful for evaluating increases and decreases in data; its benefit is that the change itself is emphasised - not the absolute value (Siegel & Castellan, 1988). This creates a standard unit of measurement that can be used to compare increases and decreases across conditions and maintenance checks.

2.1.2. Results and Discussion

Figure 7 depicts a multiple-baseline across participants. All participants entered the baseline condition at the same time, and a minimum of six baseline probes were collected before P1 was entered into the intervention phase. Subsequent participants were added in a sequential fashion - when intervention participants demonstrated *response* to the intervention for 3 consecutive days (i.e., maintaining a x2 celeration).

Figure 7 shows that P1 responded to the intervention, gaining 20 letter sounds CPM by the 3rd intervention session. Absenteeism and a mid-term school break resulted in a flat trend before P1's performance increased again and reached performance standards (76 CPM) on the 12th intervention session. Participant 2 was then entered into the intervention and also responded - gaining 20 letter sounds CPM by the 3rd intervention session. Following the mid-term school break, performance regressed to baseline levels. Upon reintroduction of the intervention, P2 immediately responded and this pattern continued for three weeks. Following this, P2 experienced goal failure resulting in, programme modification (modeling speed), and subsequently a personal best component. Despite modifications her performance maintained, and intervention was terminated at the reduced criterion of 40 CPM on the 24th intervention session.

The intervention was then introduced to P3 - who was considered responsive -

gaining 28 letter sounds CPM by the 3rd intervention session. Figure 7 shows that P3 continued to respond for two weeks, and then experienced goal failure resulting in the programme being moved to a PBC. Participant 3 achieved performance standards (68 CPM) on the 14th intervention session. Participant 4 was then entered and responded to the intervention, gaining 16 letter sounds CPM by the 3rd intervention session, and reaching performance standards (68 CPM) on the 9th intervention session. The intervention was then introduced to P5 - who was considered responsive - gaining 30 letter sounds CPM by the 3rd intervention session, and reaching near performance standards (64 CPM) on the 14th intervention session.

Figure 7 shows that P6 also responded to the intervention for the first two weeks, gaining 16 letter sounds CPM by the 3rd intervention session. The programme was then modified to include flashcards of previous errors and subsequently a PBC was introduced. Performance stabilised at 50-60 CPM and was terminated at the reduced criterion of 52 CPM on the 20th intervention session. Finally, P7 was entered and also responded to the intervention, gaining 12 letter sounds CPM by the 3rd intervention session, and reaching performance standards (68 CPM) on the 12th intervention session (see Figure 7). These results show that the intervention programme was effective in building frequencies in letter sounds to performance standards or reduced criteria for all participants. This suggests that the PT intervention programme was an effective Tier 2 intervention for targeting fluency in letter sounds.

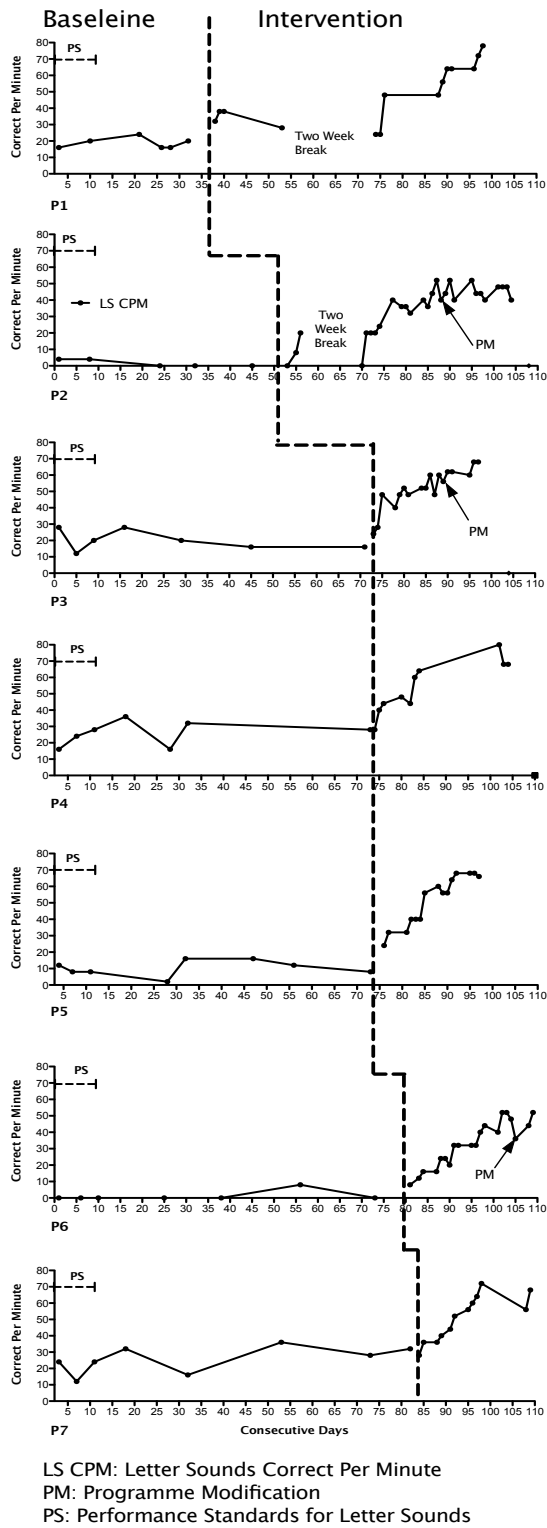


Figure 7. Experiment 1: multiple-baseline across participants design for the PT intervention in letter sounds.

2.1.2.1. Pre- Post-Test Outcomes in Letter Sounds

Table 1 presents participants' post-test standardized score, and standard score ratio gains in letter sounds on the YARC letter sound knowledge subtest (LSK-SSG; LSK-SSRG), and pre- to post-test status changes (LSK-Pre; LSK-Post). Table 1 shows the average SSG for letter sounds was 14.6 (range 6-25), and the average SSRG for letter sounds was 7.3 (range 6.3-9.2). Of the seven participants selected at pre-test, five were identified as having a severe difficulty (P1, P2, P4, P5, and P6), and two with a moderate difficulty (P3 and P7). At post-test all seven participants demonstrated average performance on the YARC LSK subtest (see Table 1). This indicates that the intervention programme resulted in at risk readers closing the gap with average peers in letter sound knowledge.

This observed effect was achieved in an average of 2.4 hours of intervention time (range 1.3-3.9 hours). An average of 18 intervention sessions were required to reach performance standards or reduced criterion in letter sounds, and daily intervention sessions lasted on average nine minutes.

Table 1

Pre-Post-Test Outcomes in YARC Letter Sound Knowledge (LSK) Subtest for Participants that Received the PT Intervention in Single Letter Sound (Experiment 1).

	LSK-SSG	LSK-SSRG	Total Hours Intervention	LSK Pre-status	LSK Post-status
P1	12	6.7	1.8	Severe difficulty	Average
P2	25	6.4	3.9	Severe difficulty	Average
P3	12	6.3	1.9	Below average	Average
P4	6	8.2	.7	Severe difficulty	Average
P5	19	8.6	2.2	Severe difficulty	Average
P6	16	7.9	2.9	Severe difficulty	Average
P7	12	9.2	1.3	Below average	Average

2.1.2.2. Long-Term Intervention Effects: Maintenance of Letter Sounds

Figure 8 shows that the average performance in letter sounds at median baseline was 14 CPM (range 0-28), and at intervention end this had increased to 62 CPM (range 40-76); a 4.4-fold increase in correct responding. Due to school absenteeism over the five scheduled maintenance checks, participants show an average of four checks (range 3-5). The maintenance checks carried out at Week 1, Week 2, and 4-6 weeks, and 8-10 weeks post- intervention demonstrate an average performance of 61 CPM (range of 44-84). This suggests that the performance

standards, or reduced criterion, achieved by participants in the intervention were largely maintained until the end of the school year (8-10 weeks post- intervention).

Figure 8 shows that participants 1, 3, 4 and 5 achieved performance standards and maintained these performance levels up to 10 weeks later. Participant 7, however, achieved the performance standard but did not maintain this level across time, showing a gradual reduction across maintenance checks. Participants 2 and 6 demonstrated at or near zero correct responding in baseline, and intervention for these participants was terminated at a reduced criterion due to time constraints (P2 40 CPM; P6 52 CPM); these also were the performance ranges achieved in the intervention. Similarly, participants' incorrect responding observed at intervention end remained stable up to 10 weeks later (see Figure 8), suggesting that the intervention was effective in reducing errors to low levels and that errors remained low across time. Taken together, these patterns suggest that the frequency levels of correct and incorrect responding achieved in the intervention may be expected in maintenance checks two months post- intervention.

Furthermore, Figure 8 shows that at the delayed maintenance check (34-38 weeks post-intervention), average performance decreased slightly to 57 CPM (range 40-72); a .9-fold decrease in correct responding achieved in the intervention. This is an important finding corroborating research that has previously reported low maintenance of academic skills (Downey, von Hippel, & Beckett, 2004) particularly evident with disadvantaged students (Alexander, Entwisle, & Olson, 2007). The current findings show high maintenance of letter sounds up to nine months post-intervention.

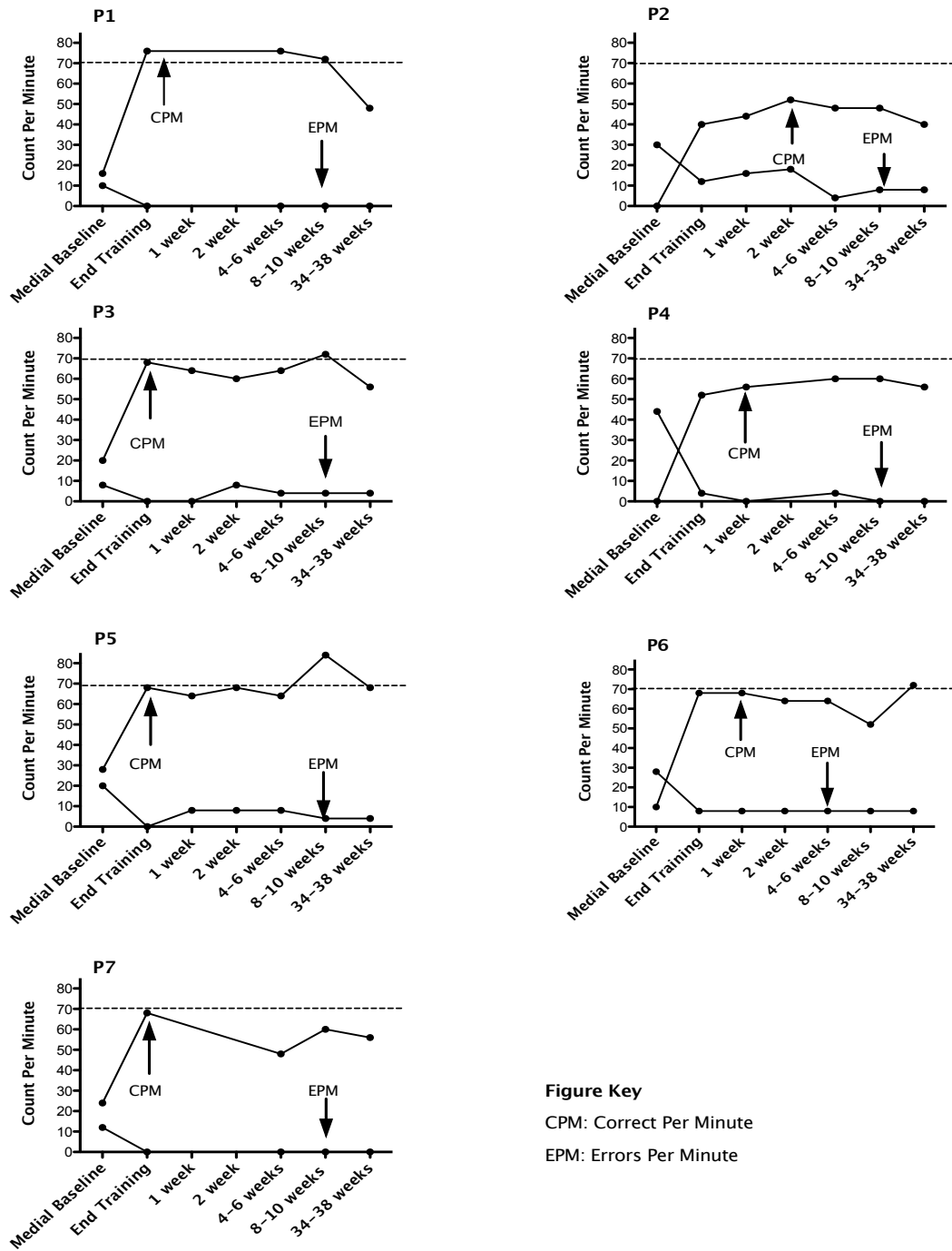


Figure Key
 CPM: Correct Per Minute
 EPM: Errors Per Minute

Figure 8. Experiment 1, maintenance data displaying performance in letter sounds at median baseline, intervention end and five post-intervention checks: 1 week; 2 weeks; 4-6 weeks; 8-10 weeks, and 34-38 weeks.

2.1.3. Experiment 1: Conclusions

The purpose of Experiment 1 was to implement a PT intervention programme as a Tier 2 intervention with at risk Senior Infant readers, to investigate its effectiveness in targeting letter sound fluency, and to evaluate maintenance of the skill over prolonged periods of time. The goal of the intervention was to build frequencies in letter sounds to performance standards, and for participants to close the gap with their average performing peers

The current research demonstrated that the intervention programme was an effective Tier 2 intervention for targeting fluency in letter sounds with young children considered to be at risk readers. The intervention procedures were effective in building frequencies in letter sounds to performance standards, or reduced criteria for all participants. Importantly, pre- post-test outcomes also indicated that the intervention programme was successful in moving at risk readers into the average range of performance in letter sounds.

Frequency building to performance standards was effective in increasing the frequency and accuracy of letter sounds for all participants. Five participants achieved at, or near performance standards (70-90 CPM) and intervention was terminated at a reduced criterion for the remaining two participants (40-52 CPM; P2 and P4). During baseline, both P2 and P4 consistently demonstrated near zero accuracy in letter sounds. However, following approximately 22 intervention sessions, performance had stabilised at 40-50 letter sounds CPM. These were considered substantial gains, as prior to the intervention both P2 and P4 had not acquired any letter sounds in 1.5 years of regular classroom instruction. In any intervention students may begin with differing levels of accuracy, this experiment shows that frequency building to performance standards can be targeted despite these differences. This is an important

consideration regarding the fit of an intervention programme to a classroom of students with diverse instructional needs.

Maintenance data shows that the performance levels achieved at intervention end are largely maintained up to 38 weeks later (with the exception of P7). This suggests that the performance standards achieved through the current PT intervention programme persist across significant periods of time.

Due to time constraints training was terminated at reduced criterion for two participants (P2 and P4), and maintenance data continued to reflect the performance levels achieved in the PT intervention. Both of these participants demonstrated at or near zero correct responding in baseline – as such their progress in the intervention was considered satisfactory, however, additional intervention time or other programme modifications may have yielded superior results.

2. 2. Experiment 2: Letter Names

The purpose of Experiment 2 was to implement a PT intervention programme as a Tier 2 intervention with Senior Infant readers identified as at risk. Through an evaluation of maintenance data, Experiment 2 intended to investigate the effectiveness of the PT intervention in establishing letter name fluency. This Experiment also aimed to evaluate the effect of the PT intervention programme in helping participants close the gap with average performing peers in the target skill of letter names. The intervention goal was achievement of performance standards in letter names (70-100 CPM) at the learning rate of x2 celeration.

2.2.1. Method

2.2.1.1. Participants and Setting

Participants and setting are identical to that reported in Experiment 1.

2.2.1.2. Universal Screening and Pre-to-Post Test Measures

The screening instrument included the Dynamic Indicators of Basic Early Literacy Letter Name Fluency subtest (DIBELS, LNF; Good & Kaminski, 2002). The DIBELS LNF subtest is a standardized criterion referenced measure that consists of upper and lower case letters presented in random order; the participant is required to name as many letters as he/she can in a one-minute interval. The reported reliability for this subtest is .86 (Cronbach's alpha), and when administered in January its predictive validity coefficient with the Woodcock Johnson Readiness Cluster standard score in May is .57 (Good & Kaminski, 2002).

2.2.1.3. Materials.

Stimuli and other materials were identical to that used in Experiment 1. In addition, "sand paper letters" and flashcards were used for discrete trials. The sandpaper letters (16 X 13cm) consisted of the letter depicted in sandpaper, accompanied by 3-4 words beginning with that letter, and illustrations to match the words (Smart Kids, 2006). The single letter laminated flashcards (9X8 cm) were handmade by the experimenter.

2.2.1.4. Design.

A multiple-baseline across participants design was employed to assess the effectiveness of a PT intervention programme as an intervention building frequencies in letter names to performance standards.

2.2.1.5. Procedure.

Baseline. Baseline procedures are identical to that described in Experiment 1.

Intervention. The independent variable involved a Tier 2 PT intervention programme that included PT as an intervention to target fluency in letter names, and a decision-making framework to monitor progress. Intervention procedures were identical to that reported in Experiment 1, with the exception of discrete trials that preceded the fluency trials. In the case of letter names, participants who were moved to a “personal best component” performance range were provided with discrete trials using sandpaper letters. Participants were presented with flash cards that corresponded with repeated errors. A number of timings were conducted with the smaller array of letters. If repeated errors continued, discrete trials consisted of using the sandpaper letter to produce the name and provide examples of words that begin with that name, for example “b is for ball, and balloon”.

2.2.1.6. Interobserver agreement (IOA). Exact Count-Per-Interval IOA data were collected for 50% of all baseline sessions demonstrating 100% agreement, and 15.5% of all intervention sessions achieving 100% agreement. Overall that is 100% agreement across 32% baseline and intervention sessions.

2.2.1.7. Procedural integrity (PI). The procedural integrity checklists were identical to those used in Experiment 1 (see Appendix C for baseline checklist, and Appendix D for intervention checklist); PI data were calculated in the same manner as Experiment 1. PI data were collected for 29% of baseline sessions demonstrating 98% adherence, and for 21% of intervention sessions demonstrating 99% adherence.

2.2.1.8. Dependent Variables and Data Analysis

Dependent variables and data analysis procedures employed were identical to those measured in Experiment 1.

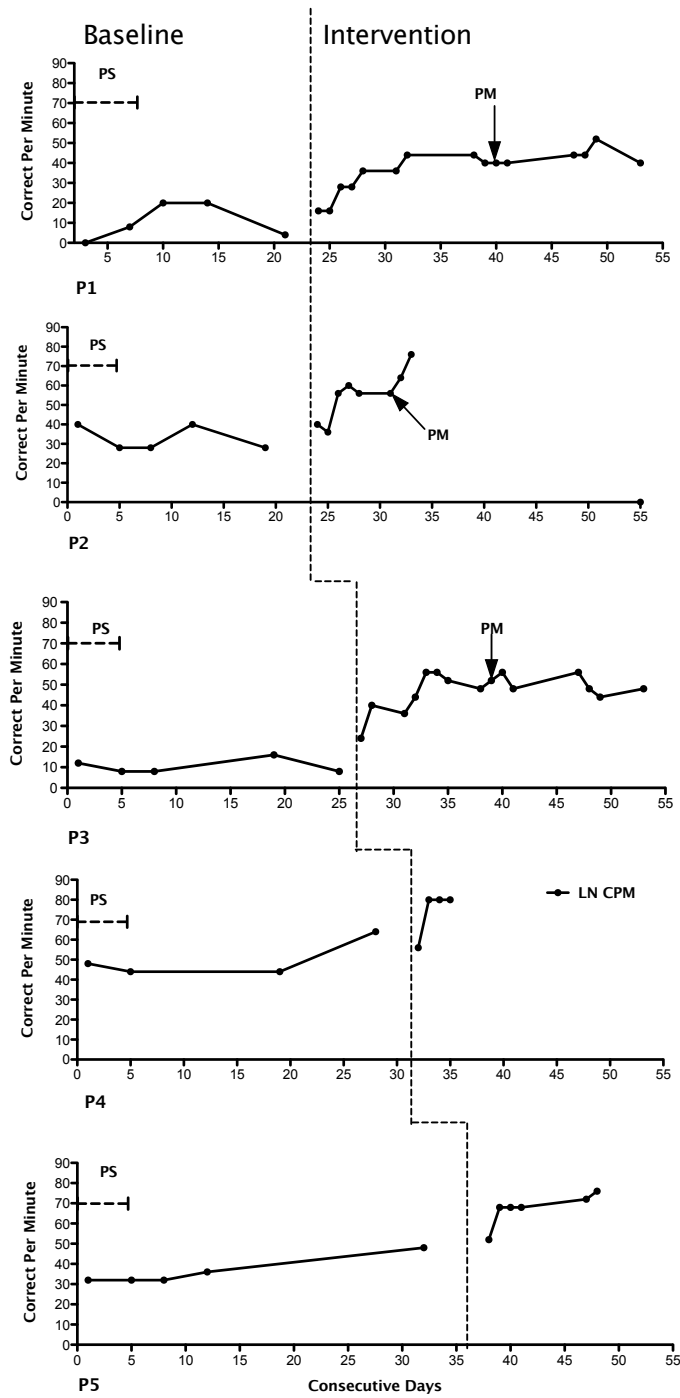
2.2.2. Results and Discussion

Figure 9 depicts a multiple-baseline across participants. All participants entered the baseline condition at the same time, and a minimum of four baseline probes were collected before P1 and P2 were entered into the intervention phase. Subsequent participants were added in a sequential fashion - when intervention participants demonstrated *response* to the intervention for three consecutive days (i.e. maintaining a x2 celeration).

Figure 9 shows that P1 and P2 responded to the intervention, respectively gaining 16 and 28 letter names CPM by the 3rd intervention session. Both participants maintained x2 celeration for the first two weeks of intervention followed by both experiencing goal failure. The programme for Participant 2 was modified (modelling speed) and she subsequently reached performance standards (76 CPM) on the ninth intervention session. The programme for Participant 1 was modified (modelling speed; flashcards) and P1 was moved to a personal best component. The PT intervention was then terminated at reduced criterion levels for P1 (48 CPM).

Participant 3 was then entered into the intervention phase; he responded to the intervention and gained 28 letter names CPM by the 3rd intervention session (see Figure 9). Participant 3 maintained x2 celeration for the first week of training and then experienced goal failure followed by programme modification (modelling speed; flashcards), and P3 moved to a personal best component. The PT intervention was then terminated at reduced criterion levels (48 CPM). Participant 4 was then entered into the intervention, he responded to the intervention gaining 34 letter names CPM,

and reached performance standards (80 CPM) by the 3rd intervention session (see Figure 9). Finally, P5 was entered and also responded to the intervention, gaining 40 letter sounds CPM and achieving performance standards by the 5th intervention session (76 CPM).



LN CPM: Letter Names Correct Per Minute
 PM: Programme Modification
 PS: Performance Standards Letter Names

Figure 9. Experiment 2: multiple-baseline across participants design for the PT intervention in letter names.

2.2.2.1. Pre- Post-Test Outcomes in Letter Names

Table 2 quantifies the participants' post-test changes in correct per minute scores for the DIBELS LNF subtest (LNF Pre-CPM/LNF Post-CPM), status changes (LNF pre-status/LNF post-status), and gains in correct per minute scores of LNF (LNF – gains CPM). Table 2 shows the average gain demonstrated in letter names measured by DIBELS subtest LNF was 25 correct per minute (range 7-34). Of the five participants selected at pre-test for intervention, three were identified as being at risk (P1, P3, and P5) and two were identified as showing some risk (P2 and P4). At post-test four participants were identified as low risk (P1, P2, P4, and P5), and one was identified as at risk (P3).

Table 2 shows that the average participant achieved these outcomes in approximately 2.3 hours of intervention time (range .4-3 hours). It took on average ten intervention sessions to reach performance standards or reduced criterion in letter names, and daily intervention sessions lasted on average nine minutes.

Table 2

Pre-Post-Test Outcomes in DIBELS Letter Naming Fluency (LNF) Subtest for Participants that Received the PT Intervention in Letter Names (Experiment 2).

	LNF Pre- CPM	LNF Post CPM	Total Hours Intervention	LNF Pre- status	LNF Post- status	LNF – Gains CPM
P1	10	30	2	At Risk	Low Risk	30
P2	20	44	1.4	Some Risk	Low Risk	24
P3	15	22	3	At Risk	At Risk	7
P4	29	61	.4	Some Risk	Low Risk	32
P5	14	48	.8	At Risk	Low Risk	34

2.2.2.2. Long-Term Intervention Effects: Maintenance of Letter Names

Figure 10 shows participants' performance in correct and incorrect responding in letter names across time. Each panel represents a participant's performance at median baseline, intervention end, one week post- intervention, two weeks post-intervention, four-to-six weeks post-intervention, eight-to-ten weeks post-intervention, and 34-38 weeks post-intervention. Due to school absenteeism over the five scheduled maintenance checks, participants show an average of four checks (range 4-5). Figure 10 shows that the average performance in letter names at median baseline was 25 CPM (range 8-46), and at intervention end this had increased to 66 CPM (range 48-80); a 2.6-fold increase in correct responding. The maintenance checks carried out at Week 1, Week 2, and 4-6 weeks, and 8-10 weeks post-intervention demonstrate an average performance of 63 CPM (range of 55-78).

However, some participants' performance patterns across maintenance checks showed fluctuations in frequencies up to the end of the school year (8-10 weeks post-intervention). Figure 10 shows that Participants 1 and 3 did not reach performance standards, and letter name intervention was terminated at the reduced criterion of 48 CPM. Both participants demonstrated variable correct responding and higher errors compared to the remaining participants. In contrast, P2 reached performance standards in letter names, but did not maintain this level across time, showing a gradual reduction across maintenance checks. Participants 4 and 5 demonstrated the highest baseline frequencies in letter sounds, and maintained performance standards across prolonged periods of time. In contrast to fluctuations observed in correct responding, participants' incorrect responding observed at intervention end remained stable up to 10 weeks later (see Figure 10). This suggests that the intervention was effective in reducing errors to low levels and that errors remained low across time.

Figure 10 shows that at the delayed maintenance check (34-38 weeks post intervention), average performance was 77 CPM (range 56-100); a 1.2-fold increase in correct responding to that achieved in the intervention. The increase in correct responding observed nine months later may reflect a shift in classroom instruction from letter sounds to letter names. Letter sounds – and not names - are taught in Senior Infant classrooms. At the delayed maintenance check, participants had moved into the next grade (1st class), and had received instruction in letter names. In this grade, spelling (letter names) is a main focus overtaking phonics instruction (letter sounds). This may explain increases in correct responding for letter names at the delayed maintenance check.

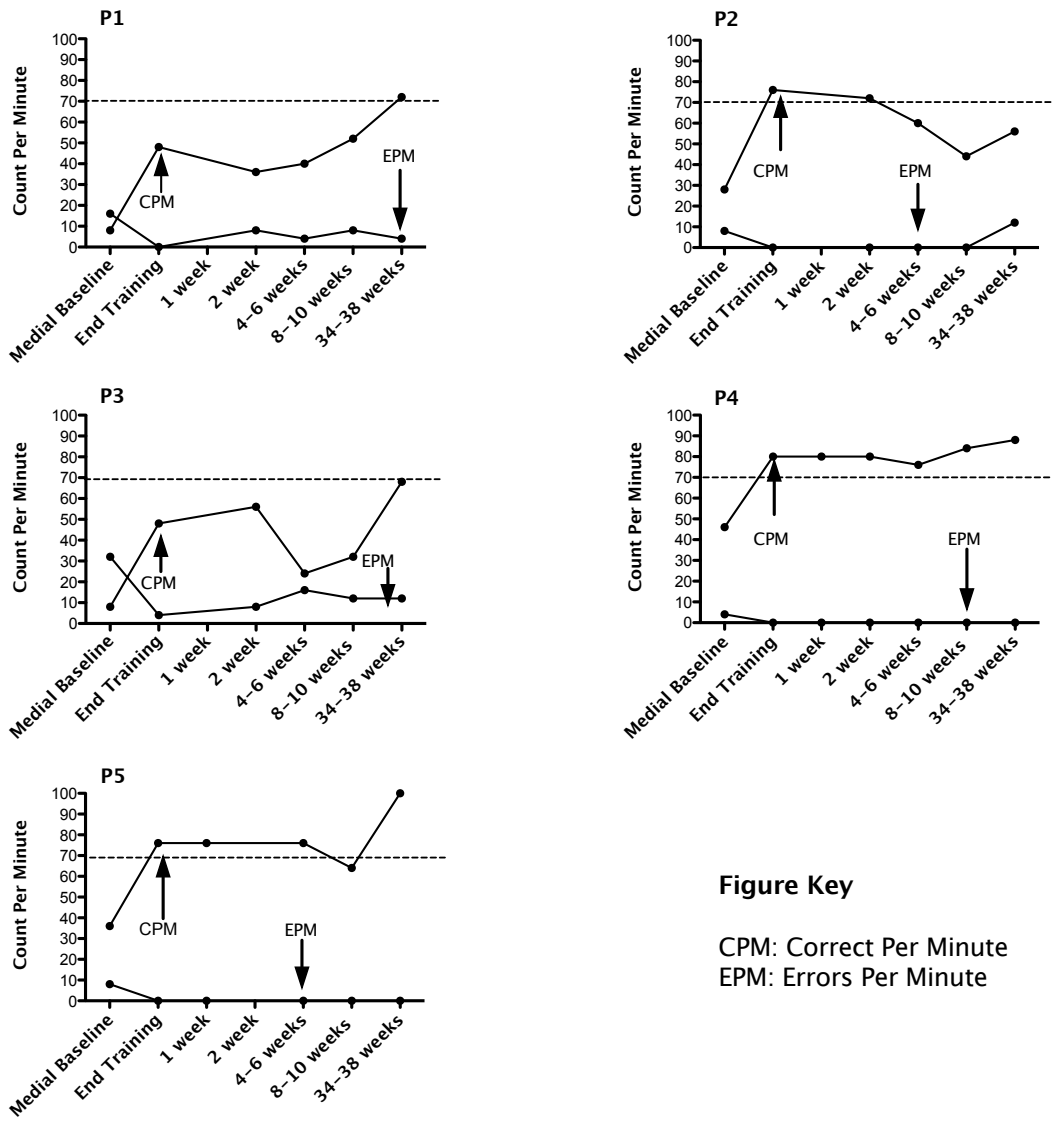


Figure 10. Experiment 2, maintenance data displaying performance in letter names at median baseline, intervention end and 5 post-intervention checks: 1 week; 2 weeks; 4-6 weeks; 8-10 weeks, and 34-38 weeks.

2.2.3. Experiment 2: Conclusions

The purpose of Experiment 2 was to implement a PT intervention programme as a Tier 2 intervention with at risk Senior Infant readers, to investigate its effectiveness in targeting letter name fluency, and to evaluate maintenance of the skill over prolonged periods of time. The goal of the intervention was to build frequencies in letter names to performance standards, and for participants to close the gap with their average performing peers.

Experiment 2 evidenced mixed results across participants. For most participants, the PT intervention programme was an effective Tier 2 intervention for targeting fluency in letter names. The intervention procedures were effective in building frequencies in letter names to performance standards, or reduced criteria for all participants. The PT intervention was terminated at the reduced criterion of 48 CPM for P1 and P3 as performance had maintained at this level for a number of days. As letter names were not targeted in core classroom instruction, the continuation of the PT intervention to performance standards was not pursued with these participants. In addition, pre-intervention levels of correct responding were 8 letter names CPM; this had increased to 48 CPM by intervention end - a 6-fold increase in correct responding – and was considered a substantial gain for both participants. Pre- post-test outcomes also indicated that the intervention programme was successful in moving at risk readers into the low risk category for letter names.

The add/subtract graphs used for the purposes of presenting the MBDs create the impression of ascending baselines for Participants 4 and 5. Such an artifact warrants discussion of two issues. First, The increasing trend observed in the baseline condition indicates that practice alone led to increases in rate. Binder (1996) notes the limited practice opportunities for newly acquired skills in education systems. The

trend in baseline attests to the power of practice in increasing rates even in the absence of feedback. The increase in trend observed in the current analysis, however, was not clinically significant on the level of the individual, as these trends were quantifiable on the SCC. Use of the SCC in the baseline condition ensured that even where a small increase was observed, this could be quantitatively compared to the effect of the intervention when it was introduced. For example, Participant 4 demonstrated a X1.2 celeration in the baseline condition, compared to a X5.1 celeration in the first four days of the intervention. Similarly, Participant 5 demonstrated a X1.5 celeration in the baseline condition, compared to a X3.4 celeration in the first four days of the intervention.

The PT intervention was terminated for two participants (P1 and P3) at reduced criteria. Despite programme modifications these participants performance stabilised at approximately 50 CPM and this performance level was maintained (P1) or decreased (P3) at the end of the school year. At the delayed post-test both participants demonstrated a large increase, as did P5. This measurement was taken in the next grade where letter names are taught in classroom instruction.

2.3. Experiment 3: Sound Isolation

The purpose of Experiment 3 was to detect a sample of Senior Infant students in difficulty with the phonemic awareness skill of sound isolation, and to apply the PT intervention programme described in Experiments 1 and 2, with these participants. The intervention aimed to build frequencies in sound isolation to performance standards, and for participants to approximate average performance on standardised assessments. Generalisation and long-term intervention effects were also investigated.

The intervention goal was achievement of performance standards in sound isolation (20-24; Freeman & Haughton, 1997) at a learning rate of x2 celeration.

2.3.1. Method

2.3.1.1. Participants and Setting

Participants and setting were identical to that in described in Experiment 1.

2.3.1.2. Universal Screening

All 12 participants were screened using the Sound Isolation subtest from the York Assessment of Early Reading Comprehension (YARC; Hulme, et al., 2009) to identify a pool of individuals in need of additional instruction. The sound isolation subtest assesses one component of phoneme awareness – phoneme isolation. It was designed to measure phoneme isolation skills and to identify students who are experiencing difficulties with phoneme awareness skills. It consists of 12 test items and six teaching items. Six of the test items assess isolation of the initial phoneme, and six of the final phoneme. The reported reliability for the test is .88 (Cronbach's alpha) and its predictive validity coefficient with the Single Word Reading Test (Foster, 2007) is .62 (Hulme et al., 2009). The universal screener also served as a pre-test to compare to post-test outcomes in investigating the effect of the intervention.

A student is considered to have a severe difficulty in phonemic awareness skill of sound isolation if he or she achieves a standard score between 70-79, and to have a moderate difficulty if he or she achieves a standard score between 80-84.

Five of the 12 participants were identified as having a severe difficulty (P2, P4, P5, $n=3$, age range 5.7-6.8, $M = 6.3$ years) or a moderate difficulty (P1, P3, $n=2$, age 5.8 years) in the phonemic awareness skill of sound isolation. These participants were selected for intervention in sound isolation.

2.3.1.3. Materials

The stimulus materials were the sound isolation section of the “Phonological Coding: Word and Syllable/Phonemic Awareness” curriculum (Freeman & Haughton, 1997). These materials were selected for the intervention as they reflected the content domain of the YARC subtest selected as outcome measures (sound isolation) and they provided multiple exemplars of the phonemic skills targeted in the intervention. This curriculum is used in precision teaching centres of excellence in the United States (e.g., Haughton Learning Center; Morningside Academy).

Other materials were identical to that in Experiment 1.

2.3.1.4. Design

A multiple-baseline across participants with a multiple probe element across target skills design was employed to assess the effectiveness of a PT programme as an intervention building frequencies in sound isolation of initial and final phoneme.

2.3.1.5. Procedure

Baseline. Baseline probes demonstrated current rates of responding in the absence of the PT intervention to compare to rates of responding following intervention. Stimulus sheets from the “Phonological Coding” curriculum (Freeman & Haughton, 1997) were also used during baseline sessions. Each target skill was probed 1-2 times per week, and rates of responding recorded on a daily per minute standard celeration chart. During this phase rates of responding were obtained for each target skill in timing lengths of 30 seconds, as they required instructor presentation of stimuli. Where a skill had multiple levels of difficulty probes were obtained for each level of the skill. For example, for the skill sound isolation there are three levels of difficulty: initial phoneme → final phoneme → medial phoneme (Puffpaf, 2009).

Baseline procedures were identical to that described in Experiment 1.

Intervention. The independent variable involved a Tier 2 PT intervention programme that included PT as an intervention to target fluency in sound isolation, and a decision-making framework to monitor progress. Participants received the intervention on a one-to-one basis in a resource room near their regular classroom. Each intervention session was timed to calculate approximate time of overall intervention delivered. The intervention was delivered on a daily basis during school time, five days per week. Each intervention session consisted of daily fluency trials in the form of three-to-five 30-second timings sound isolation (discrimination and production of initial and final phonemes).

Intervention procedures were identical to that in Experiment 1.

2.3.1.6. Interobserver agreement (IOA). Exact Count-Per-Interval IOA data were collected for 23% of all baseline sessions, and 30% of all intervention sessions, achieving 100% exact count per interval agreement. Overall 100% agreement across 26.5% baseline and intervention sessions was reported.

2.3.1.7. Procedural Integrity (PI). The procedural integrity checklists were identical to those used in Experiment 1 (see Appendix C for baseline checklist, and Appendix D for intervention checklist); PI data were calculated in the same manner as Experiment 1. PI data were collected for 22.4% of baseline sessions demonstrating 98.8% adherence, and for 15.4% of intervention sessions demonstrating 95.9% adherence.

2.3.1.8. Dependent Variables and Data Analysis

Dependent variables were identical to those measured in Experiment 1. Data analysis was identical to that reported in Experiment 1, with the exception that gains

achieved by the 3rd intervention session are not reported in the current experiment.

2.3.2. Results and Discussion

Figure 11 depicts a multiple-baseline across participants ($n = 5$) with multiple probes across target skills (initial and final phonemes) design. All participants entered the baseline condition at the same time, and a minimum of three baseline probes were collected before P1, P2, and P3 were entered into the intervention phase for initial phoneme intervention (i.e. discrimination and production of the first phoneme in words presented orally to the participant). Subsequent participants were added in a sequential fashion - when intervention participants demonstrated *response* to the intervention for three consecutive days (i.e., maintaining a x2 celeration). Participants 1, 2, and 3 responded to the intervention, reaching performance standards (24-26 CPM) on the 5th, 5th, and 8th intervention sessions respectively. Participant 4 was then entered, and responded to the intervention, reaching performance standards (24 CPM) on the 10th session (see Figure 11). Finally, Participant 5 was entered and did not respond to the intervention; therefore the programme was modified to include familiar stimuli and visual cues. This modification was successful and P1 responded to the intervention, achieving performance standards (24 CPM) on the 9th intervention session.

When the PT intervention for the initial phoneme was terminated, a minimum of two true baseline (TBL) probes of final phoneme were taken to establish performance levels in context of the intervention in initial phoneme received. These TBL were used to set performance goals for the subsequent intervention in final phoneme i.e., the median of the TBL was used to set a x2 celeration goal for the subsequent week of intervention. Figure 11 shows that participants 1 and 2 were entered into the intervention for sound isolation of final phoneme. Participant 2

responded to the intervention. Participant 1 failed to respond for two sessions followed by a performance increase before reaching criterion levels of frequency (24 CPM, 24 CPM, respectively). Participants 3, 4, and 5 all responded to the intervention, reaching performance standards (20-26 CPM) on the 4th, 3rd, and 9th intervention session respectively.

2.3.2.1. Sound Isolation – Initial and Final Phoneme Interaction

During initial phoneme baseline, intervention, and post- intervention sessions, multiple probes of final phoneme were also measured. The purpose of this was to investigate the effects of the PT intervention targeting the initial phoneme, and its potential generalisation to the discrimination and production of final phonemes. A generalisation effect of the PT intervention targeting the initial phoneme on the final phoneme was observed for most participants, but to varying magnitudes. Figure 11 depicts the performances of P1 and P2 showing a generalisation effect to the final phoneme within three sessions of targeting the initial phoneme. Participant 3 demonstrated limited generalisation from the initial to final phoneme, and it remained static across time. Participant 4 demonstrated delayed generalisation to the final phoneme, and P5 demonstrated a generalisation effect with the final phoneme, but most clearly when he reached performance standards in initial phoneme (see Figure 11).

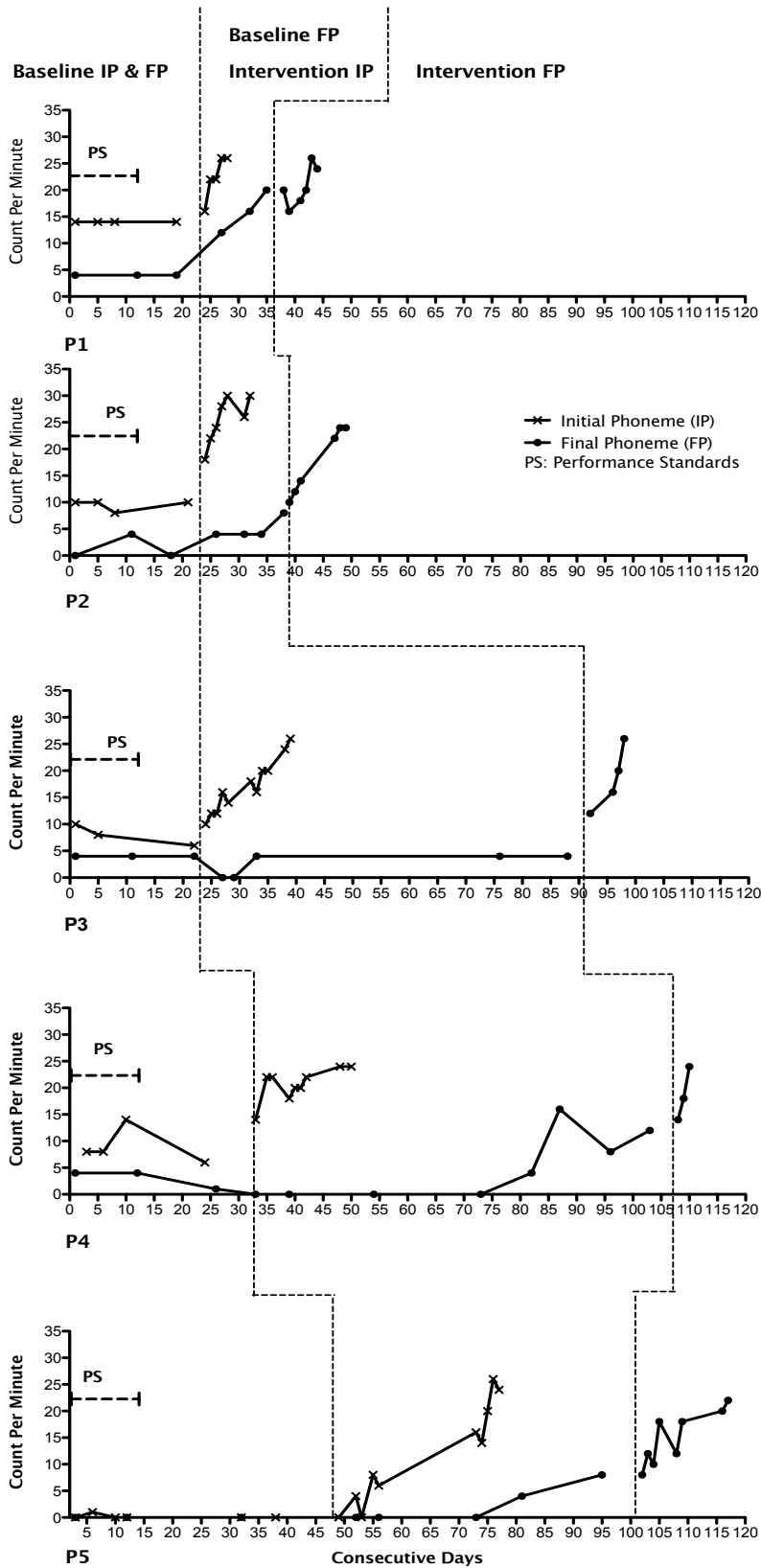


Figure 11. Experiment 3: multiple-baseline across participants with probes across skills design for the PT intervention in sound isolation.

2.3.2.2. Pre- Post-Test Outcomes in Sound Isolation

Table 3 illustrates participants' post-test gains in: standard scores for the sound isolation subtest (SI-SSG) and resulting standardised score ratio gains (SI-SSRG) status changes (SI pre-status/SI post-status). Table 3 shows the average gain in YARC standard scores for sound isolation was 28.8 (range 23-34), with an average SSRG per hour of intervention of 15.6 (range 12.8-17). Of the five participants at pre-test, three were identified as having a severe difficulty (P1, P2, P4) and two with a moderate difficulty (P1, P3). At post-test all five participants demonstrated performances within the average range (see Table 3).

This observed effect was achieved in an average of 1.9 hours of intervention time (range 1.7-2 hours). An average of 14 intervention sessions were required to reach performance standards in sound isolation of both initial and final phonemes, and daily intervention sessions lasted on average eight minutes.

Table 3

Pre-Post-Test Outcomes in YARC Sound Isolation (SI) Subtest for Participants that Received the PT Intervention in Sound Isolation (Experiment 3).

	SI-SSG	SI-SSRG	Total hours Intervention	SI Pre-Status	SI Post-Status
P1	23	13.5	2	Below Average	Average
P2	34	17	2	Severe Difficulty	Average
P3	23	12.8	1.7	Below Average	Average
P4	23	18.8	1.8	Severe Difficulty	Average
P5	32	16	1.7	Severe Difficulty	Average

2.3.2.3. Long-Term Intervention Effects: Maintenance of Sound Isolation

Figure 12 shows participants' performance in correct and incorrect responding in sound isolation of initial and final phonemes across time. Each panel represents a participant's performance at median baseline, intervention end, one week post-intervention, two weeks post-intervention, four-to-six weeks post-intervention, eight-to-ten weeks post-intervention, and 34-38 weeks post-intervention. Due to school absenteeism over the five scheduled maintenance checks, participants show an average of four checks (range 3-5). Figure 12 shows that the average performance in initial phoneme at median baseline was 8 CPM (range 0-14), and at intervention this had increased to 25 CPM (range 24-26), representing a 3-fold increase in correct responding. The maintenance checks carried out at Week 1, Week 2, 4-6 weeks, and 8-10 weeks post-intervention demonstrate an average performance of 28 CPM (range of 27-30).

Figure 12 also shows that the average performance in final phoneme at median baseline was 8 CPM (range 0-14), and at intervention end this had increased to 24 CPM (range 20-26); a 3-fold increase in correct responding. The maintenance checks carried out at Week 1, Week 2, 4-6 weeks, and 8-10 weeks post-intervention demonstrate an average performance of 24 CPM (range of 22-27). This suggests that the performance standards, or reduced criterion, achieved by participants in the intervention were largely maintained until the end of the school year (8-10 weeks post-intervention).

Similarly, participants' incorrect responding observed at intervention end remained stable up to 10 weeks later (see Figure 12), suggesting that the intervention was effective in reducing errors to low levels and that errors remained low across time. Taken together, these patterns suggest that the frequency levels of correct and

incorrect responding achieved in the intervention can be expected in maintenance checks two months post-intervention. Performance at the delayed maintenance check showed a 1.1-fold decrease in correct responding, and at nine months post-intervention, remained at performance standards for initial and final phoneme.

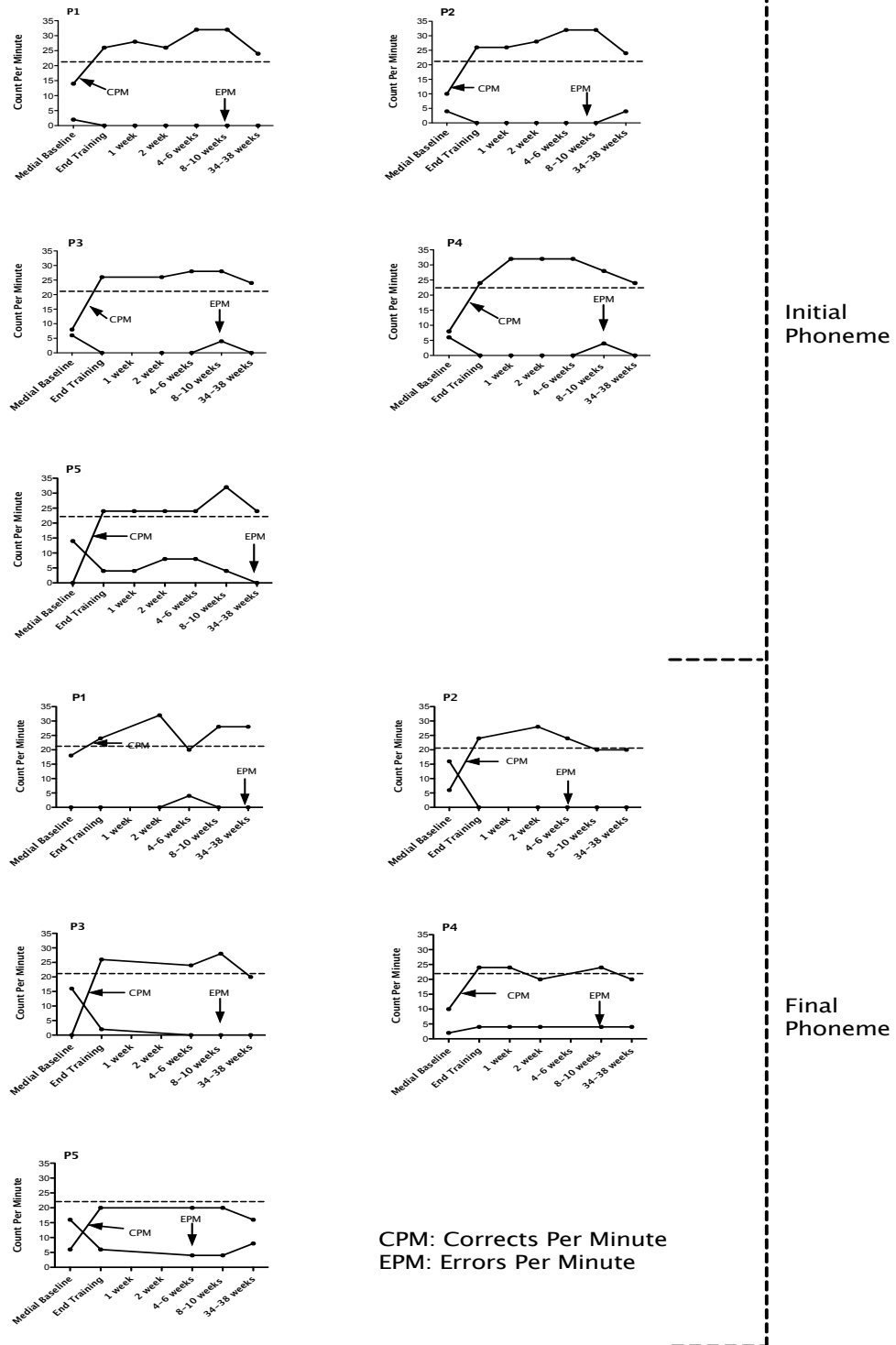


Figure 12. Experiment 3, maintenance data displaying performance in sound isolation at median baseline, intervention end and five post-intervention checks: 1 week; 2 weeks; 4-6 weeks; 8-10 weeks, and 34-38 weeks.

2.3.3. Experiment 3: Conclusions

The purpose of Experiment 3 was to identify a group of Senior Infant students in need of additional instruction in the phonemic awareness skill of sound isolation, and to implement the PT intervention programme with these participants. The goal of the intervention was to build frequencies in sound isolation to performance standards, and for participants to close the gap with their average performing peers. Generalisation and long-term intervention effects were also investigated.

A group of participants in need of additional instruction in sound isolation was identified using the sound isolation subtest from the YARC assessment, and the PT intervention programme was subsequently implemented with these students. The intervention was successful in building frequencies in initial and final phonemes to performance standards for the majority of participants. At post-test, all participants were in the average performance range, indicating that these participants had closed the gap with average performing peers in the phonemic awareness skill of sound isolation, and long-term effects suggest that the skills were maintained up to nine months post- intervention.

Experiment 3 demonstrated that the PT intervention may be used to effectively target fluency in PS skills of sound isolation with emergent readers, regardless of initial performance in these skills. All participants in the current study scored in the below average and severe difficulty range at pre-test in sound isolation and all scored in the average range at post-test. Experiment 3 demonstrated that phoneme isolation (initial and final) was taught to performance standards with most participants, suggesting that despite significant deficits in this skill it is possible to teach phoneme manipulation skills to emergent readers.

The current research also examined whether targeting one skill would generalise to specific related skills. Experiment 3 showed clear generalisation effects for two participants and a delayed effect for the remaining two participants. This suggests that building frequencies in initial phoneme isolation may foster final phoneme isolation in the absence of intervention. Importantly, intervention effects were still evident up to nine months post-intervention.

2.4. Experiment 4: Sound Deletion

The purpose of Experiment 4 was to identify a sample of Senior Infant students in difficulty with the phonemic awareness skill of sound deletion, and to apply the PT intervention programme with these participants. The intervention aimed to build frequencies in sound deletion to performance standards (15-25 CPM; Freeman & Haughton, 1997), and for participants to approximate an average performance. Generalisation and long-term intervention effects were also investigated.

2.4.1. Method

2.4.1.1. Participants and Setting

Participants and setting are identical to that described in Experiment 1.

2.4.1.2. Universal Screening Phase

All 12 participants were screened using the Sound Deletion (SD) subtest from the York Assessment of Early Reading Comprehension (YARC; Hulme, et al., 2009) to identify a pool of individuals in need of additional instruction. The sound deletion subtest assesses one component of phoneme awareness – sound deletion. It was designed to measure phoneme deletion skills and to identify students who are experiencing difficulties with phoneme awareness skills. The test consists of 12 test

items and seven teaching items. Two of the test items assess deletion of a word from a compound word; three assess deletion of the final phoneme; four of the initial phoneme; and, three of the medial phoneme. The reported reliability for the test is .93 (Cronbach's alpha) and its predictive validity coefficient with the Single Word Reading Test (Foster, 2007) is .76 (Hulme et al., 2009). The screener also served as a pre-test to compare to post-test in investigating the effect of the intervention.

A student is considered to have a severe difficulty in phonemic awareness skill of sound isolation if he or she achieves a standard score between 70-79, and to have a moderate difficulty if he or she achieves a standard score between 80-84.

Four participants were identified as having a severe difficulty (P3, P2, P4, $n=3$, age range 6-6.5, $M = 6.3$) or a moderate difficulty (P1, $n=1$, age 5.8 years) in the phonological/phonemic awareness skill sound deletion. These participants were selected for intervention in sound deletion.

2.4.1.3. Materials

The stimulus materials were the Auditory Analysis (sound deletion) section of the "Phonological Coding: Word and Syllable/Phonemic Awareness" curriculum (Freeman & Haughton, 1997). Other materials were identical to those described in Experiment 1.

2.4.1.4. Design

A multiple-baseline across participants with a multiple probe element across target skills design was employed to assess the effectiveness of a PT intervention programme targeting fluency the phonemic awareness skill of sound deletion.

2.4.1.5. Procedure

Baseline. Baseline procedures were identical to that described in Experiment 3.

Intervention. The independent variable involved a Tier 2 PT intervention programme that included PT as an intervention to build frequencies in sound deletion of linguistic units, and a decision-making framework to monitor progress. Participants received the intervention on a one-to-one basis in a resource room near their regular classroom. Each intervention session was timed to calculate approximate time of overall intervention delivered. The intervention was delivered on a daily basis during school time, five days per week. Each intervention session consisted of daily fluency trials in the form of three-to-five 30-second timings sound isolation (discrimination and production of initial and final phonemes). Intervention procedures were identical to that reported in Experiment 3.

2.4.1.6. Interobserver agreement (IOA). Exact Count-per-Interval IOA data was collected for 36% of all baseline sessions achieving 98.9% agreement, and 25% of all intervention sessions, achieving 88% agreement. Overall, 93.5% agreement across 30.8% baseline and intervention sessions was reported.

2.4.1.7. Procedural integrity (PI). The procedural integrity checklists were identical to those used in Experiment 1 (see Appendix C for baseline checklist, and Appendix D for intervention checklist); PI data were calculated in the same manner as Experiment 1. PI data were collected for 14.4% of baseline sessions demonstrating 97.8% adherence, and for 21% of intervention sessions demonstrating 98% adherence.

2.4.1.8. Dependent Variables and Data Analysis

Dependent variables were identical to those measured in Experiment 3, with the exception of the accuracy measure YARC Sound Deletion subtest which assessed sound deletion in Experiment 4. Data analysis was identical to that described in Experiment 1, with the exception that gains achieved by the 3rd intervention session

were not reported in the current experiment.

2.4.2. Results and Discussion

Figure 13 represents a multiple-baseline across participants ($n = 4$) with multiple probes across four skills: compound words (CW); multisyllabic words (MS); initial phoneme (IP), and final phoneme (FP). Similar to the sound isolation paradigm in Experiment 3, while compound words was measured during baseline, intervention and post- intervention phases, multiple probes of the following related skills were implemented: multi-syllabic words, initial phoneme, and final phoneme. Probes of skills higher in the hierarchy were collected during intervention e.g., probes of final phoneme were taken while initial phoneme was targeted in intervention. This was to establish true baselines of subsequent skills in the hierarchy, and to investigate the individual effects of building frequencies in deletion of compound words, multisyllabic words, and initial phonemes and its potential generalisation to the discrimination and production of deletion of phonemes up the hierarchy. For example, would fluency in deletion of compound words generalise to multi-syllabic words or initial/final phonemes before these skills are directly targeted?

Figure 13 shows that all participants entered the baseline condition concurrently; a minimum of four baseline probes were collected before P1 and P2 were entered into the intervention for compound words. Subsequent participants were added in a sequential fashion - when intervention participants demonstrated *response* to the intervention for three consecutive days (i.e., maintaining a x2 celeration). Both participants responded to the intervention, respectively reaching performance standards (20 CPM) on the 3rd and 4th intervention session. Similarly when P3 was next entered, he responded and achieved performance standards (20 CPM) on the 3rd intervention session.

Participant 4 responded to the intervention demonstrating an immediate increase in level followed by maintenance in performance for five days (see Figure 13). Articulation difficulties appeared to incur a ceiling on his performance; a reduced criterion of 18 CPM was set, and P4 achieved this in two sessions. When an intervention effect for compound words had been demonstrated across three participants (P1, P2, P3), and minimums of two TBL taken for sound deletion of syllables from multisyllabic words (SD-MS), Participants 2 and 3 were entered into the intervention for multisyllabic words.

Figure 13 shows that both Participants 2 and 3 maintained performance at 8 and 10 responses CPM respectively for three days, followed by programme modifications (visual cue; sliceback to two phoneme words). Both participants, however, continued to experience goal failure. Stimulus materials were re-evaluated, and the impact of targeting multi-syllabic words on initial and final was considered. The PT intervention in multisyllabic words for Participants 2 and 3 was terminated due to confounding items identified in the stimulus materials. This skill was not pursued with the remaining participants. Participants 2 and 3 reached less than performance standards in multisyllabic words (15-20 CPM). After seven intervention sessions P2 produced 14 CPM; after 9 intervention sessions P3 produced 8 CPM. Minimums of two TBL probes of initial phoneme were obtained, and P1 and P2 were entered into the intervention condition for initial phonemes.

Participant 1 responded to the intervention for initial phonemes, reaching performance standards (26 CPM) on the 11th intervention session (see Figure 13). Participant 2 was next to be entered, he responded and reached performance standards on the 3rd intervention session. Participant 3 was next entered, followed by P4; both responded and reached performance standards on the 4th intervention session. The PT

intervention for the initial phoneme was then terminated, and minimums of two true baseline probes taken for the final phoneme, and participants entered in different stages (P3, P2, P1, and P4; respectively). All participants responded to the intervention (with the exception of P3 whom had already demonstrated an increase), and reached performance standards in deletion of the final phoneme.

2.4.2.1. Sound Deletion – Interaction of Compound and Multi-syllabic

Words, and Initial and Final Phonemes

During baseline, intervention, and post-intervention sessions for compound and multisyllabic words, multiple probes of initial and final phonemes were measured. Similarly, while the initial phoneme was in intervention, further probes of the final phoneme were taken. The purpose of this was to investigate potential generalisation effects of building frequencies in compound and multi-syllabic words on initial and final phoneme. Figure 13 shows that P1 and P4 did not demonstrate generalisation from compound word to phonemes, nor from initial to final phonemes; both required explicit fluency instruction at each level. Figure 13 shows that Participants 2 and 3 demonstrated generalisation from word/syllable level, to the level of the phoneme (P2 – initial; P3 – initial and final). Both participants, however, received the PT intervention in multi-syllabic words as well as compound words, and it is therefore unclear as to which linguistic unit occasioned generalisation effects at the level of the phoneme. This suggests that building frequencies in compound words and multisyllabic words can generalise to the level of the phoneme. This should not, however, be a reason to target multi-syllabic words before targeting phonemes. Participants demonstrated inconsistent correct responding and experienced goal failure due to the varied linguistic complexity of the multi-syllabic words in the curriculum used. In addition, phonemic awareness is more closely related to reading

achievement than is phonological awareness, and can be targeted despite at or near zero correct responding in baseline.

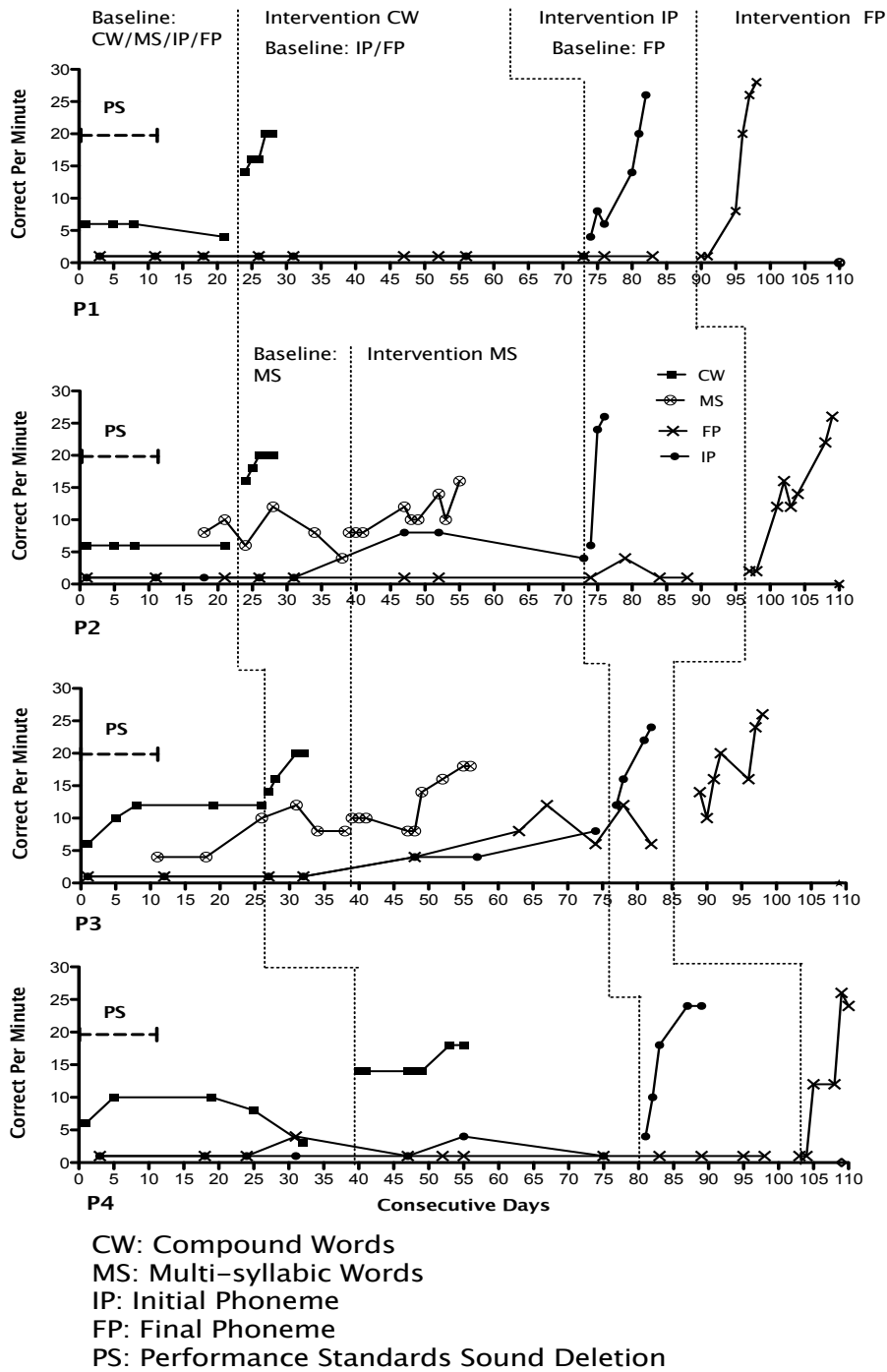


Figure 13. Experiment 4: multiple-baseline across participants with multiple probes across skills design for the PT intervention in sound deletion.

2.4.2.2. Pre- Post-Test Outcomes in Sound Deletion

Table 4 quantifies participants' post-test gains in: standard scores for the sound deletion subtest (SD-SSG) and resulting standardised score ratio gains (SD-SSRG), status changes (SD Pre-status/SD Post-status). Average gains in YARC standard scores for the subtest sound deletion was 9.25 (range 0-21) with an average standard score ratio gain per hour of intervention of 4.8 (range 0-11.7). This observed effect was achieved with the average participant in approximately two hours of intervention time (range 1.8-2.3 hours). Daily intervention sessions on average lasted eight minutes, while it took on average 16 intervention sessions to reach criterion levels of frequency in sound deletion of compound words, and, initial and final phonemes. Of the four participants at pre-test, three were identified as having a severe difficulty (P2, P3, P4) and one with a moderate difficulty (P1). Table 4 shows that at post-test two participants demonstrated average performances (P2 and P3), while the remaining two remained in their pre-intervention category (P1 – moderate difficulty; P4 – severe difficulty).

Table 4

Pre-Post-Test Outcomes in YARC Sound Deletion (SD Subtest for Participants that Received the PT Intervention in Single Letter Sound (Experiment 4)).

	SD-SSG	SD-SSRG	Total hours Intervention	SD Pre-Status	SD Post-Status
P1	4	2	2	Below Average	Below Average
P2	12	5.5	2.2	Severe Difficulty	Average
P3	21	11.7	1.8	Severe Difficulty	Average
P4	0	0	2.3	Severe Difficulty	Severe Difficulty

2.4.2.3. Long-Term Intervention Effects: Maintenance of Sound Deletion

Figure 14 shows participants' performance in correct and incorrect responding in sound deletion of compound words, for initial phonemes, and for final phonemes across time. Each graph represents a participant's performance at median baseline, intervention end, one week post- intervention, two weeks post- intervention, four-to-six weeks post-intervention, eight-to-ten weeks post-intervention, and 34-38 weeks post- intervention. Due to school absenteeism over the five scheduled maintenance checks, participants show an average of five checks (range 4-5). Figure 14 shows that the average performance in compound words at median baseline was 9 CPM (range 4-12), and at intervention end this had increased to 20 CPM (range 18-20); a 2.2-fold increase in correct responding. The maintenance checks carried out at week 1, week 2, and 4-6 weeks, and 8-10 weeks post- intervention demonstrate average performance of 20 CPM (range of 16-24).

Figure 14 shows that the average performance in initial phoneme at median baseline was 4 CPM (range 0-6), and at intervention end this had increased to 25 CPM (range 24-26); a 6.3-fold increase in correct responding. The maintenance

checks carried out at Week 1, Week 2, 4-6 weeks, and 8-10 weeks post-intervention demonstrated average performance of 14 CPM (range of 0-28). At the 4-6 week maintenance check, all participants deleted the final phoneme instead of the initial phoneme, resulting in a score of 0 CPM. Notably, correct responding returned to performance standard levels at the next two retention checks.

Figure 14 also shows that the average performance in final phoneme at median baseline was 3 CPM (range 0-8), and at intervention end this increased to 26 CPM (range 26-26); an 8.6-fold increase in correct responding. The maintenance checks carried out at Week 1, Week 2, 4-6 weeks, and 8-10 weeks post- intervention demonstrate an average performance of 22 CPM (range of 16-28). This suggests that the performance standards, or reduced criterion, achieved by participants in the intervention were largely maintained until the end of the school year (8-10 weeks post- intervention).

Performance at the delayed maintenance check showed a 4-fold decrease in correct responding; Participants 2 and 4 deleted the initial phoneme instead of the final phoneme, resulting in a score of 0 CPM. As deletion of both initial and final probes was measured at the same maintenance checks, this may have resulted in the wrong phoneme exerting stimulus control.

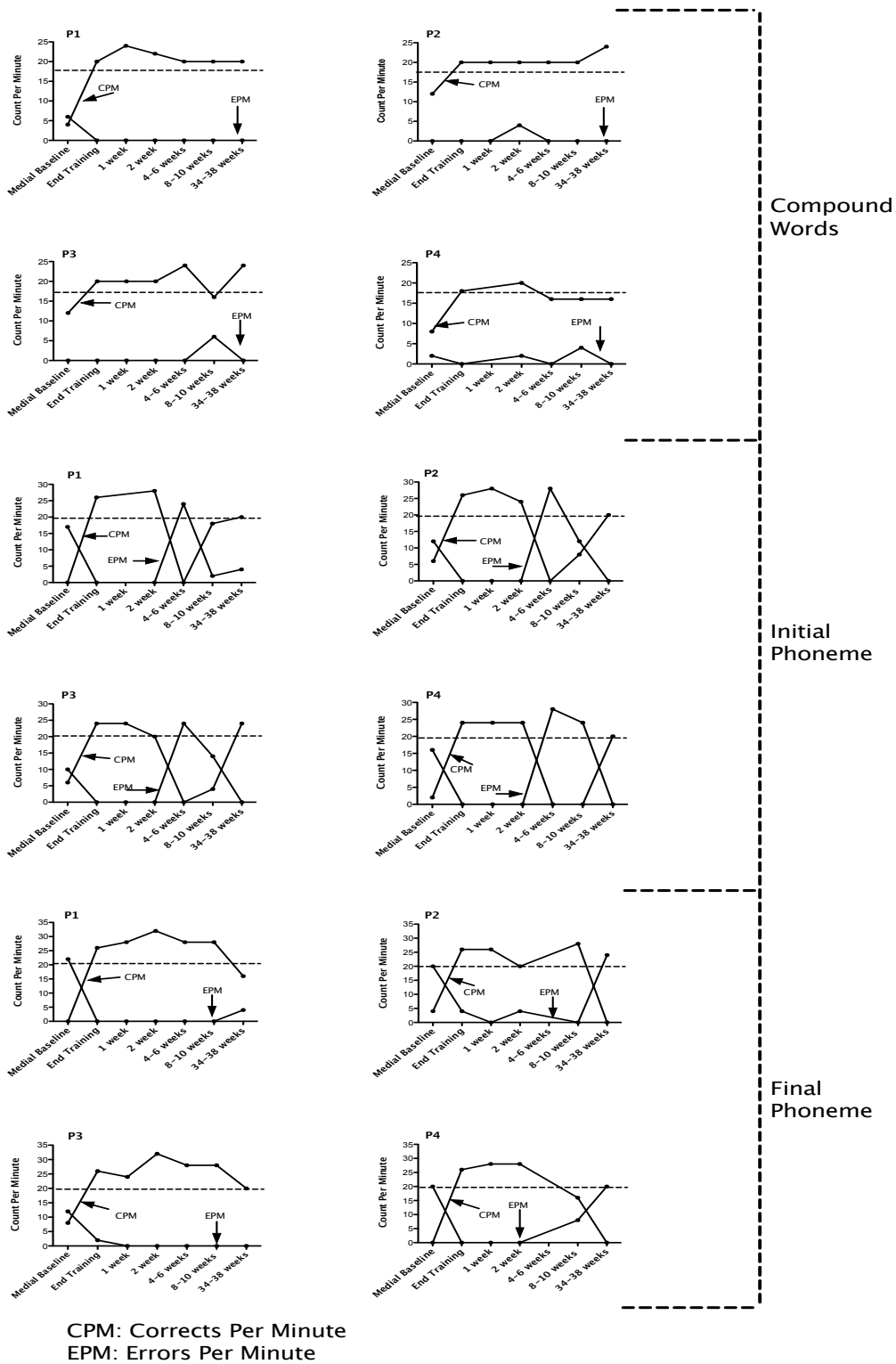


Figure 14. Experiment 4, maintenance data displaying performance in sound deletion at median baseline, intervention end and five post-intervention checks: 1 week; 2 weeks; 4-6 weeks; 8-10 weeks, and 34-38 weeks.

2.4.3. Experiment 4: Conclusions

The purpose of Experiment 4 was to identify a group of Senior Infant students in need of additional instruction in the phonemic awareness skill of sound deletion, and to implement the PT intervention programme with these participants. The goal of the intervention was to build frequencies in sound deletion to performance standards, and for participants to close the gap with their average performing peers. Generalisation and long-term intervention effects were also investigated.

A group of participants in need of additional instruction in sound deletion were identified using the YARC Sound Deletion subtest, and the PT intervention programme was implemented with these students. The intervention was successful in building frequencies in compound words, initial and final phonemes to performance standards for all participants. There were inconsistent gains on the norm-referenced measure, precluding evaluation of “closing the gap” with average peers, but long-term effects suggest that the skills are maintained up to nine months post-intervention.

A core goal of the study was to identify participants with difficulties in PS awareness. The SD measure was effective in identifying a pool of students to target with the PT intervention in this skill. However, there is a possibility that the measures and criteria used (i.e., standard score cut-off points) may have selected “false positives” for intervention. While a number of studies have used sound deletion as a measure of PS skills (e.g., Hogan, Catts, & Little, 2005; Savage et al., 2009), the fact that SD is considered one of the most difficult in a hierarchy of PS skills, using SD as a proxy for PS may have over-identified participants who may naturally have developed such awareness if tested at a later date.

Pre-post-test results from Experiment 4 were inconsistent across participants, despite evidencing large standard score ratio gains for two participants (P2 and P3),

the remaining two (P1 and P4) did not achieve average status in Sound Deletion at post-test. This was somewhat surprising, as all had reached performance standards in sound deletion. The content of the YARC Sound Deletion subtest, however, may have placed a floor effect on performance due to a limited number of items (12 items, with three devoted to medial phoneme - an untargeted skill in the current study).

An important consideration in educational programmes is generalisation of skills taught. The current research examined whether targeting one skill would generalise to specific related skills. Of interest was the possibility that targeting larger units such (e.g., compound words/syllable) may generalise to smaller units (e.g., syllable/phoneme). Experiment 4 showed that targeting compound words did not generalize to increases in correct deletion of syllables, or immediately to initial phonemes. However, for Participants 2 and 3 who received the PT intervention in compound words and multisyllabic words, generalisation to final and/or initial phoneme was demonstrated. While this is an interesting finding, the current results do not support targeting multisyllabic words before targeting phonemes. During the study when targeting multisyllabic words participants' acceleration (growth in frequency) was not commensurate with trajectories of similar skills, participants' performance plateaued and both experienced goal failure. Though this may have been an artifact of the stimulus materials used in the current study, the number of syllables ranged from 2-3, and some were compound words (already trained), invariably participants' were not able to delete from three phoneme words, thus confounding measurement.

2.5. Chapter 2: Results and Discussion.

This section of the Chapter evaluates the effects of the PT intervention programme on global outcome measures rather than the specific measures used within

each experiment and described above. In addition, a measure of social validity of the research was considered an important evaluation; participant views of the intervention are presented subsequent to the generalised reading outcomes.

Experiments 1, 2, 3, and 4 were all conducted with participants from one Senior Infant class in a rural DEIS school. In January of the Senior Infant year, all participants were administered a number of screening assessments (i.e., the YARC subtests of - Letter Sound Knowledge; Sound Isolation; Sound Deletion; and, the DIBELS Letter Naming Fluency subtest). The screeners identified students in need of additional instruction in these skills, and the PT intervention provided specifically targeted these skills. Some participants' demonstrated difficulties across more than one assessment, therefore some participants participated in more than one experiment. This section of the chapter provides a synthesis of participant outcomes in additional reading measures and reports social validity findings.

In the interest of clarity, participants have been assigned a pseudo name to allow the reader to identify and evaluate participant performance across the experiments. For example, Con was identified to be in the range of severe difficulty in letter sounds, sound isolation, and sound deletion; he subsequently participated in Experiments 1, 3 and 4. Table 5 shows Participant (pseudo) names, gender, language status, age at pre-test, and participant identification numbers across experiments.

Table 5

Participant Characteristics, and Participant Identification Across Experiments 1, 2, 3 and 4.

Name	Gender	English Language Status	Age at Pre-test	Participant ID Exp. 1 <i>Letter Sounds</i>	Participant ID Exp. 2 <i>Letter Names</i>	Participant ID Exp. 3 <i>Sound Isolation</i>	Participant ID Exp. 4 <i>Sound Deletion</i>
Jay	Male	EL2	5.8	P1		P3	
Kim	Female	EL1	5.7	P2		P4	
Lea	Female	EL1	5.5	P3	P1		
Abe	Male	EL1	6.8	P4		P5	
Gus	Male	EL1	5.9	P5	P3		
Con	Male	EL1	6.5	P6		P2	P2
Ela	Female	EL1	5.9	P7	P2		
Ike	Male	EL1	5.8			P1	P1
Del	Male	EL1	6		P4		P4
Ben	Male	EL1	6.5		P5		P3

EL1: English as a first language

EL2: English as a second language

Each Experiment used subtests that assessed the exact nature of the foundational reading skills targeted in that experiment e.g., Experiment 1 assessed and intervened at the letter sound level. The skills targeted in each experiment are considered foundational skills in reading, but may not be necessarily sufficient to establish word reading. It was important to also examine the effects of the intervention on global measures than the specific ones used in each experiment.

Also, because all participants took part in more than one experiment, it is challenging to disentangle the relative contribution of the foundational skill targeted to overall reading development. As overall reading development was the distal goal, it

was therefore important to measure related outcomes in word reading, letter sound fluency, and nonsense word decoding in a pre- post-test context. To this end, all participants were administered three additional assessments as pre- post-tests to evaluate the contribution of the PT intervention targeting foundational reading skills on overall reading development. These findings are presented as generalised outcomes in reading measures.

In addition, social validity was central to the research question in order to measure the suitability of the Tier 2 intervention to existing school curricula and practices. As such, the study aimed to investigate acceptability and viability of the intervention within the applied setting. Two Likert-type social validity questionnaires were constructed to assess the following elements of the intervention: (a) the goals of treatment; (b) the treatment procedures, and (c) the outcomes produced by treatment procedures. Questionnaires were administered to teachers (see Appendix E) and parents (see Appendix F). Both questionnaires contained six items/statements each rated on a 5-point Likert-type scale (strongly agree, agree, don't know, disagree, and strongly disagree). The questionnaires also had a section for additional comments. Social validity results are presented subsequent to generalised reading outcomes.

2.5.1. Generalised Outcomes In Reading Measures for Participants in Experiments 1, 2, 3, and 4.

Generalised outcomes in reading are presented as pre- post-test gains in real and nonsense word measures. Participants in Experiments 1, 2, 3, and 4 were administered three additional measures as pre- post-tests: the YARC Early Word Reading (EWR) subtest; the CBM Letter Sound Fluency (LSF) subtest, and the Dynamic Indicators of Basic Early Literacy (DIBELS) Nonsense Word Fluency (NWF) subtest.

Word reading was assessed using the Early Word Reading (EWR) subtest also from the YARC assessment battery (Hulme, et al. 2009). The EWR subtest contains 30 words (15 regular, 15 irregular), its reported reliability is .98 (Chronbach's alpha) and its predictive validity coefficient with the Single Word Reading Test (Foster, 2007) is .88 (Hulme, et al. 2009). The Curriculum-Based Measure Letter Sound Fluency Test (CBM LSF; Fuchs & Fuchs, 2003) was used to measure outcomes in letter sound fluency. The LSF is reported to have a test-retest coefficient of .89 in kindergarten, concurrent criterion related validity of .71 for word identification, and .66 for decoding subtests in the Woodcock Reading Mastery Tests (Speece & Case, 2001, as cited in Fuchs & Fuchs, 2003).

Decoding was assessed using the DIBELS NWF subtest, a randomly ordered nonsense vowel-consonant (VC) and consonant-vowel-consonant (CVC) words (e.g., "vaj"); its reported alternate form reliability is .83 in first grade. The NWF subtest is also reported to demonstrate concurrent, criterion related validity with the Woodcock-Johnson Psycho-Educational Battery-Revised Readiness Cluster standard scores of .51 in first grade (Good & Kaminski, 2002).

Table 6 displays the outcomes in the YARC EWR subtest for participants in Experiments 1, 2, 3, and 4. Participant results are grouped according to the PT intervention received e.g., single letter sounds. Table 6 depicts participants' total hours of intervention received; standard score gains (EWR-SSG); standard score ratio gains (EWR-SSRG); pre-test status (EWR Pre-Status); post-test status (EWR Post-Status); and, percentile gains (EWR Percentile). Table 6 shows that participants gained an average of 4.4 (range 0-22) standard score gains in early word reading; and an average standard score ratio gain of one (range -1.35-4.8). This means that

participants gained approximately one standard score in word reading, as a ratio to hours of intervention received across Experiments 1, 2, 3, and 4.

Table 6 shows that at pre-test, eight participants were in the severe difficulty category (Abe, Ben, Con, Ela, Gus, Jay, Kim, and Lea) and two in the average category (Del and Ike). At post-test, one participant moved from severe difficulty to below average (Ben), and one to average (Lea). Four participants made no standard score gains (Abe, Con, Gus, and Ike) in word reading. Gus also failed to make standard score gains in word reading, and notably on both letter sound and letter name measures - despite the PT intervention targeting these skills. Though Ike's raw score had increased by seven, his standard score had dropped by five; however, he still remained in the average category of word reading. Table 6 also shows there was an average increase of 9.6 percentile points across participants.

Table 6

Outcomes in the YARC Early Word Reading Subtest Across Experiments 1, 2, 3, and 4.

	Total Hours Intervention	EWR- SSG	EWR- SSRG	EWR Pre-Status	EWR Post- Status	EWR Percentile Gains
Participants that Received the PT Intervention in Single Letter Sounds and Sound Isolation						
Abe	4.9	1	.2	Severe Difficulty	Severe Difficulty	0
Ben	3.6	4	1.1	Severe Difficulty	Severe Difficulty	3
Kim	5.6	-2	-.36	Severe Difficulty	Severe Difficulty	-1
Participants that Received the PT Intervention in Letter Names and Sound Deletion						
Ben	3.9	10	2.6	Severe Difficulty	Below Average	7
Del	2.7	13	4.8	Average	Average	29
Participants that Received the PT Intervention in Single Letter Sounds and Letter Names						
Ela	2.6	1	.4	Severe Difficulty	Severe Difficulty	1
Gus	3.8	0	0	Severe Difficulty	Severe Difficulty	0
Lea	3.9	22	3.6	Severe Difficulty	Average	68
Participants that Received the PT Intervention in Other Combinations						
Con*	7.7	0	0	Severe Difficulty	Severe Difficulty	0
Ike**	3.7	-5	-1.35	Average	Average	-11
Averages	4.2	4.4	1			9.6

*Con received the PT intervention in single letter sounds, sound isolation, and sound deletion.

**Ike received the PT intervention in sound isolation and sound deletion.

Table 7 displays the outcomes in the CBM LSF subtest, and the DIBELS NWF subtest, for participants in Experiments 1, 2, 3, and 4. Table 7 depicts participants' letter sound fluency pre-test (LSF Pre-Test) and post-test score (LSF Post-Test); and, nonsense words read correctly at pre-test (NWF-WRC Pre-Test) and post-test (NWF-WRC Post-Test). Results are described as the average correct per minute (CPM) gains from pre- to post-test, and resulting fold change in average performance. Table 7 shows participants that received the PT intervention in letter sounds (P1, P12, P13, P7, P10, and P14) gained an average of 31 (range 24-44) letter sounds CPM; this represents an average 3.2-fold increase. In contrast, those participants that not receive the PT intervention in letter sounds (P3, P6, and P4) gained an average of eight (range -33-10) letter sounds CPM; this represents an average .85-fold decrease. Table 7 also shows all participants gained an average of 1.2 correctly decoded nonsense words; this represents a 5-fold increase in number of nonsense words correctly decoded from pre- to post-test.

Table 7

Outcomes in the Curriculum Based Measure Letter Sound Fluency (LSF) and DIBELS Nonsense Word Fluency (NWF) Subtest Across Experiments 1, 2, and 3, and 4.

	Total Intervention Sessions in Exp. 1, 2, 3, and 4.	LSF Pre- Test	LSF Post-Test	NWF-WRC Pre-Test	NWF- WRC Post-Test
Participants that Received the PT Intervention in Single Letter Sounds and Sound Isolation					
Abe	39	2	46	0	1
Ben	29	9	48	0	0
Kim	36	0	36	0	0
Participants that Received the PT Intervention in Letter Names and Sound Deletion					
Ben	30	80	47	1	1
Del	22	47	57	2	6
Participants that Received the PT Intervention in Single Letter Sounds and Letter Names					
Ela	23	13	47	0	0
Gus	24	48	48	0	0
Lea	31	20	60	0	4
Participants that Received the PT Intervention in Other Combinations					
Con*	52	6	30	0	0
Ike**	27	26	26	0	6
Averages	31			.3	1.5

*Con received the PT intervention in single letter sounds, sound isolation, and sound deletion.

**Ike received the PT intervention in sound isolation and sound deletion.

2.5.2. Social Validity Results.

Eight parents strongly agreed and one agreed with statements relating to the goals of intervention and outcomes of the treatment procedures. Two parents

commented on an improved attitude towards school and four on changes in attitude to homework (e.g., “more enthusiastic”, “volunteering to do it”). Five parents commented on an increased interest in books, and two parents believed that their child was reading more since taking part in the intervention. Four parents reported a marked improvement in their child’s overall progress. In relation to the treatment procedures, six parents commented positively about the treasure chest and nine about the star charts. Five parents referred to the enjoyment that their child felt from taking part in the intervention.

Both teachers strongly agreed with statements relating to the goals of intervention, the treatment procedures, and the outcomes of the treatment procedures. Teachers also expressed clear benefits observed for the students who took part in the intervention in terms of student engagement and motivation, acquisition of core reading skills and the consolidation of existing skills leading to an overall improvement in reading ability. Teachers commented on students “increased self-esteem resulting from experiencing success and positive feedback”.

2.6. Chapter 2: General Limitations

Limitations of the experiments reported in the current chapter are associated with the applied nature of the research and the multiple components of the intervention programme. Limitations also pertain to the individual experiments, and subsequent overall data analysis of the studies combined. The current research was conducted in the applied setting, and attempted to provide both skill specific and hybridised Tier 2 interventions. Tier 2 interventions were skill specific, this means that students received interventions on early reading skills that that were directly related to universal screening outcomes. This meant that all students did not receive

the same intervention, nor was the PT intervention limited to only one target skill but targeted multiple skills across participants. As a result it is difficult to isolate the effects of the intervention on one specific skill. In addition, a hybrid model of Tier 2 intervention was applied, therefore although interventions were standardised, the PT intervention programme provided for a consistent approach to intervention modification where student performance data indicated such a need (i.e., falling below a $\times 2$ celeration over two consecutive days).

Another issue to consider in relation to the application of the intervention programme was its multi-componential nature. The three main components included the intervention, progress monitoring, and decision-making rules. In addition, each of these components was composed of further subcomponents. For example, the intervention component comprised of frequency building trials (and sometimes discrete trials), goal setting, and reinforcement and error correction. Each of these components has been shown to be effective instructional strategies, however, the relative contribution of each to the overall success of the intervention is unclear.

A limitation in regards to progress monitoring was that progress was evaluated solely on the skill targeted, additional curriculum based measures were not collected. Therefore, it is unclear if building frequencies in skills targeted generalised to overall reading behaviour. The decision making rule was based solely on celeration (i.e., $\times 2$), while this allowed for a consistent approach to programme modification, variations in the celeration value (e.g., $\times 1.5$) may have yielded a different programme modification profile. Taken together, the applied nature of the research and the multi-component nature of the PT intervention programme had a number of implications for disentangling the differential effects of this programme.

Two main issues must be considered in regards to the overall data analysis. First, due to the experimental design (SCED), one cannot be assured that pre- post-test gains were the result of the intervention; because there is no control group, there is no comparative group with which to compare post-test gains. Second, because participants took part in more than one experiment it is difficult to untangle the relative contribution of each skill targeted to observed post-test gains in word and nonsense word reading. Limited increases were observed in nonsense words gained at post-test, and were unevenly distributed across participants. The same outcome applies to word reading. This may suggest that increasing frequencies in the skills targeted was not sufficient for decoding behaviour to emerge. This may be due in part to the nature of the phonemic awareness skills targeted. Sound isolation and sound deletion are not functionally related to decoding (i.e., they are not focused on segmenting or blending sounds), and therefore may have limited generalisation effects to decoding. During the intervention generalisation was not programmed for as the research was specifically concerned with evaluating the effects of building frequencies in foundational skills.

A potential limitation of targeting more than one foundational skill is that that the PT intervention in one skill may have accelerated growth in another. This consideration may be offset in part by the fact that participants did not receive the PT intervention in more than one skill at a time. In addition, stable baselines were observed across experiments, indicating that if a cross-contamination of the PT intervention did occur, effects were negligible.

Another limitation is the fit of the Tier 2 intervention within an overall RTI framework. The evaluation of the current intervention within an RTI framework was beyond the scope of the current research. Both PT and RTI are concepts and practices

that are wholly unfamiliar to school systems in Ireland, therefore the current research was conducted in a supplemental (and perhaps isolated) nature to the general education provided in the classroom. The overall content of the core reading programme is unclear however, the curriculum used was *Jolly Phonics*, which is a synthetic phonics program that targets letter sounds and blending. The Department of Education and Skills mandates that reading instruction be allocated for 90 minutes per day, however the duration of daily core reading was not measured. In addition, the degree to which it was implemented with fidelity is unknown for this setting. As such evaluation of the Tier 2 intervention is not within the context of an RTI framework.

Alignment of Tier 2 to the core reading curriculum is an important consideration (Gersen et al., 2009). In the absence of fidelity information on the core reading curriculum, the degree of alignment of the Tier 2 intervention, is difficult to evaluate. Letter sounds can be considered as highly aligned, however in the *Jolly Phonics* programme letter names are not targeted at the same time. This may partly explain difficulties experienced by some participants in building frequencies in letter sounds to performance standards, and may have confounded measurement of each skill (this is true for one participant, who frequently produced the letter sound when being measured for letter names). The phonemic awareness skill of sound isolation is aligned to *Jolly Phonics*, where there is an emphasis on hearing and identifying phonemes in words. Sound deletion may be considered in part aligned with *Jolly Phonics* for the same reason (i.e., hearing and identifying words). However, this skill is considered the most difficult of all the phonemic awareness skills, and not targeted at Senior Infants level.

To address this limitation alignment of the intervention programme to the core reading curriculum was a central focus of the next phase of the research. Thus, Chapter 3 provides an account of the PT intervention programme targeting a different set of foundational reading skills to those described in Chapter 2, with a more diverse student population within an urban school setting.

Chapter 3: Introduction

Chapter 2 described the PT intervention programme targeting fluency in letter sounds, letter names, and phonemic awareness skills of sound isolation and sound deletion in a rural DEIS school. The current chapter attempts to address a number of limitations identified in the experiments reported in Chapter 2. Specifically, the setting involved a school in a rural area, with a small number of students identified as at risk from one classroom. In addition, only one participant spoke English as a second language (EL2). Other classrooms in urban area have much larger class sizes, and a higher ratio of students with EL2. This phase of the research specifically targeted an urban school that accommodated three Senior Infant classrooms that consisted of students from diverse backgrounds.

Similarities and differences are observed in the intervention focus between this phase and the last in the rural DEIS school. Letter sound fluency was screened and targeted for intervention in Experiment 5, and using the same intervention protocol MLGPCs were targeted for intervention in Experiment 6. Letter names were not targeted across experiments in this chapter, as they were not aligned to the general reading curriculum. In addition, sound isolation and sound deletion were not targeted, as their functional relation to decoding behaviour remains unclear. As an alternative, blending sounds was screened and targeted for intervention in Experiment 7. Blending sounds was considered to have a more direct relationship with decoding behaviour, and is closely aligned with the core reading curriculum (i.e., *Jolly Phonics*).

3.1. Experiment 5: Letter Sounds

The purpose of Experiment 5 was to identify a pool of Senior Infant students in need of additional instruction in letter sounds, and to implement a PT intervention programme as a Tier 2 intervention with these students. Through an evaluation of maintenance data, Experiment 5 aimed to investigate the effectiveness of the PT intervention in establishing letter sound fluency. This Experiment also aimed to evaluate the effect of the PT intervention programme in helping participants close the gap with average performing peers in the target skill of letter sounds. The intervention goal involved achievement of performance standards in letter sounds (70-100 CPM) at a learning rate of x2 celeration.

3.1.1. Method

3.1.1.1. Participants and Setting

Three Senior Infant classes (kindergarten equivalent; $n=75$) within a DEIS (“Delivering Equality of Opportunity in Schools”) school participated in this phase of the experiment. DEIS schools are situated in educationally disadvantaged areas, with a high percentage of the school population experiencing socio-economic disadvantage. Parents received information letters through the school with an option to consent to participation. A total of 72 participants ($M = 5.7$ years, range 5.1-6.8 years) participated in the universal screening phase.

3.1.1.2. Universal Screening Phase

The screening instrument used was identical to that in Experiment 1 (Chapter 2). Of the 72 participants screened, five were excluded, as they were allocated through the school to receive Reading Recovery as a direct reading intervention. Of the remaining 67, ten participants were identified as having a severe difficulty (P1, P2, P3, P4, P5, P6, P7, P8, and P9; age range 5.5-6.3 years; $M = 5.9$ years) or a

moderate difficulty (P3, age 5.6 years) in letter sound knowledge. All participants were typically developing (four males and six females); five spoke English as a first language (EL1), and five as a second language (EL2). Two EL2 participants (P8 and P9) had significant speech difficulties in their native language, and both attended weekly speech and language therapy sessions to address speech difficulties.

3.1.1.3. Dependent Variables and Data Analysis

Dependent variables and data analysis were identical to that described in Experiment 1 (Chapter 2) with one exception. Post-intervention rates in letter sounds were measured up to 28 weeks later to investigate maintenance of treatment effects.

3.1.1.4. Materials

Stimulus and other materials were identical to that described in Experiment 1 (Chapter 2).

3.1.1.5. Design

Two multiple-baseline across participants designs were employed to assess the whether the PT intervention programme would be effective in building frequencies in letter sounds to performance standards. Six participants were assigned to one multiple-baseline design (P1, P2, P3, P4, P5, and P6), and four to another (P7, P8, P9, and P10).

3.1.1.6. Procedure

Baseline. Baseline procedures were identical to that in Experiment 1 (Chapter 2).

Intervention. The independent variable involved a Tier 2 PT intervention programme that included PT as an intervention to target fluency in letter sounds, and a decision-making framework to monitor progress. Intervention procedures were identical to that in Experiment 1 (Chapter 2). Participants received the one-to-one

intervention five days per week, daily sessions lasted on average 6.6 minutes, and the average total intervention duration was 1.5 hours. Each intervention session consisted of discrete trials in letter sound discrimination and production, and frequency building to performance standards.

Decision Rules and Programme Modifications. Decision rules and programme modifications were identical to that in Experiment 1 (Chapter 2).

3.1.1.7. Interobserver agreement (IOA). IOA data were calculated as described in Experiment 1 (Chapter 2). Exact Count-Per-Interval IOA data were collected for 40.5% of all baseline sessions demonstrating 97.8% agreement, and 21% of all intervention sessions achieving 96% agreement.

3.1.1.8. Procedural integrity. The procedural integrity checklists were identical to those used in Experiment 1 (see Appendix C for baseline checklist, and Appendix D for intervention checklist); PI data were calculated in the same manner as Experiment 1 (Chapter 2). Independent observers collected data on procedural integrity for 35% of baseline and 21.8% of intervention sessions. Overall procedural adherence was 100% for baseline and intervention sessions.

3.1.1.9. Data analysis.

The data analysis procedures employed were identical to that described in Experiment 1 (Chapter 2).

3.1.2. Results and Discussion

Figure 15 depicts two multiple-baseline designs (MBD 1, MBD 2) across participants ($n = 6$; $n = 4$) designs for the PT intervention in letter sounds. Figure 15 shows that in MBD 1 (P1, P2, P3, P4, P5, and P6) all participants entered the baseline condition within five days of each other, and a minimum of four probes were collected before P1, P2, and P3 were entered into the intervention phase. Participants

1, 2, and 3 responded to the intervention, respectively gaining 16, 12, and 24 letter sounds CPM by the 3rd session. Participant 1 changed schools during the experiment, but Participants 2 and 3 continued to respond to the intervention, reaching performance standards of 68 letter sounds CPM on the 13th (P2), and 11th (P3) session. Participants 4 and 5 were subsequently entered into the intervention and both responded to the intervention gaining 24 and 16 letter sounds CPM by the 3rd session. Participant 4 continued to respond to the intervention for three weeks and then experienced goal failure. The programme was modified (timing trials began with the same start point on stimulus worksheet), and she subsequently achieved performance standards of 68 CPM on the 14th session. Participant 5 continued to respond to the intervention reaching performance standards of 68 CPM letter sounds on the 13th session (see Figure 15).

Participant 6 was last entered into the intervention phase in MBD 1 (see Figure 15). Participant 6 responded to the intervention for three sessions, then a two-week school holiday commenced. On returning to school, P6's performance decreased and on the 6th intervention session goal failure was experienced. The programme was modified (timing trials beginning with the same start point on stimulus worksheet, and use of flashcards); and P6 was subsequently moved to a personal best component. Participant 6 achieved a reduced criterion of 64 letter sounds CPM on the 18th intervention session.

Figure 15 shows that in MBD 2 (P7, P8, P9, and P10) all participants entered the baseline condition at the same time, and a minimum of six baseline probes were collected before P7, and P8 were entered into the intervention phase. Both participants responded to the intervention respectively gaining 28 and 34 letter sounds CPM by the 3rd session, and achieving performance standards (68 CPM) on the 8th

session. Participants 9 and 10 were subsequently entered into the intervention (see Figure 15), and both responded by gaining 16 and 31 letter sounds CPM, and achieved performance standards (68 CPM) on the 11th (P9) and 5th (P10) intervention sessions respectively.

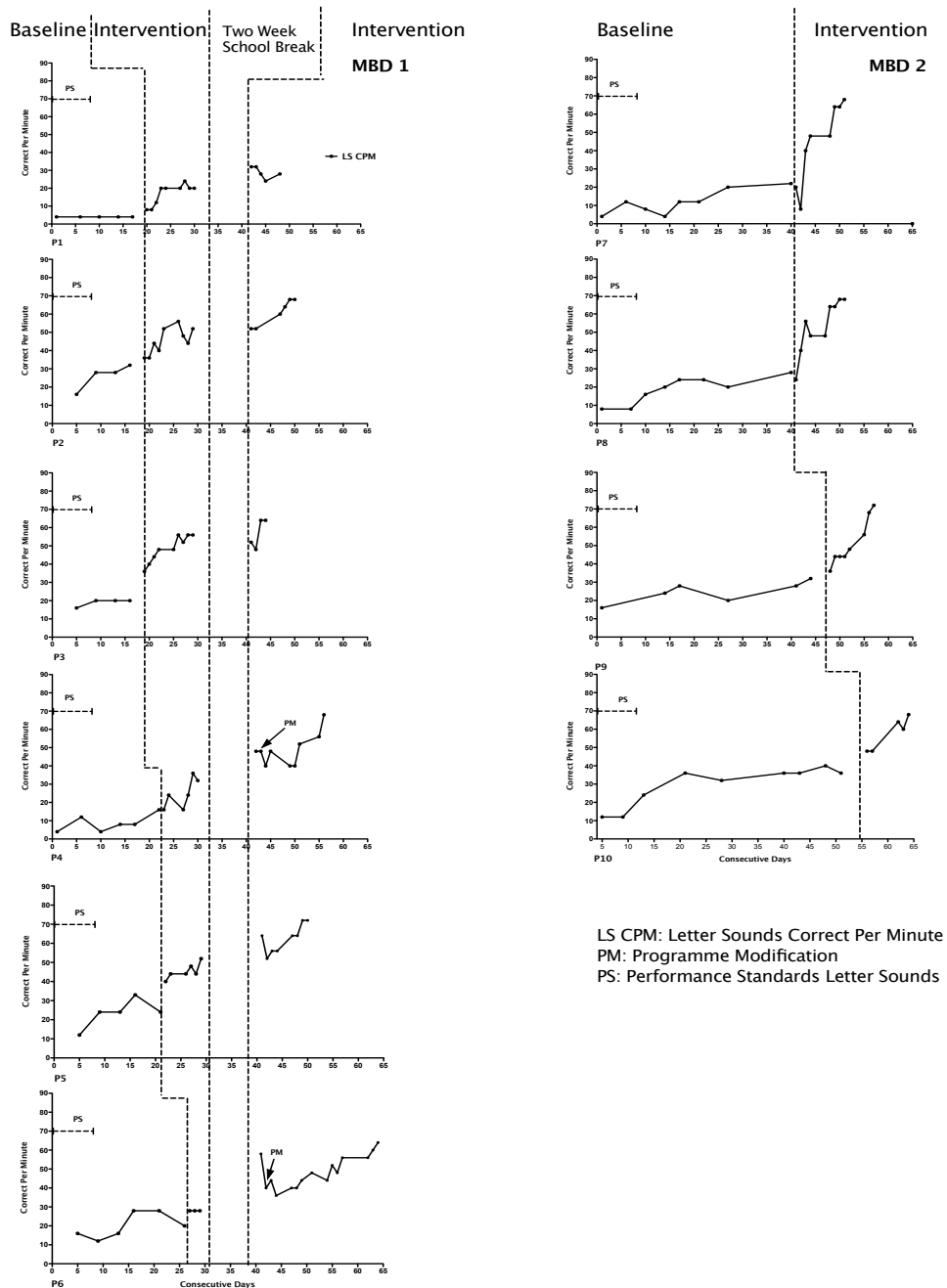


Figure 15. Experiment 5: multiple-baseline across participants design for the PT intervention in letter sounds.

3.1.2.1. Long-Term Intervention Effects: Maintenance of Letter Sounds

Figure 16 shows the maintenance data for letter sounds. Each panel represents a participant's performance at median baseline, end of intervention, one week, two weeks, four-to-six weeks, 10-12 weeks, and 26-28 weeks post-intervention. Due to school absenteeism over the five scheduled maintenance checks, participants show an average of four checks (range 3-5). Figure 16 shows that the average performance in letter sounds at median baseline was 20 CPM (range 8-33), and at the end of the intervention this had increased to 67 CPM (range 64-68); a 3.4-fold increase in correct responding. The maintenance checks carried out at Week 1, Week 2, and 4-6 weeks, and 8-10 weeks post-intervention demonstrate average performances of 61 CPM (range 56-84 CPM). This suggests that the performance standards, or reduced criterion, achieved by participants in the intervention were largely maintained until the end of the school year (10-12 weeks post-intervention). Figure 16 shows that at the delayed maintenance check (26-28 weeks post-intervention), average performance was 62 CPM (range 48-80); a 0.9-fold decrease in correct responding to that achieved in the intervention.

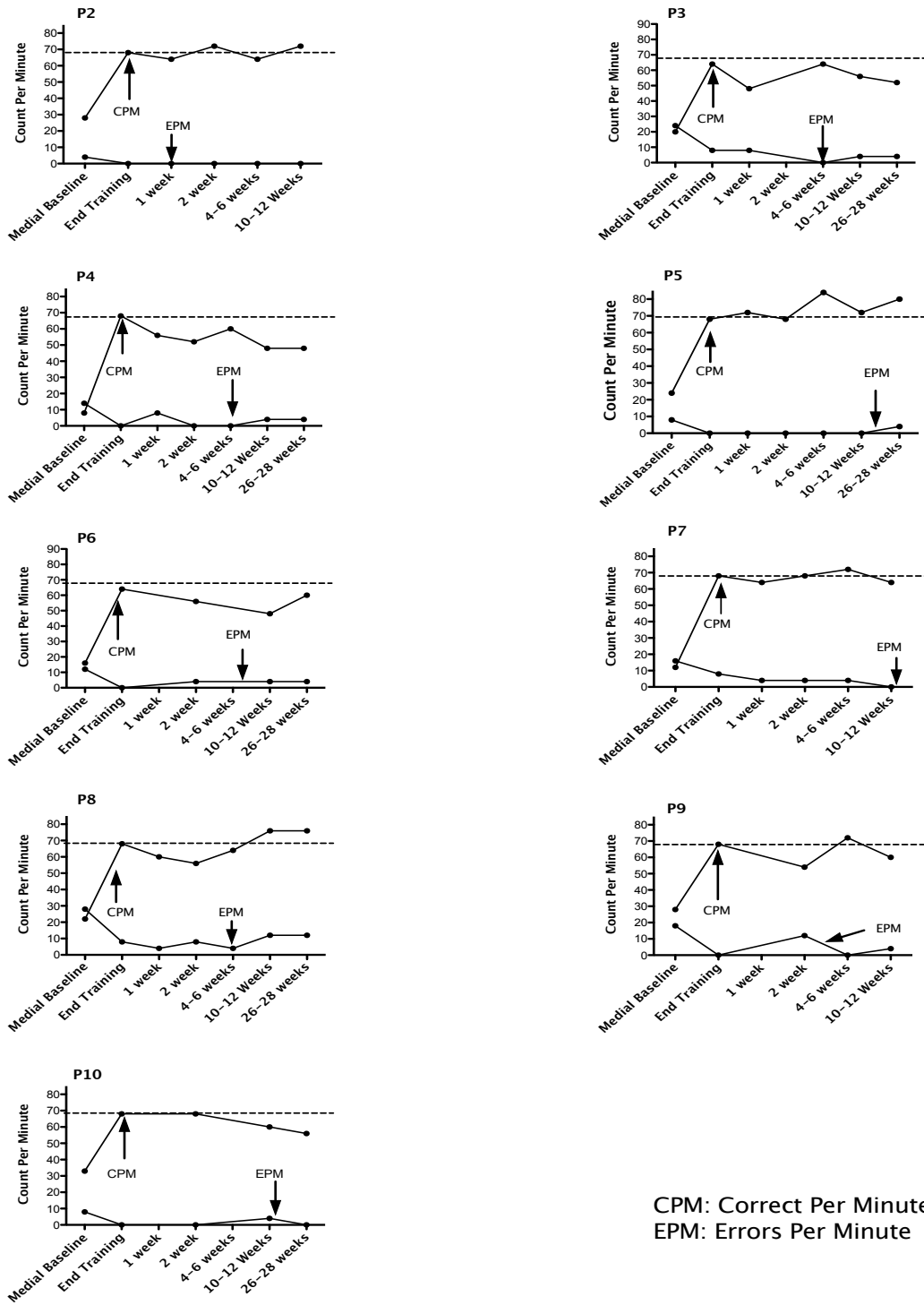


Figure 16. Experiment 5, maintenance data displaying performance in letter sounds at median baseline, intervention end and five post-intervention checks: 1 week; 2 weeks; 4-6 weeks; 10-12 weeks, and 26-28 weeks.

3.1.3. Experiment 5: Conclusions

The purpose of the Experiment 5 was to implement a PT intervention programme as a Tier 2 intervention with Senior Infant readers identified as at risk, to investigate its effectiveness in building frequencies in letter sounds, and to evaluate maintenance of the skill over prolonged periods of time. The goal of the intervention was to build frequencies in letter sounds to performance standards, and for participants to close the gap with their average performing peers.

The current experiment demonstrated that the intervention programme was an effective Tier 2 intervention for targeting fluency in letter sounds with students identified as at risk readers. The intervention procedures were effective in building frequencies in letter sounds to performance standards, or reduced criteria for all participants. Frequency building to performance standards was effective in increasing the frequency and accuracy of letter sounds for all participants. Seven participants achieved at, or near performance standards (70-90 CPM) and the intervention was terminated at a reduced criterion for the other two (64 CPM).

Maintenance data shows that the performance levels achieved at the end of the intervention are largely maintained up to 12 weeks later. At the delayed post-test there is a 0.9-fold decrease.

The add/subtract graphs create the impression of ascending baselines for Participants 2 and 8; this issue has been addressed in detail in the limitations of Experiment 2 (Chapter 2). Specifically the use of the SCC permitted precise quantification of increases in both the baseline and intervention conditions. The increase in trend observed in the baseline was not clinically significant on the level of the individual (i.e., the increase in celeration in the intervention was far greater than that observed in the baseline).

3.2. Experiment 6: Multi Letter Grapheme Phoneme Conversions

The purpose of Experiment 6 was to investigate if the procedures utilised in Experiment 1 (Chapter 2) could also be used to build frequencies in Multi Letter Grapheme Phoneme Conversions (MLGPCs) with participants who had previously received the PT intervention programme for single letter sounds. The intervention goal involved achievement of performance standards in MLGPCs (70-100 correct per minute; K. Brooks Newsome, personal communication, February 5, 2012) at a learning rate of x2 celeration.

3.2.1. Method

3.2.1.1. Participants and Setting

Participants and setting in the current experiment were identical to those in Experiment 5; four participants (P2, P3, P5, and P8) who had completed the PT intervention for letter sounds were selected for intervention targeting MLGPCs.

3.2.1.2. Materials

Stimulus materials were designed by the experimenter. Each A4 stimulus sheet consisted of printed MLGPCs (six across and 13 down) in size 16 font. The MLGPCs selected for intervention were /oo/, /ou/, /sh/, /ch/, /ee/, and /th/. These MLGPCs were specifically targeted as Carnine and colleagues (1997) had identified each as a highly frequent MLGPC in children's literature. Ten multiple exemplars of each stimulus sheet were created, and three stimulus workbooks were compiled and bound using all 10 exemplars (two experimenter copies and one participant copy). This resulted in randomised exemplars for each measurement occasion, controlling for practice effects.

Other materials were identical to those described in Experiment 1 (Chapter 2).

3.2.1.3. Design

A multiple-baseline across participants design was employed to assess whether the PT intervention programme would be effective in building frequencies in MLGPCs to performance standards.

3.2.1.4. Procedure

Baseline. Baseline procedures were identical to those described in Experiment 1 (Chapter 2).

Intervention. The independent variable involved a Tier 2 PT intervention programme that included PT as an intervention to target fluency in MLGPCs, and a decision-making framework to monitor progress. Intervention procedures were identical to those described in Experiment 1 (Chapter 2).

3.2.1.5. Interobserver agreement (IOA). IOA data were calculated as described in Experiment 1 (Chapter 2). Exact Count-per-Interval IOA data were collected for 38% of all baseline sessions demonstrating 93.7% agreement, and 25% of all intervention sessions achieving 97% agreement. Overall, that is 95% agreement across 31.5% baseline and intervention sessions.

3.2.1.6. Procedural integrity. The procedural integrity checklists were identical to those used in Experiment 1 (see Appendix C for baseline checklist, and Appendix D for intervention checklist); PI data were calculated in the same manner as Experiment 1 (Chapter 2). Independent observers collected data on procedural integrity for 28% of baseline sessions and for 16.7% of intervention sessions evidencing 100% adherence.

3.2.1.7. Data analysis

The PT intervention programme was assessed using a multiple-baseline design. Data were evaluated with respect to gains in correct responding between

median baseline and the 3rd intervention session. Due to the short nature of the intervention, data are presented as a mean outcome across participants. Response to intervention was indexed in an identical manner to that in Experiment 1 (Chapter 2).

3.2.2. Results

Figure 17 depicts a multiple-baseline across participants ($n = 4$) design for the PT intervention programme targeting MLGPCs. Figure 17 shows that all participants entered the baseline condition at the same time, and a minimum of eight baseline probes were collected before P3 was entered into the intervention phase. Participant 3 responded to the intervention, gaining 24 MLGPCs CPM by the 3rd session. Participants 5, 8 and 2 were then entered into the intervention, and all responded to the intervention. Participant 5 gained 48 MLGPCs CPM by the 3rd session; P8 gained 36 CPM; and P2 gained 44 CPM (see Figure 17). All participants maintained or exceeded a x2 celeration for the duration of the intervention. Due to time constraints, the intervention was terminated for all participants at reduced criteria ($M = 59$; range 48-68 CPM) after approximately six intervention sessions (range 4-7 sessions). The observed effect was achieved with the average participant in 30 minutes and daily intervention sessions lasted approximately 5.5 minutes.

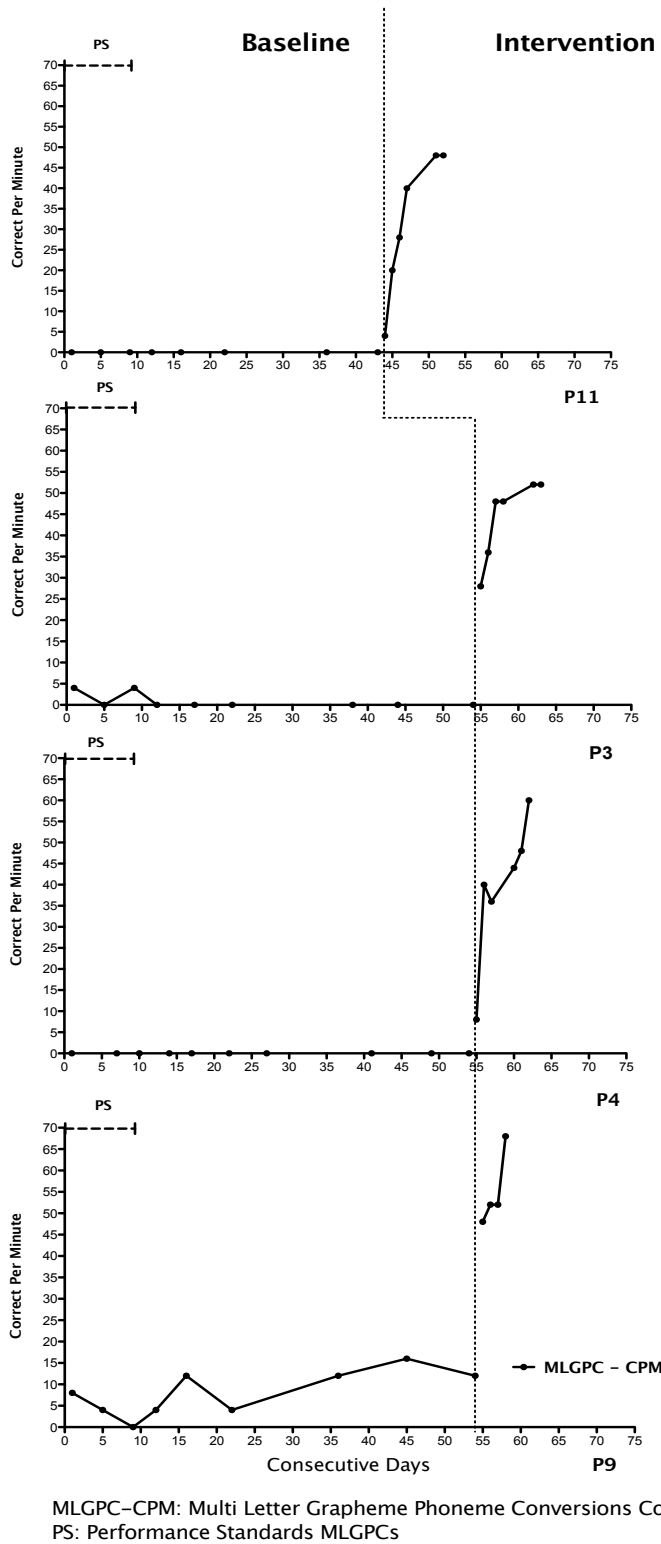


Figure 17. Experiment 6: multiple-baseline across participants design for the PT intervention in multi-letter grapheme phoneme conversions.

3.2.2.1 Long Term Intervention Effects: Multi-letter Grapheme Phoneme

Conversions

Figure 18 shows the maintenance data for MLGPCs. Each panel represents a participant's performance at median baseline, end of intervention, one week, two weeks, 10-12 weeks, and 26-28 weeks post-intervention. Due to school absenteeism over the four scheduled maintenance checks, participants show an average of three checks (range 2-3). Figure 18 shows that the average performance in MLGPCs at median baseline was 2 CPM (range 0-8 CPM), and at the end of intervention this had increased to 59 CPM (range 40-68 CPM); a 29-fold increase in correct responding. The 2-week maintenance check shows no decrease in performance, and at the delayed maintenance check (26-28 weeks post-intervention), average performance was 72 CPM (range 56-96); a 1.2-fold decrease in correct responding to that achieved in the intervention (see Figure 18). These findings suggest that the reduced criteria achieved by participants in the intervention were largely maintained up to seven months post-intervention.

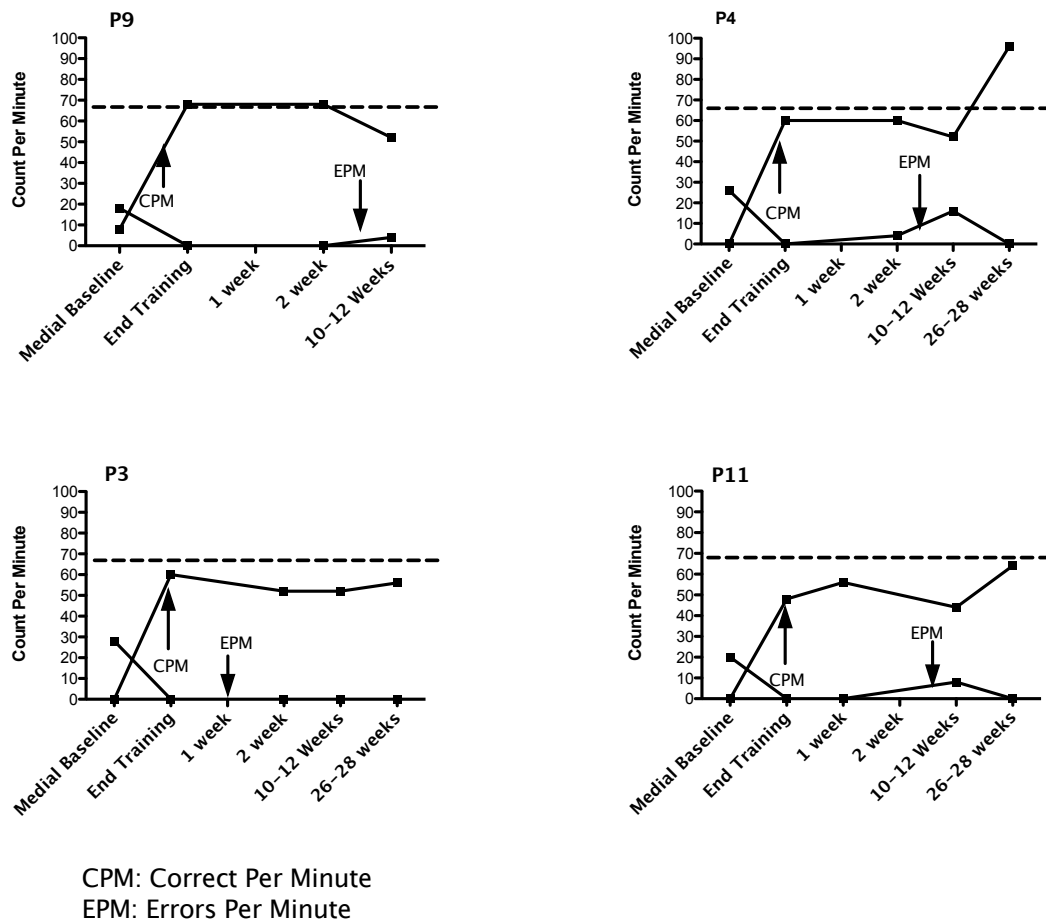


Figure 18. Experiment 6 maintenance data displaying performance in multi-letter grapheme phoneme conversions at median baseline, intervention end and five post-intervention checks: 1 week; 2 weeks; 4-6 weeks; 10-12 weeks, and 26-28 weeks.

3.2.3. Experiment 6: Conclusions

The purpose of Experiment 6 was to investigate if the same procedures used to build frequencies in single letter sounds could be used to build frequencies in MLGPCs to performance standards. The results clearly show that the procedures were effective in increasing frequencies in MLGPCs, albeit to reduced criteria ($M = 59$; range 48-68). The baseline condition shows at or near zero levels of correct responding for MLGPCs across all participants. A strong intervention effect was evident in the consistent performance patterns within condition, i.e., immediate and large increases in correct responding. The PT intervention was terminated at reduced

criteria for all participants due to time constraints; it would seem logical that higher frequencies would have been achieved if the intervention was continued. The frequencies achieved in the intervention, however, are largely maintained up to 28 weeks later, illustrating the potency of the PT intervention in teaching MLGPCs in isolation.

While the effects of the intervention appear to be strong, the functional use of targeting MLGPCs in isolation to occasion decoding behaviour is questionable. Two untimed application checks were conducted in the intervention condition that tested decoding words that consisted of the trained phonemes. Participants were observed not to recognise the MLGPCs when embedded in real words, and tried to decode the MLGPCs not as whole units but as single GPCs (i.e., single letter grapheme-phoneme conversions). These observations are anecdotal, however, as the checks were casual in nature and thus yielded incomplete data.

Maintenance data show that the performance levels achieved at the end of the intervention were largely maintained up to 28 weeks later, and in some cases increased further. This suggests that the performance standards achieved through the current PT intervention programme persisted across significant periods of time.

A main limitation to this experiment may involve only one replication of effect in the intervention condition (three replications are desirable in multiple-baseline designs). After the first participant had demonstrated an intervention effect, due to time constraints, the three remaining participants were added into the intervention condition together. One replication of effect does not permit demonstration of experimental control, as it does not rule out the possibility of history as a threat to validity. A further limitation, however, involves the utility of targeting MLGPCs in isolation for decoding behaviour. The full extent of this is unclear as the

generalisation probes were casual in nature, untimed, and probed only during the intervention. This also precludes comparison between baseline and intervention conditions.

3.2.3.1. Pre-Post-Test Outcomes in YARC Letter Sound Knowledge (LSK) and CBM Letter Sound Fluency (LSF) Subtests for Participants in Experiments 5 and 6

Table 8 presents participants' post-test standardized score, and standard score ratio gains in letter sounds on the YARC letter sound knowledge subtest (LSK-SSG; LSK-SSRG), and pre- to post-test status changes (LSK-Pre; LSK-Post). Table 5 shows the average gain in YARC standard scores for Letter Sound Knowledge for those trained in single letter sounds was 13.6 (range 5-20) with an average SSRG per hour of intervention of 15.8 (range 4.1-35.6). Table 8 also shows the average gain in YARC standard scores for Letter Sound Knowledge for those trained in single letter sounds and MLGPCs was 36.5 (range 33-44) with an average SSRG per hour of intervention of 20.6 (range 17.1-25.8). Of the nine participants that completed Experiment 5, eight were identified as having a "severe difficulty" (P2, P3, P4, P5, P6, P7, P8, and P9), and one with a "moderate difficulty" (P10). At post-test six participants demonstrated average performance (P3, P5, P6, P7, P8, and P10), and one participant (P2) excellent performance (see Table 8).

In regard to the remaining two participants, P4 moved from "severe difficulty" at pre-test to "below average" at post-test, and was subsequently selected by the special education providers in the school for more intensive reading instruction in the form of Reading Recovery. Participant 9 remained in the "severe difficulty" category though demonstrated gains of five standard scores; this student was an EL2 with a significant language difficulty in his native language. These results indicate that the

intervention programme resulted in most participants closing the gap with average peers in letter sound knowledge.

Table 8 shows the observed effect was achieved in approximately 1.5 hours of intervention time (range 0.5-2.4 hours). An average of 12 intervention sessions were required to reach performance standards in single letter sounds (and) MLGPCs, and daily intervention sessions on average lasted 6.6 minutes.

Table 8

Pre-Post-Test Outcomes in YARC Letter Sound Knowledge (LSK) Subtest for Participants That Received the Intervention in Single Letter Sounds, or Single Letter Sounds and Multi Letter Grapheme Phoneme Conversions (Experiments 5 and 6).

	LSK-SSG	LSK-SSRG	Total Hours Intervention	LSK Pre-status	LSK Post-status
Participants that Received the Intervention in Single Letter Sounds (P4, P6, P7, P9, and P10)					
P4	11	6.2	1.8	Severe Difficulty	Below Average
P6	16	6.7	2.4	Severe Difficulty	Average
P7	20	26.5	.8	Severe Difficulty	Average
P9	5	4.1	1.2	Severe Difficulty	Severe Difficulty
P10	16	35.6	.5	Severe Difficulty	Average
Participants that Received the Intervention in Single and Multi Letter Grapheme Phoneme Conversions (P2, P3, P5, and P8)					
P2	44	25.8	1.7	Severe Difficulty	Excellent
P3	33	18.8	1.8	Below Average	Average
P5	36	17.1	2.1	Severe Difficulty	Average
P8	33	20.6	1	Severe Difficulty	Average

3.3. Experiment 7: Blending Sounds

Phonemic awareness (PA) is a critical foundational reading skill (Snowling & Hulme, 2010), however, the construct of PA refers to awareness of and ability to

manipulate the smallest speech sounds, phonemes. This skill can take many forms, for example, sound isolation or sound deletion (described in Experiments 3 and 4, Chapter 2). These skills may be necessary but not sufficient for early decoding behaviour. Indeed, because sound deletion emerges later in development than blending sounds (Pufpaff, 2009) it is unclear if sound deletion is even necessary for early decoding. Conversely, the ability to blend sounds is a pre-requisite for decoding behaviour, which is a composite of phoneme identification and phoneme blending. Accordingly, this phase of the research screened for difficulties in blending sounds, and subsequently targeted this skill using the PT intervention programme.

The purpose of Experiment 7 was to identify a pool of Senior Infant students in need of additional instruction in blending sounds, and to implement a PT intervention programme as a Tier 2 intervention with these students. The intervention goal was achievement of performance standards in blending sounds (10-15 correct per minute; Freeman & Haughton, 1997) at the learning rate of x2 celeration. Through an evaluation of maintenance data, Experiment 7 aimed to investigate the effectiveness of the PT intervention in establishing blending sound fluency. This Experiment also aimed to evaluate the effect of the PT intervention programme in helping participants close the gap with average performing peers in the target skill of blending sounds.

3.3.1. Method

3.3.1.1. Participants and Setting

Participants and setting are identical to that described in Experiment 4 (Chapter 2).

3.3.1.2. Universal Screening Phase

The screener utilized was the Blending Words (BW) subtest of the Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, &

Rashotte, 1999). This instrument includes 20 items that measure blending sounds into words; the reported reliability for this is .88 (Cronbach's alpha) and its predictive validity coefficient with the Woodcock Reading Mastery–Revised subtests of Word Identification and Word analysis is .62 and .61, respectively (Wagner et al., 1999). A participant is considered to be very poor in blending words if he or she achieves a standard score between 1-3, to poor if he or she achieves a standard scores score between 4-5, or to be below average if he or she achieves a standard score between 6-7.

Of the 67 participants screened, four participants (P1, P2, P3, and P4) were identified as very poor, poor, or below average in blending sounds into words (age range 5.5-6.7 years, $M = 6.1$ years). All participants were typically developing (three males, one female), one spoke English as a first language (EL1), and three as a second language (EL2). Two EL2 participants (P4 and P5) had significant speech difficulties in their native language, and both attended speech and language therapy sessions weekly.

3.3.1.3. Dependent Variables and Data Analysis

Dependent variables and data analysis were identical to that described in Experiment 1 (Chapter 2).

3.3.1.4. Materials

The stimulus materials included the Unit II blending section of “Phonological Coding: Word and Syllable Awareness” curriculum (Freeman & Haughton, 1997); and, Unit II, blending section of the “Phonological Coding: Phonemic Awareness” curriculum (Freeman & Haughton, 1997). These materials were selected for the intervention as they reflected the content domain of the CTOPP subtests selected as outcome measures (blending words), and they provided multiple exemplars of the

phonemic skills targeted in the intervention. This curriculum has been used in precision teaching centres of excellence in the United States (e.g., Fit LearningTM, Haughton Learning Center, Morningside Academy). Other materials used were identical to those described in Experiment 1 (Chapter 2).

3.3.1.5. Design

A multiple-baseline across participants design was employed to assess if the PT intervention programme would be effective in building frequencies of blending 2-3 phonemes into words.

3.3.1.6. Procedure

Baseline. Baseline probes demonstrated current rates of responding in the absence of the PT intervention to compare to rates of responding following intervention. Stimulus sheets from the “Phonological Coding” curriculum (Freeman & Haughton, 1997) for intervention were also used in baseline. Baseline procedures were identical to those outlined in Experiment 2 (Chapter 2), with the exception of stimulus presentation. Prior to the baseline probe, the experimenter modeled a correct response with a practice word. For example, the experimenter would say “my turn, /c/, /a/, /t/, /cat/”), and the participant was given the opportunity to try and blend sounds to make a word e.g., “your turn, /d/, /o/, /g/, what word?” The experimenter provided praise for a correct response, or error correction for an incorrect response (until the student emitted the correct response). Probes were timed during a 15 second duration. The response was recorded as correct if the whole word was decoded (i.e., all phonemes accurately blended), and incorrect if the phonemes were incompletely or inaccurately blended. Other baseline procedures were identical to those in Experiment 1 (Chapter 2). Blending sounds was probed one or two times per week until a minimum of four probes were collected.

Intervention. The independent variable involved a Tier 2 PT intervention programme that included PT as an intervention to target fluency in blending 2-3 phonemes into words, and a decision-making framework to monitor progress. Intervention procedures were identical to that described in Experiment 3 (Chapter 2), however, in the blending words paradigm the experimenter presented the stimuli verbally and the participant would blend the sounds together (e.g., experimenter: “/c/, /a/, /t/”, participant: “/cat/”).

Decision Rules and Programme Modifications. Decision Rules were identical to Experiment 1 (Chapter 2). Programme changes consisted of a hierarchy of modifications, beginning with adding a visual cue and reducing the timing length to 15 seconds (Programme modification; PM 1); using the modifications from PM 1 and reducing the number of phonemes to 2 (PM 2); using the modifications from PM 2 and adding a Random Between Stimulus Prompt (RBSP; PM 3). RBSP consists of practicing items learned in error correction in previous timings, and testing random stimuli. The final modification consisted of reintroducing 3 phoneme words while keeping the other modifications present (PM 4).

3.3.1.7. Interobserver agreement (IOA). IOA data were calculated as described in Experiment 1 (Chapter 2). Exact Count-Per-Interval IOA data were collected for 19.4% of all baseline sessions demonstrating 85.7% agreement, and 18% of all intervention sessions achieving 90% agreement.

3.3.1.8. Procedural integrity. The procedural integrity checklists were identical to those used in Experiment 1 (see Appendix C for baseline checklist, and Appendix D for intervention checklist); PI data were calculated in the same manner as Experiment 1 (Chapter 2). Independent observers collected data on procedural

integrity for 33% of baseline and 25.9% of intervention sessions. Overall adherence was 100% for these baseline and intervention sessions.

3.3.2. Results and Discussion

Figure 19 depicts a multiple-baseline across participants ($n = 4$) design for the PT intervention programme in blending phonemes into words. Figure 19 shows that the remaining participants entered the baseline condition concurrently, and a minimum of four baseline probes were collected before P1 was entered into the intervention. Participant 1's performance stabilised for two days with high errors before being modified to a reduced timing length of 15 seconds, and the use of a visual cue - performance remained at the same level despite the modification (PM 1). It was observed that the participant had difficulty blending 3 phoneme words, so the programme was modified to include only 2 phoneme words in addition to the modifications above (PM 2); however while performance increased significantly for 2 phoneme words errors still remained (see Figure 19). The programme was again modified (PM 3) to include RBSP; this resulted in the student exceeding criterion levels of frequency for two phoneme words. The programme was finally modified to target 2-3 phoneme words again (PM 4; while retaining the above modifications). Participant 1 reached a reduced criterion of 12 CPM after three days intervention (see Figure 19), but only with the modifications, the intervention was then terminated in consideration of time constraints after the ninth intervention session.

Figure 19 shows that P2 was next entered into intervention, his performance remained stable for the first three days, however it remained within the x2 celeration goal for the intervention before reaching performance standards on the 4th session - despite a significant speech difficulty in his native language. Participant 3 was next

entered into the intervention. Figure 19 shows that P3 demonstrated a stable performance for three days before the programme was modified to a reduced timing length of 15 seconds, the use of a visual cue, and targeting two phoneme words (PM 1). Intervention was terminated on the 7th session seven due to time limits and in recognition of the constraints imposed by the participant's speech difficulty in his native language. Figure 19 shows that P4 was last to enter the intervention phase and responded to the intervention reaching performance standards for blending 2-3 phoneme words following six sessions.

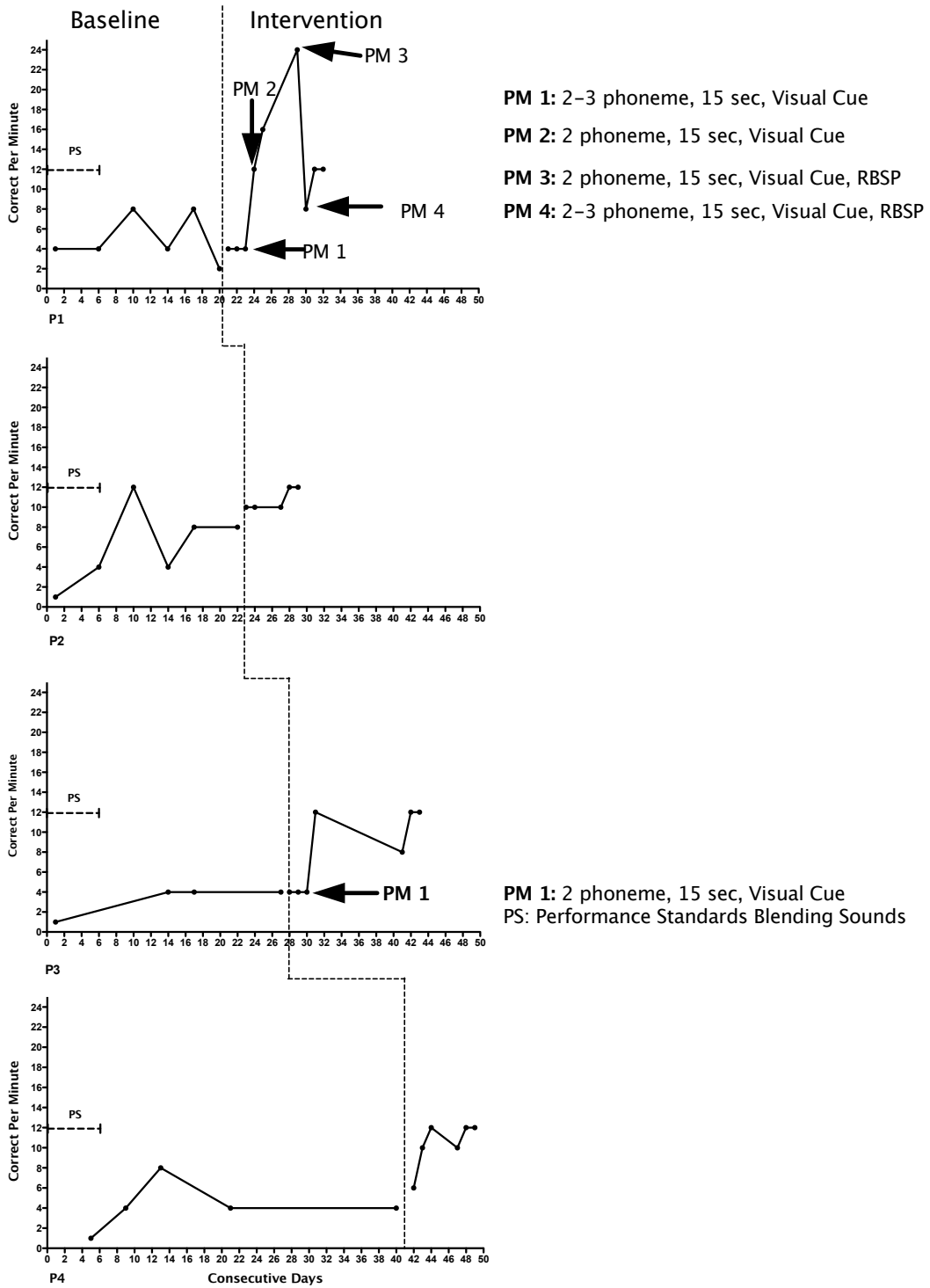


Figure 19. Experiment 7: multiple-baseline across participants design for the PT intervention in blending sounds.

3.3.2.1. *Pre-Post-Test Outcomes in Blending Sounds*

Table 9 quantifies the participants' post-test changes in standard scores for the CTOPP blending words (BW-SSG) subtest, post-test changes in standard score ratio gains (BW-SSRG), pre-post-test status changes (BW-Pre Status/BW-Post Status), and hours of intervention are also presented. Table 9 shows the average gain in standard scores for blending words was seven (range 5-10), and the average standard score ratio gains was 8.5 (range 5.9-11.8). Table 9 also shows that of the four participants selected at pre-test for intervention, one was identified as being "very poor" (P1), one as "poor" (P3), and two identified as "below average" (P2 and P4). At post-test three participants were identified as "average" (P1, P3, and P4), and one (P2) as "above average" (see Table 6).

The observed effect was achieved in approximately 0.8 hours of intervention time (range 0.6-1.3 hours). An average of seven intervention sessions were required to reach performance standards in blending phonemes into words, and daily intervention sessions lasted on average 7.5 minutes.

Table 9

Gains in CTOPP Blending Words (BW) Subtest for Participants That Received the Intervention in Blending Sounds (Experiment 7).

	BW-SSG	BW-SSRG	Total Hours Intervention	BW-Pre Status	BW-Post Status
P1	10	7.8	1.3	Very Poor	Average
P2	8	11.8	.7	Below Average	Above Average
P3	5	5.9	.8	Poor	Average
P4	5	8.3	.6	Below Average	Average

3.3.2.2. Long-Term Intervention Effects: Maintenance of Blending Sounds into Words

Figure 20 shows the maintenance data for blending sounds into words. Each panel represents a participant's performance at median baseline, end of intervention, one week, two weeks, four-to-six weeks, 10-12 weeks, and 26-28 weeks post-intervention. Due to school absenteeism over the five scheduled maintenance checks, participants show an average of four checks (range 4-5). Figure 20 shows that the average performance in blending sounds at median baseline was 5 words CPM (range 4-6), and at intervention end this had increased to 12 words CPM (range 12-12); a 2.4-fold increase in correct responding. The maintenance checks carried out at Week 1, Week 2, and 4-6 weeks, and 8-10 weeks post-intervention demonstrate average performance of 16 words CPM (range of 15-19 CPM). At the delayed maintenance check (26-28 weeks post-intervention), average performance was 16 words CPM; a 1.4-fold increase in correct responding to that achieved in the intervention (see Figure 20).

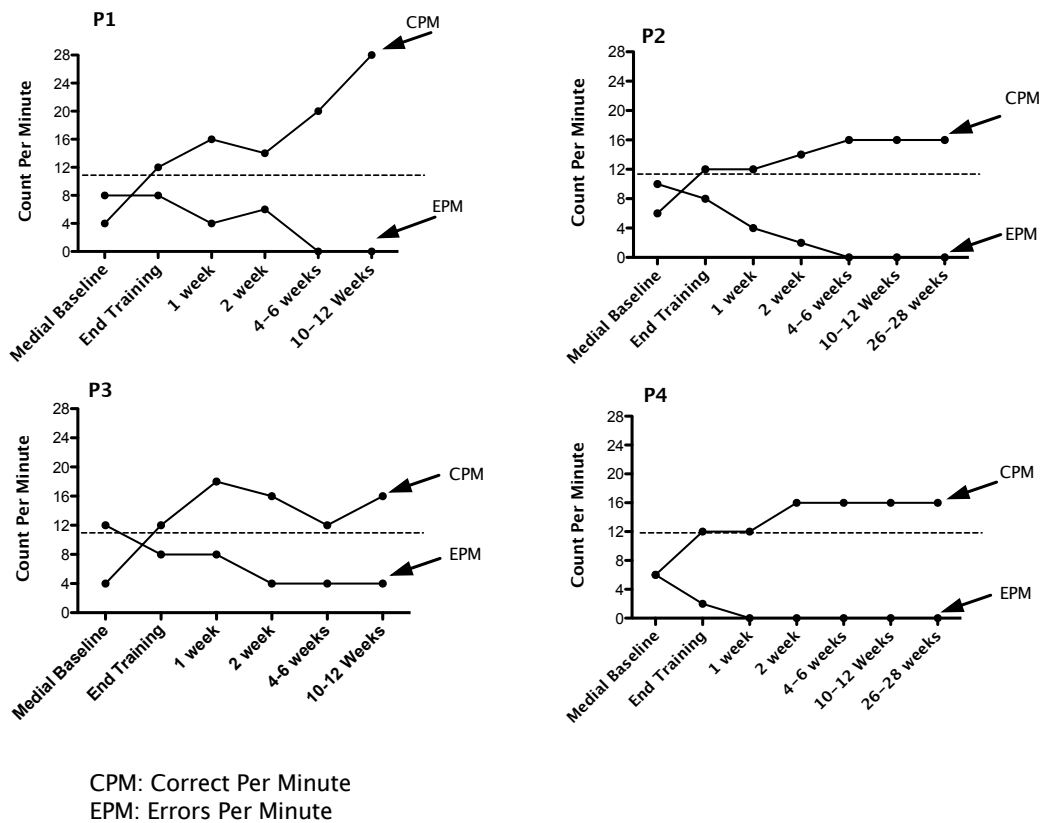


Figure 20. Experiment 6, maintenance data displaying performance in blending sounds at median baseline, intervention end and five post-intervention checks: 1 week; 2 weeks; 4-6 weeks; 10-12 weeks, and 26-28 weeks.

3.3.3. Experiment 7: Conclusions

The purpose of Experiment 7 was to implement a PT intervention programme as a Tier 2 intervention with Senior Infant students identified as at risk and to investigate its effectiveness in building frequencies in blending sounds to performance standards. Experiment 7 aimed to assess maintenance of the skill over prolonged periods of time, and to evaluate if the PT intervention resulted in participants closing the gap with average performing peers in the skill of blending sounds.

The current research demonstrated that the PT intervention programme was an effective Tier 2 intervention for building frequencies in blending sounds with students

identified as at risk of difficulties in this skill. The intervention procedures were effective in building frequencies in blending sounds to performance standards, or reduced criteria for all participants. Pre- post-test outcomes also indicated that the PT intervention programme was successful in moving the majority of at risk readers into the “average” or “above average” range of performance in blending sounds.

Frequency building to performance standards was effective in increasing the frequency and accuracy of blending sounds for all participants. All participants achieved performance standards (12-16 CPM) in words correctly blended.

Maintenance data shows that the performance levels achieved at intervention end were maintained up to 28 weeks later, and in some cases increased further.

A number of limitations were identified. Two participants required (multiple) modifications to achieve performance standards in blending sounds; the intervention stimuli were comprised of words that contained either single GPCs (2-3 phonemes) and single GPCS and MLGPCs (2-3 phonemes). These participants consistently demonstrated errors in blending 2-3 phoneme words that contained MLGPCs, and responded to the modification of reducing the number of phonemes to two. Through a number of modifications P1 was successful in blending 2-3 phoneme words. It is possible that the participants in Experiment 7 would benefit from PT intervention directly targeting decoding, but did not receive such intervention due to time constraints.

An additional limitation to this experiment involved the screening measure used to select students for the intervention, the post-test outcomes of which must be interpreted with caution. All four participants demonstrated average or above average performance at post-test, however, three continued to demonstrate difficulties in blending multiple phonemes or words consisting of MLGPC blends. The CTOPP

measure was standardised with an American population, and therefore the norms may not be applicable to the current student sample. In addition, earlier items on the subtest measure blending larger units such as syllables, therefore an average post-test status may not reflect an optimal blending repertoire for blending words at the level of the phoneme.

3.4. Chapter 3: Results and Discussion

This section of the Chapter evaluates the effects of the PT intervention programme on global outcome measures rather than the specific measures used within each experiment and described above. In addition, a measure of social validity of the research was considered an important evaluation; participant views of the intervention are presented subsequent to the generalised reading outcomes.

Experiments 5, 6, and 7 were all conducted with participants selected from three Senior Infant classrooms in an urban DEIS school. In September of the Senior Infant year all participants were administered two screening assessments (i.e., the YARC Letter Sound Knowledge subtest and the CTOPP Blending Words subtest). These screeners identified students in need of additional instruction in letter sounds and blending sounds, and a PT intervention was provided that targeted these areas specifically in Experiments 5, 6, and 7.

Some participants demonstrated low performances across both screening instruments, and as a result required the PT intervention for more than one foundational reading skill. As a result seven participants received intervention within more than one multiple-baseline design. Similar to Chapter 2, participants have been assigned a pseudo name to allow the reader to identify and evaluate participant performance across the Experiments. For example, Alf was identified as in severe difficulty in letter sounds and in blending sounds; therefore he participated in Experiments 5 and 7.

Table 10 shows participant (pseudo) names, gender, language status, age at pre-test, and participant identification numbers across experiments.

Table 10

Participant Characteristics, and Participant Identification Across Experiments 5, 6, and 7.

Name	Gender	English Language Status	Age at Pre-test	Participant ID Exp. 5 <i>Letter Sounds</i>	Participant ID Exp. 6 <i>MLGPCs</i>	Participant ID Exp. 7 <i>Blending Sounds</i>
Jen	Female	EL1	5.5 yrs.	P1		
Hana	Female	EL2	6 yrs.	P2	P2	
Jim	Male	EL1	5.6 yrs.	P3	P3	
Fay	Female	EL1	6.3 yrs.	P4		
Bev	Female	EL2	5.6 yrs.	P5	P5	
Ira	Female	EL1	6.3 yrs.	P6		
Alf	Male	LEP	6 yrs.	P7		P1
Cal	Male	EL2*	5.5 yrs.	P8	P8	P2
Dan	Male	EL2*	6.7 yrs.	P9		P3
Eve	Female	EL2	6.3 yrs.	P10		P4

EL1: English as a first language

EL2: English as a second language

EL2*: English as a second language with a significant speech difficulty in native language.

LEP: Limited English proficiency

Because seven participants took part in more than one experiment (Alf, Bev, Cal, Dan, Eve, Hana, and Jim), it is challenging to disentangle the relative contribution of the foundational reading skill targeted to overall reading development. As general reading improvement was the distal goal, it was important to measure related outcomes in word reading, letter sound fluency, and nonsense word decoding in a pre- post-test context, and as generalisation probes. Measures were administered to investigate the contribution of the PT intervention in foundational reading skills of

letter sounds and blending sounds on overall reading development; these findings are presented as generalised outcomes in reading measures.

3.4.1. Generalised Outcomes in Reading Measures for Participants in Experiments 5, 6, and 7.

Generalised outcomes in reading are presented as pre- post-test gains in real and nonsense word measures, and summaries of generalisation probes taken during baseline, intervention and post-intervention phases. Participants in Experiments 5, 6, and 7 were administered four additional measures as pre- post-tests - the YARC Early Word Reading (EWR) subtest; the CBM Letter Sound Fluency (LSF) subtest; a recoded version of the Dynamic Indicators of Basic Early Literacy (DIBELS) Nonsense Word Fluency (NWF) subtest; and a Real Word (CVC) list.

The YARC EWR and CBM LSF subtests were utilised and described in Chapter 2. Nonsense word decoding was assessed using a recoded version of the DIBELS NWF subtest. Previous research (Duhon et al., 2010) highlighted that this measure, as it is scored, functions as a letter sound fluency measure - not a decoding measure. Duhon and colleagues (2010) recoded the scoring system so that only the letter sounds correctly blended were scored. The recoded measure yields a count per minute of letter sounds correctly blended, and words correctly blended. An experimenter made probe for decoding real words was used, the words were consonant-vowel-consonant (CVC) words in size 14 font; the list yields a count per minute of words correctly blended.

Table 11 displays the outcomes in the YARC EWR subtest for participants in Experiments 5, 6, and 7. Participant results are grouped according to the intervention received e.g., single letter sounds. Table 11 depicts participants' total hours of intervention received; the total number of intervention sessions; standard score gains

(EWR-SSG); standard score ratio gains (EWR-SSRG); pre-test status (EWR Pre-Status); and, post-test status (EWR Post-Status). Table 11 shows that participants gained an average of 11.2 (range 0-27) standard score gains in early word reading; and an average standard score ratio gain of 6.9 (range 0-14.3). Overall, participants gained approximately 6.9 standard scores in word reading, per hour of intervention received across Experiments 5, 6, and 7.

Table 11 shows that at pre-test, seven participants were in the severe difficulty category (Alf, Dan, Eve, Fay, Hana, Ira, and Jim) and two in the “below average” category (Bev and Cal) in word reading. At post-test, five participants were categorised within as “average” range of scores (Bev, Cal, Eve, Hana, and Jim), however, three remained in the “severe difficulty” category (Alf, Dan, and Fay). These results suggest that the intervention was effective in moving most participants into the average range of performance in word reading.

Table 11

Outcomes in YARC Early Word Reading (EWR) Subtest Across Experiments 5, 6, and 7.

	Total Hours Intervention	Total Intervention Sessions	EWR- SSG	EWR- SSRG	EWR pre- status	EWR post- status
Participants that Received the Intervention in Single Letter GPCs						
Fay	1.7	15	0	0	Severe Difficulty	Severe Difficulty
Ira	2.4	18	9	3.8	Severe Difficulty	Average
Participants that Received the Intervention in Single and Multi Letter GPCs						
Bev	2.1	21	24	11.4	Below Average	Average
Hana	1.7	19	12	7	Severe Difficulty	Average
Jim	2.5	18	10	5.7	Severe Difficulty	Average
Participants that Received the Intervention in Single Letter GPCs and Blending Sounds into Words						
Alf	2.7	23	3	1.1	Severe Difficulty	Severe Difficulty
Cal*	2.3	21	27	11.9	Below Average	Average
Dan	2	18	1	.5	Severe Difficulty	Severe Difficulty
Eve	1	12	15	14.3	Severe Difficulty	Average
Averages	2	18	11.2	6.9		

* Cal received intervention in blending words, single letter sounds, and multi letter grapheme phoneme conversions

Table 12 displays the outcomes in the CBM LSF subtest, the recoded DIBELS NWF subtest, and the Real Word list for participants in Experiments 5, 6, and 7. Table 12 depicts participants' letter sound fluency pre-test (LSF Pre-Test) and post-test scores (LSF Post-Test); real words read correctly at pre-test (RW-WRC Pre-Test) and post-test (RW-WRC Post-Test); and, nonsense words read correctly at pre-test (NWF-WRC Pre-Test) and post-test (NWF-WRC Post-Test). Results are described as the average correct per minute (CPM) gains from pre- to post-test, and resulting fold change in average performance. At post-test, the average participant gained 42 (range 29-56) letter sounds CPM; 11.3 (range 2-18) real words segmented and blended CPM; and, nine (range 0-19) nonsense words segmented and blended CPM (see Table 12).

These pre- post-test gains represent an average 4-fold increase in letter sound fluency, an average 6-fold increase in number of real words correctly decoded, and an 18.8-fold increase in number of nonsense words correctly decoded.

Table 9 shows that two participants (Alf and Fay) failed to decode any nonsense words correctly at post-test, and made minimal gains in decoding real words, as did Dan. These three participants gained on average three (range 2-4) real words blended CPM; the remaining participants gained an average 14 CPM (range 10-18). Similarly, they gained on average 1.3 (range 0-4) real words blended CPM; the remaining participants gained an average 13 CPM (range 8-19).

Alf, Dan, and Fay also remained in the “severe difficulty” category in the YARC EWR subtest, while all other participants moved to the “average” category at post-test. Taken together, pre- post assessment outcomes converge on a similar finding: increasing frequencies in letter sounds for Fay, and, increasing frequencies in letter and blending sounds for Alf and Dan did not generalise to decoding real or

nonsense words. These students would require an intervention that targeted decoding words and programmed for generalisation. Fay was subsequently selected by the school to receive the Reading Recovery intervention. However, both Alf and Dan did not meet the inclusion criteria for this intervention.

Table 12

Outcomes in Words Read Correctly on the Recoded DIBELS Nonsense Word Fluency Subtest (NWF) and Real Word CVC Probe for Experiments 5, 6, and 7.

	LSF Pre-Test	LSF Post-Test	RW-WRC Pre-Test	RW-WRC Post-Test	NWF-WRC Pre-Test	NWF-WRC Post-Test
Participants that Received the Intervention in Single Letter GPCs						
Fay	7	63	0	2	0	0
Ira	12	55	2	12	2	10
Participants that Received the Intervention in Single and Multi Letter GPCs						
Bev	17	52	2	20	0	19
Hana	19	58	8	24	2	11
Jim	21	50	0	12	0	11
Participants that Received the Intervention in Single Letter GPCs and Blending Sounds into Words						
Alf	4	46	4	8	0	0
Cal*	11	57	0	18	0	17
Dan	17	51	0	4	0	4
Eve	21	75	4	14	0	13
Averages	14	56	2	12	.5	9.4

* Cal received intervention in blending words, single letter sounds, and multi letter grapheme phoneme conversions

These outcomes show that in approximately two hours (range 1-2.7) of intervention, over 18 (range 12-23) intervention sessions, the average participant gained 6.9 standard scores in word reading, 42 letter sounds CPM, 11.3 real words blended CPM, and nine nonsense words blended CPM.

3.4.2. Partial and Complete Decoding: Generalisation Probes Across Experiments 5, 6 and 7

A possible limitation in reporting outcomes of pre- post-test measures when using SCED is that one cannot be certain that such outcomes resulted from the intervention. In addition, as participants took part in multiple experiments it may be unclear which foundational skill occasioned any observed generalised responding to word reading. In an effort to address these limitations, two of the pre- post-test measures were also used as generalisation checks within and across experiments. Two types of generalisation checks were administered across baseline and intervention conditions in Experiments 5, 6, and 7. These checks were the recoded DIBELS NWF(*R*) subtest and the experimenter made Real Word (RW) list.

The NWF(*R*) subtest measured of the number of grapheme-phoneme conversions (GPCs) correctly blended (partial decoding), and the RW list measured the number of words read correctly (WRC; complete decoding). Reporting these two metrics is important as each add valuable information. For example, because partial decoding may emerge before full decoding (Adams, 1990), reporting the number of GPCs correctly blended may be a more sensitive measure of early decoding development. Conversely, reporting the number of words read correctly provides a more comprehensive picture of complete decoding. The outcomes of the generalisation checks are referred to as “partial decoding” and “complete decoding”.

It was expected that the increasing frequencies in foundational skills of blending sounds and letter sounds could be accompanied by concurrent gains in partial and/or complete decoding. Because many participants demonstrated at or near zero decoding in baseline, partial blending was anticipated to emerge even if complete decoding did not.

As some participants (Alf, Cal, Dan, and Eve) received intervention in both letter sounds and blending sounds, the pre- post-outcomes on the NWF(R) and RW list does not elucidate the relative contribution of each skill to nonsense and real word decoding. For this reason, observation of performance on generalisation checks administered in each experiment illustrated at what point partial and complete decoding emerged. In other words, it would help clarify which foundational skill occasioned the development of partial and complete decoding.

Figure 21 shows the generalisation outcomes for three participants (Jen, Fay, and Ira) that received the PT intervention in letter sounds only. All participants demonstrate at or near zero partial and complete decoding in the baseline condition. When the intervention is implemented Ira and Jen showed a concurrent increase in partial and complete decoding. Jen changed school during the intervention and as a result the intervention was stopped short and the overall treatment effect unclear. In 2.4 hours of intervention Ira had made gains of 10 WRC per minute, and demonstrated a steady increase in partial decoding, with the largest gains toward towards the end of the intervention (see Figure 21). Despite 1.7 hours of the intervention Fay did not exhibit even partial decoding (Fay also showed zero gains on additional measures of word reading and was subsequently selected for Reading Recovery intervention). These results suggest that increasing frequencies in letter sounds fosters partial and complete decoding for some, but not all at risk readers. In

Fay's case, the generalisation checks quickly identified her need for more intensive instruction in decoding and blending sounds.

The generalisation checks administered to Fay highlighted a limitation relating to the sensitivity of the CTOPP measure as a universal screener to identify students at risk of difficulty in blending sounds, and consequently evaluating intervention outcome effects. Using the norm-referenced cut off scores in the universal screening phase, Fay was identified as average in blending sounds and was subsequently only targeted for intervention with letter sounds. Her generalisation checks, however, demonstrated that she could not decode CVC real or nonsense words (see Figure 21). In an effort to identify all students with blending/decoding difficulties employing additional universal screening measures may provide greater sensitivity.

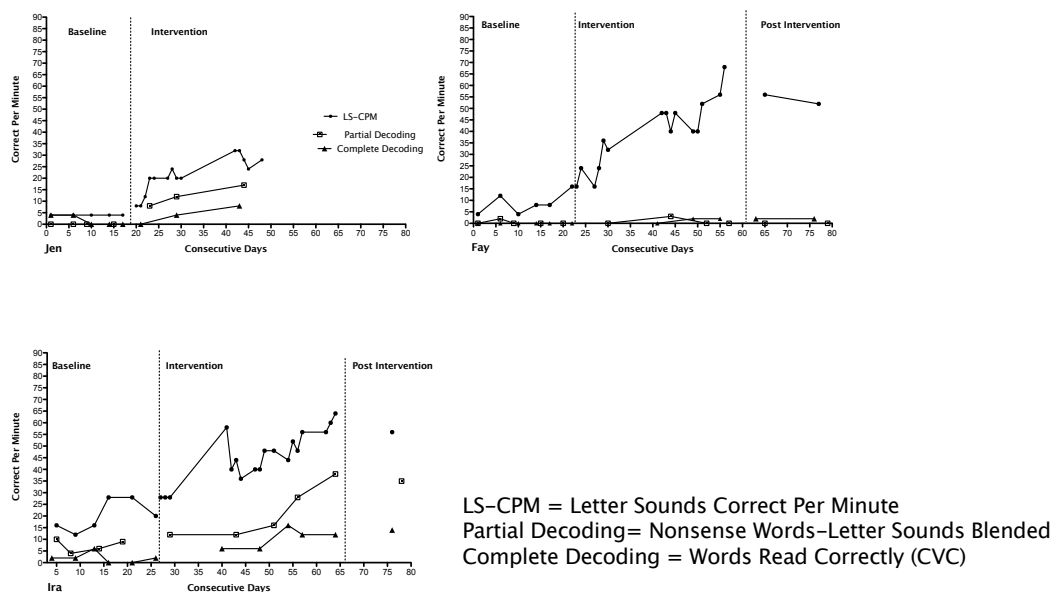


Figure 21. Generalisation outcomes for participants that receive the PT intervention for letter sounds. Generalisation checks for partial and complete decoding administered within and across baseline and intervention conditions.

Figure 22 shows the generalisation outcomes for three participants (Bev, Hana, and Jim) trained in single and multi-letter grapheme phoneme conversions (MLGPCs). All three demonstrated low complete decoding in the letter sound baseline condition. Hana showed an ascending baseline in both letter sounds and in partial decoding; this suggests that the opportunity to practice even in the absence of feedback or reinforcement led to increases in the target skill (letter sounds) and in partial decoding. When trained in letter sounds and MLGPCs for a total of 1.7 hours, however, Hana demonstrated a steady increase in partial and complete decoding, gaining 16 WRC per minute. Bev received 2.1 hours of intervention and showed a concurrent increase in partial and complete decoding, gaining 18 WRC per minute (see Figure 22). Jim received 2.5 hours of intervention and showed a concurrent increase in partial and complete decoding, gaining 12 WRC per minute. These results also suggest that building frequencies in letter sounds and MLGPCs occasions partial and complete decoding.

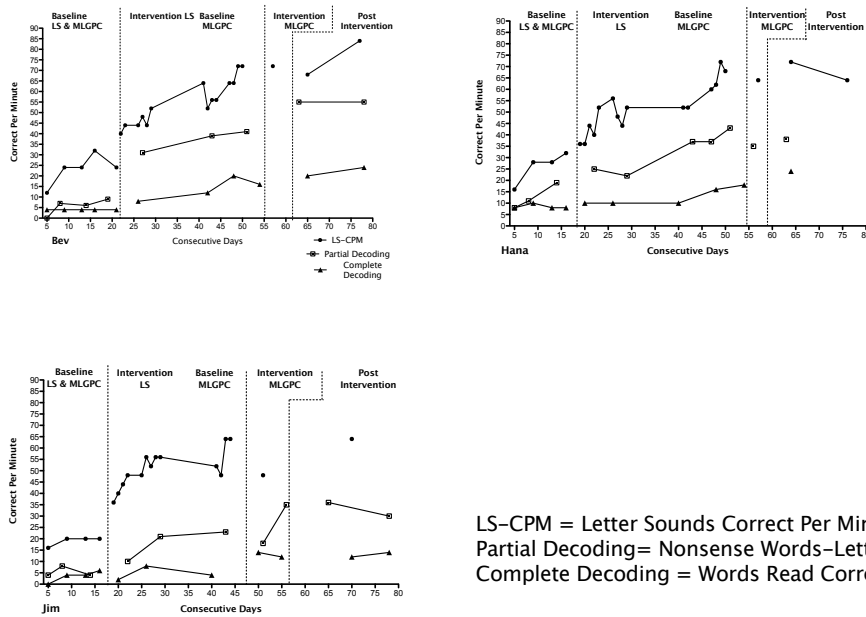


Figure 22. Generalisation outcomes for participants that receive the PT intervention for letter sounds and multi-letter grapheme phoneme conversions (MLGPCs). Generalisation checks for partial and complete decoding administered within and across baseline and intervention conditions.

Figure 23 shows the generalisation probes for participants provided with the PT intervention for both blending sounds and letter sounds. Alf, Cal, Dan, and Eve demonstrate at or near zero partial or complete decoding in baseline conditions for letter sounds and blending sounds. In the PT intervention condition for blending sounds Alf and Dan did not show generalisation to partial decoding; Cal demonstrated some increase; Eve's performance, however, did increase (see Figure 23). In the letter sounds intervention condition, all participants showed concurrent increases in partial decoding. Generalisation to complete decoding, however, did not emerge for Alf and Dan. Contrastingly, Cal demonstrated a concurrent increase in complete decoding, gaining 18 WRC correctly per minute after 2.3 hours of intervention. Similarly Eve exhibited generalisation effects to complete decoding, gaining 10 WRC per minute after one hour of intervention (see Figure 23).

These findings suggest that for the students identified as having difficulty with both blending sounds and letter sounds, the PT intervention programme was required to build frequencies in both skills before either partial or complete decoding emerged. Moreover, such participants may benefit from additional PT intervention that directly targets decoding. These results are promising for the diverse educational needs presented within classrooms; Eve, Dan, and Cal were all English language learners, and Alf had limited English proficiency. In addition, Cal and Dan were identified as having a severe speech and language delay in their native language, both attended weekly speech and language therapy sessions.

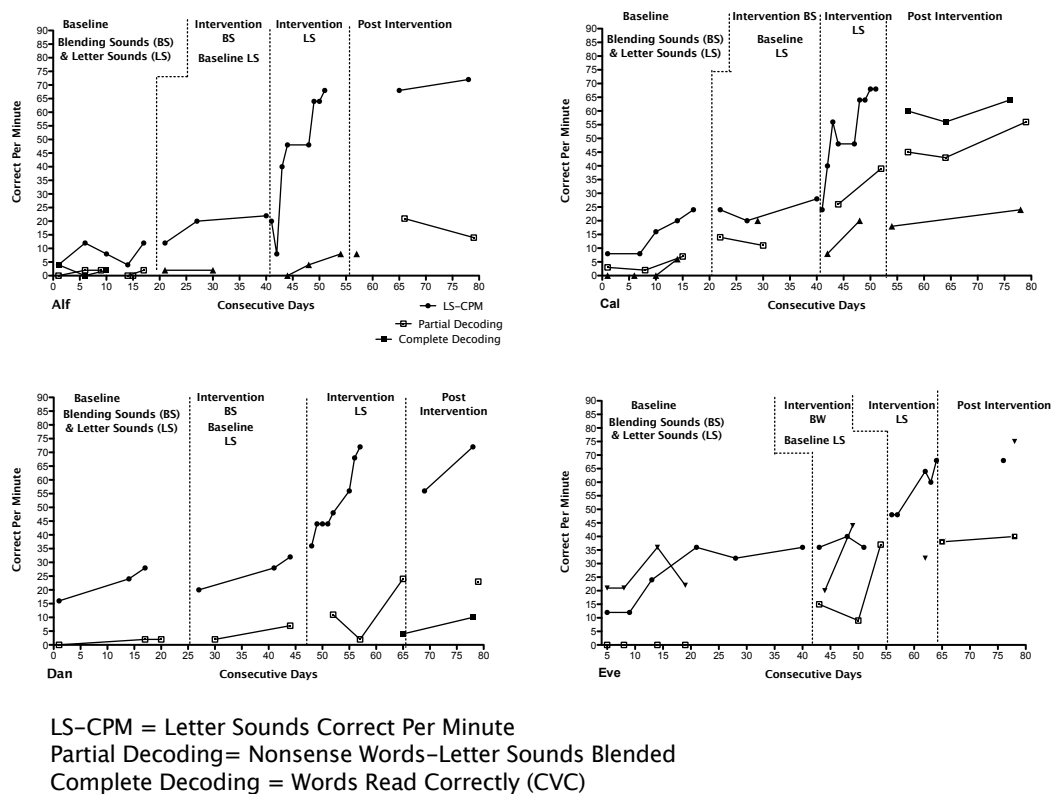


Figure 23. Generalisation outcomes for participants that receive the PT intervention for letter sounds and blending sounds. Generalisation checks for partial and complete decoding administered within and across baseline and intervention conditions.

Taken together the generalisation outcomes suggest that increasing frequencies in foundational reading skills of blending sounds and letter sounds (single

and multi) results in non-programmed generalisation to partial and complete decoding of CVC words. This represents an efficient form of instruction for at risk readers. Very short intervention durations achieved high frequencies in target skills that were maintained up to seven months post-intervention. In addition, the generalised gains to real and nonsense word reading were largely maintained up to five weeks later.

3.5. Social Validity

Social validity was central to the research question to assess the suitability of the Tier 2 intervention to existing school curricula and practices. As such, the study aimed to investigate acceptability and viability of the intervention within the applied setting. Participant, parent, and teacher views were solicited through both experimenter designed and standardised social validity and reading enjoyment measures. The parent and teacher social validity measures were identical to those used in Chapter 2 (see Appendices E and F).

An additional Likert-type questionnaire used in previous research (Musti-Rao, 2005) was administered to participants (see Appendix G). This measure contained six items/statements rated on a 3-point Likert-type scale, four yes/no items followed by open ended statements, and two open ended statements. All questionnaires included a section for additional comments.

3.5.1. Parental Interview

All eight parents strongly agreed or agreed with statements relating to the goals of intervention and outcomes of the treatment procedures. Two parents remarked on an enriched attitude towards school (“happier to be going every day”) and two commented on improvements in homework. Four parents observed on an increased interest in books (“can’t get enough of books”), and they believed that their

child was reading more since taking part in the intervention. In relation to the treatment procedures, five parents commented positively about the impact of treasure chest, and three on the stickers earned. Seven parents mentioned that their child would show off the star chart, and three stated they would praise their child's efforts in return. Five parents described a perceived increase in confidence in their child, and five commented that their child was proud of themselves for their achievement. The parents of the EL2 students with significant language difficulties remarked on their child's improved speech in both languages, and an increase in speech.

3.5.2. Participant Interview

All participants reported "liking" taking part in the intervention, when questioned on what they liked best four cited the stickers earned, four reported the treasure chest, and three using the SCC. In regard to being part of the group, earning rewards, and learning new skills, eight participants reported liking these elements "a lot", and one "a little". All participants reported that they had learned important things in the intervention, when probed specifically on the most important things learned, four commented on "letter sounds", and three on "listening". All participants reported that what they learned helped them do better work in school, and seven at home. All participants reported that they used the skills learned in the intervention, when queried exactly where they used the skills, five commented on reading, and three on in school/classroom. Seven participants wanted the programme to have lasted longer, and two did not.

Chapter 4 – Introduction

Chapter 3 reported the first screening phase (September, 2012) in the Senior Infants grade across three classrooms in an urban DEIS school. Of the 67 participants eligible for the intervention, 10 were identified with difficulties in letter sounds, and four with difficulties in blending sounds. Experiments 5 and 7 (Chapter 3) showed that the intervention was effective for participants in achievement of at or near performance standards in letter sounds and blending sounds. Generalisation gains to decoding real and nonsense words were also demonstrated for most participants.

The current chapter is concerned with the second universal screening phase in the DEIS school described in Chapter 3. The second phase of screening was conducted in January 2013 within the Senior Infants classes. During the first phase, described in Chapter 3, all students within the three Senior Infant classrooms were screened to determine difficulties in letter sounds and blending sounds. During the second phase of universal screening, one Senior Infant classroom was excluded, because a class-wide reading intervention (*Literacy Lift Off*) was to be utilised to address reading skills among students.

During this second phase of universal screening, norm-referenced measures and standard score criteria alone were deemed inadequate to demonstrate sensitivity in detecting a new pool of participants in need of supplemental instruction in foundational reading skills. It was considered unlikely that students identified in the average range of performance in September 2012 (and therefore not identified as at risk in the first phase of screening) would demonstrate below average performance in the second phase of screening. This was speculated as since September 2012 students across the three classrooms had received class-wide instruction in phonics (letter

sounds and blending the sounds together), and would therefore likely demonstrate maintained or increased performance on the screening measures.

Despite demonstrating average performance, however, participants may have fluency deficits in foundational reading skills that may hinder reading skill development. Such deficits in fluent performance may not be detected through standardised assessment alone. For this reason, the use of fluency probes in reading skills (i.e., nonsense words, CVC words, letter sounds, and blending sounds) were administered in addition to standardised measures of letter sound knowledge, in an effort to capture the differential fluency performance across the classrooms.

In the current chapter, letter sound fluency (Experiment 8) was measured during the screening phase and provided a target for intervention. MLGPCs in isolation were not targeted within this phase, due to the questionable utility of targeting MLGPCs in isolation. Specifically, when presented with words that contained the MLGPCs participants attempted to decode a word by breaking the MLGPC into single GPCs rather than recognising it as a whole linguistic unit. This indicated that although the intervention was successful in building frequencies in MLGPCs, students did not recognise them when encountered in real words, and this resulted in inaccurate decoding. To address this issue, MLGPCs were embedded in real words, and decoding was directly targeted with the PT intervention (Experiment 9). Similar to the previous phase, blending sounds (Experiment 10) was screened and targeted for intervention.

4.1. Experiment 8: Letter Sounds

In the current thesis previous experiments investigated the effects of the PT intervention programme on letter sounds in both a rural and an urban school (Experiments 1, Chapter 2 and 5, Chapter 3). In both of these experiments differing

levels of accuracy was observed across participants during the pre-testing phase. Results showed that the PT intervention programme was effective in increasing frequencies in letter sounds to performance standards, or reduced criteria, for all participants. Pre-existing accuracy levels appeared to impact the intervention time required to achieve performance standards i.e., low levels of accuracy in letter sounds required more intervention time, and in some of these cases the intervention was terminated at reduced criteria. Both of these experiments were implemented during the first universal screening phase in their respective schools (January 2011 in the rural DEIS school for Experiment 1; September 2011 in the urban DEIS school for Experiment 5).

The second phase of universal screening occurred later in the school year; due to classroom instruction in phonics it was expected that participants would be closer to total accuracy. It was anticipated that participants in Experiment 8 would possibly demonstrate differential performance patterns and outcomes than those participants in the previous two experiments targeting letter sounds. Thus this experiment potentially provides for a further analysis of intervention time required to reach performance standards in letter sounds where there is near or total accuracy.

The purpose of Experiment 8 was to implement a PT intervention programme as a Tier 2 intervention with Senior Infant students demonstrating frequency deficiencies in letter sounds. The focus of the intervention was to build frequencies in letter sounds to performance standards (70-90 CPM) at a x2 celeration. Through an evaluation of post-intervention maintenance data, Experiment 8 sought to investigate the effectiveness of the PT intervention in establishing letter sound fluency.

4.1.1. Method

4.1.1.1. Participants and Setting

Two Senior Infant classes (equivalent to kindergarten; $n=46$) within a DEIS (“Delivering Equality of Opportunity in Schools”) were selected for the universal screening phase of the current experiment. Nine students already receiving a reading intervention or learning support were excluded, resulting in a sample of 36 students (range 5.5–6.10 years, $M = 6$ years) participating in the universal screening phase.

4.1.1.2. Universal Screening Phase

Screening instruments used included the Letter Sound Knowledge (LSK) subtest of the York Assessment of Early Reading Comprehension (YARC; Hulme, et al. 2009), the Curriculum-Based Measure Letter Sound Fluency (LSF) test (Fuchs & Fuchs, 2003), and the letter sound probes described in Experiments 1 (Chapter 2) and Experiment 5 (Chapter 3). In previous experiments examining letter sounds, a participant was selected for Tier 2 intervention if he or she was identified as below average in letter sounds on the YARC LSK measure. As expected, universal screening results showed that all participants scored in the average range of performance. Average performance, however, does not reflect fluent performance (Binder, 1990).

To detect students with potential fluency deficits in letter sounds, participants’ performance frequencies in letter sounds were screened on three occasions over a two-week period to establish a baseline level. Letter sound probes were identical to those described in Experiment 1; the LSF measure was also administered on three separate occasions. Those participants who demonstrated a frequency range at 50% or lower than the LSF performance standards for letter sounds were selected for subsequent intervention. Eight participants were selected for intervention in PT

intervention with letter sounds (P1, P2, P3, P4, P5, P6, P7, and P8; age range 5.8-6.9 years, $M = 6$ years). Participants were typically developing (four females and four males); three spoke English as a first language (P1, P2, P6), and five spoke English as a second language (P3, P4, P6, P7, and P8).

4.1.1.3. Dependent Variables and Data Analysis

Dependent variables and data analysis were identical to that described in Experiment 1 (Chapter 2) with one exception. Post-intervention frequency data were collected up to 15 weeks post-intervention to investigate maintenance of intervention effects.

4.1.1.4. Materials

Stimulus and other materials were identical to those utilised in Experiment 1 (Chapter 2).

4.1.1.5. Design

A multiple-baseline across participants design was employed to assess whether the PT intervention programme would be effective in building frequencies in letter sounds to performance standards.

4.1.1.6. Procedure

Baseline. Baseline procedures were identical to those described in Experiment 1 (Chapter 2).

Intervention. The independent variable involved a Tier 2 PT intervention programme that included PT as an intervention to target fluency in letter sounds, and a decision-making framework to monitor progress. Intervention procedures were identical to those employed in Experiment 1 (Chapter 2), with the exception that discrete trials were not used with any participants.

Decision Rules and Programme Modifications. Decision rules and programme

modifications were identical to those employed in Experiment 1.

4.1.1.7 Interobserver agreement (IOA). IOA data were calculated as described in Experiment 1 (Chapter 2). Exact Count-per-Interval IOA data were collected for 70% of all baseline sessions demonstrating 90% agreement, and 29% of all intervention sessions achieving 94% agreement. Dividing the number of intervals where there is agreement by the total number of intervals and multiplying by 100 calculates this index

4.1.1.8. Procedural integrity. The procedural integrity checklists were identical to those used in Experiment 1 (see Appendix C for baseline checklist, and Appendix D for intervention checklist); PI data were calculated in the same manner as Experiment 1. Independent observers collected data on procedural integrity for 31.8% of baseline sessions and for 35.7% of intervention sessions evidencing 100% adherence.

4.1.2. Results and Discussion

Figure 24 depicts a multiple-baseline across participants ($n = 8$) design for the PT intervention programme in letter sounds. Figure 24 shows that all participants entered the baseline condition at the same time, and a minimum of five baseline probes were collected before P1 and P2 were entered into the intervention phase. Both participants responded to the intervention, gaining 20 (P1), and 28 letter sounds CPM by the 3rd session. Subsequently, P3 and P4 were entered into the intervention phase and both responded to the intervention, gaining 28 and 20 letter sounds CPM respectively, by the 3rd session. Participants 5 and 6 next entered intervention, and both responded to the intervention, gaining 16 and 29 letter sounds CPM respectively, by the 3rd session. Figure 24 shows that P7 was next entered into the intervention phase, and responded, gaining 29 letter sounds CPM by the 3rd session. Finally, P8

was entered into the intervention and responded, gaining 35 letter sounds CPM by the 3rd session (see Figure 24). All participants maintained or exceeded a x2 celeration for the duration of the intervention before achieving performance standards (M = 82; range 72-88 CPM) after approximately seven intervention sessions (range 4-12).

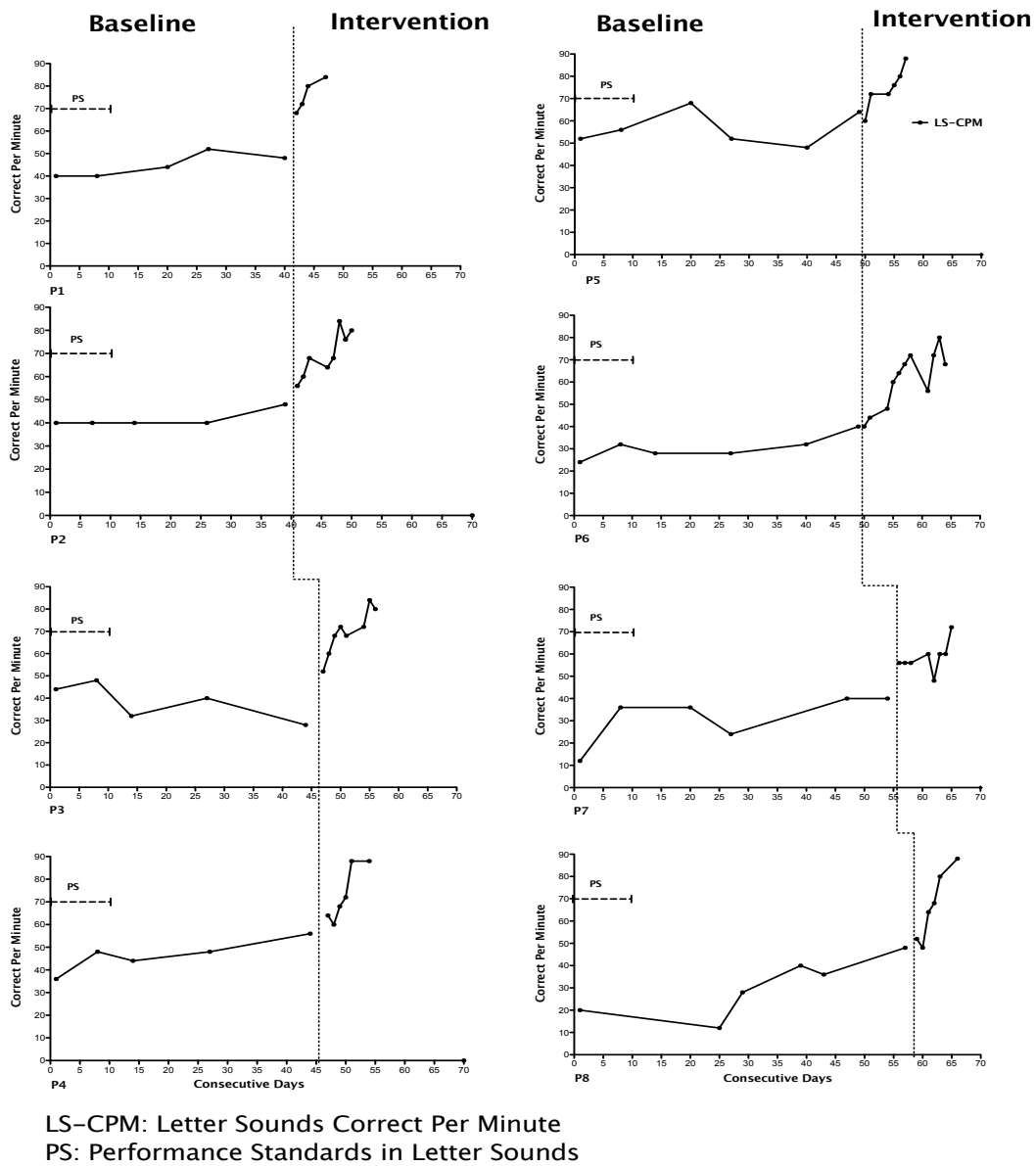


Figure 24. Experiment 8: multiple-baseline across participants design for the PT intervention in letter sounds.

4.1.2.1. Pre- Post-Test Outcomes in Letter Sounds

Table 13 presents participants' post-test standardized score gains (SSG), and standard score ratio gains (SSRG) in letter sounds on the YARC letter sound knowledge subtest (LSK-SSG; LSK-SSRG); total hours of intervention received; rate gains (RG) of correct letter sounds on the CBM letter sound fluency subtest (LSF-RG); and rate ratio gains (RRG) on the CBM LSF subtest. Table 13 shows the average standard score gain for letter sounds was 13 (range 0-21), and the average standard score ratio gain for letter sounds was 14 (range 0-41.7). Table 8 also shows the average rate gains in letter sounds was 51 letter sounds CPM (range 27-77), and the average rate-ratio gains were 52 letter sounds CPM (range 19-106; see Table 13). The observed effect was achieved with the average participant in approximately 1.2 hours of intervention time (range 0.5-1.5 hours). Daily intervention sessions on average lasted 5.6 minutes, across an average of seven (range 4-12) days of intervention to reach performance standards in letter sounds.

Table 13

Outcomes in YARC Letter Sound Knowledge (LSK) and CBM Letter Sound Fluency (LSF) Subtests for Participants That Received the PT Intervention in Letter Sounds (Experiment 8).

	LSK-SSG	LSK-SSRG	Total Hours Intervention	LSF- RG	LSF-RRG
P1	21	15	1.4	27	19.3
P2	21	14	1.5	48	32
P3	6	6	1	77	77
P4	7	7	1	43	43
P5	20	41.7	.5	51	106
P6	16	14.5	1.1	62	56
P7	0	0	.7	44	36.7
P8	--	--	.7	58	48.3

4.1.2.2. Long-Term Intervention Effects: Maintenance of Letter Sounds

Figure 25 shows maintenance data collected for letter sounds. Each panel represents a participant's performance at median baseline, end of intervention, 1 week, 2 weeks, and 14-15 weeks post-intervention. Due to school absenteeism over the three scheduled maintenance checks, participants show an average of two checks (range 1-3). Figure 25 shows that the average performance in letter sounds at median baseline was 42 CPM (range 31-56), and at intervention end this had increased to 80 CPM (range 68-88); a 1.9-fold increase in correct responding. The maintenance checks implemented at Week 1, Week 2, and 18-20 weeks post-intervention demonstrated an average performance of 61 CPM (range of 44-84); a 0.9-fold decrease in correct responding. Despite this decrease in frequencies, five participants

remained within the performance standard range for letter sounds across the maintenance checks (P1, P2, P4, P5, and P8). Participants 3 and 6 demonstrated a decrease in frequencies to 60 letter sounds CPM three months post-intervention, and only the first maintenance check was measured for P7 due to absence from school (see Figure 25).

These findings suggest that the performance standards achieved by participants in the intervention phase were maintained with most participants until the end of the school year (14-15 weeks post-intervention).

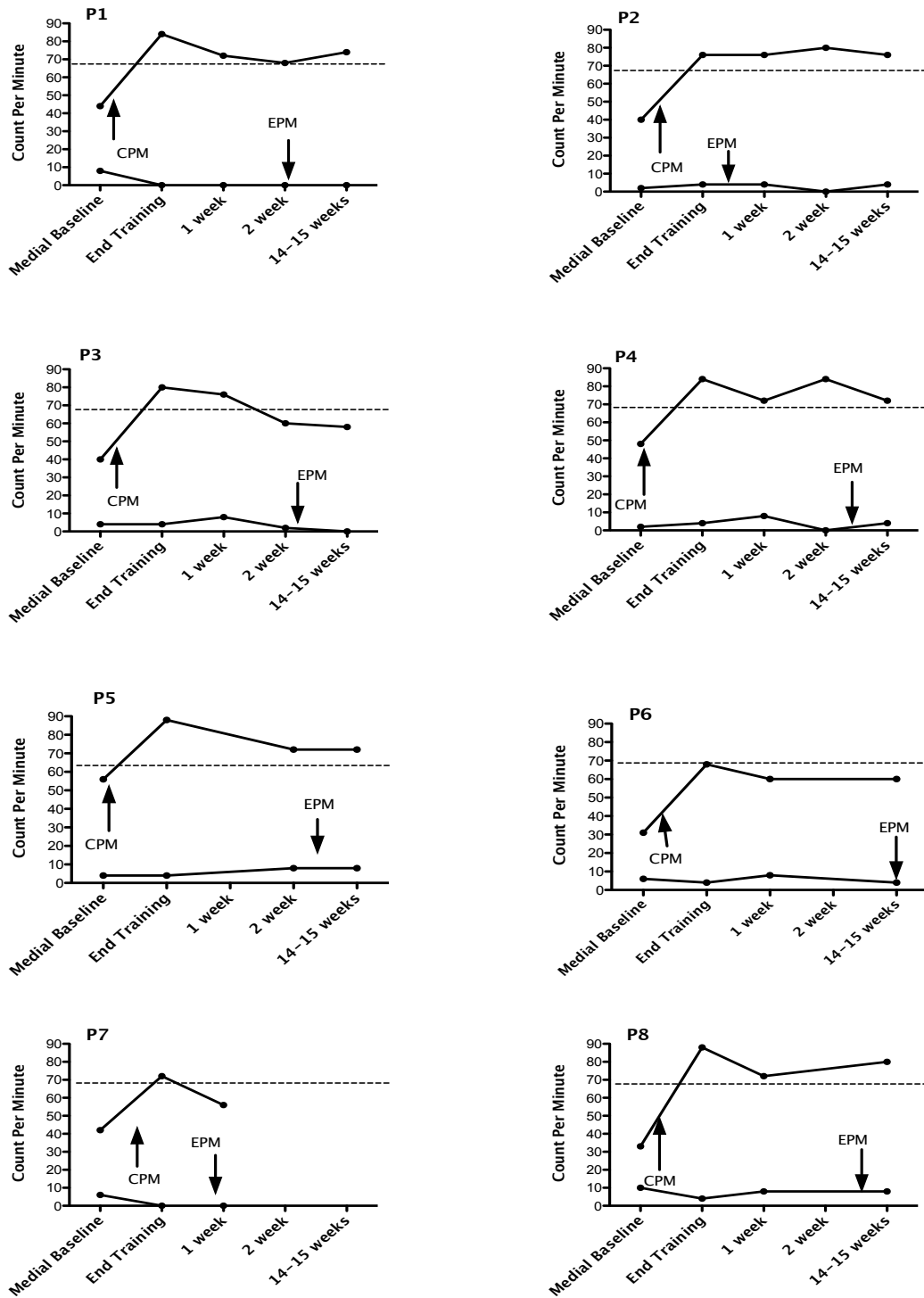


Figure 25. Experiment 8, maintenance data displaying performance in letter sounds at median baseline, intervention end and three post-intervention checks: 1 week; 2 weeks, and 14-15 weeks.

4.1.3. Experiment 8: Conclusions

The purpose of Experiment 8 was to apply a PT intervention programme as a Tier 2 intervention with Senior Infant students demonstrating frequency deficiencies in letter sounds. The focus of the intervention was to build frequencies in letter sounds to performance standards at a x2 celeration, and to evaluate post-intervention maintenance data.

The combined findings of the multiple-baseline design, the pre- post-test outcomes, and maintenance effects suggest that the PT intervention programme was both an effective and efficient form of intervention to target performance standards and fluency in letter sounds. In an average of 1.2 hours of intervention time (range 0.5-1.5 hours) over approximately seven sessions (range 4-12), participants achieved performance standards in letter sounds ($M = 82$; range 72-88 CPM). Participants also on average achieved rate gains of 51 letter sounds CPM (range (27-77) on the Curriculum-Based Measure letter sound fluency subtest, and an average standard score ratio gain of 14 (range 0-41.7) in the YARC letter sound knowledge subtest. Finally, these intervention outcomes were largely maintained 14-15 weeks post-intervention.

A number of limitations may be considered within Experiment 8. The add/subtract graphs create the impression of ascending baselines for Participants 4 and 8; this issue has been addressed in detail in the limitations of Experiment 2 (Chapter 2). Specifically the use of the SCC permitted precise quantification of increases in both the baseline and intervention conditions. The increase in trend observed in the baseline was not clinically significant on the level of the individual (i.e., the increase in celeration in the intervention was far greater than that observed in the baseline).

4.2. Experiment 9: Decoding (Segmenting and Blending Words)

In Experiment 6 (Chapter 3) MLGPCs were targeted using the PT intervention programme. While the procedures used were effective in increasing frequencies in MLGPCs, the utility of targeting them in isolation was questioned. Specifically, participants did not recognise MLGPCs as whole linguistic units when embedded in a word, and decoded them as single GPCs resulting in inaccurate decoding. Previous research has targeted fluency intervention in decoding words comprised of specific MLGPCs. For example, Martens, Werder, Hier, and Koenig (2013) provided fluency intervention for three MLGPCs (*aw*, *oi*, and *au*) embedded in real words with three 2nd grade students.

The authors used a series of multiple probe across MLGPC designs to evaluate generalised effects on reading of untrained words in lists, trained and untrained words in target passages, and novel words in generalisation passages. In terms of reading accuracy, the authors demonstrated that all three students showed generalised increases from trained to untrained words on lists, and in both target and generalisation passages. Findings related to oral reading fluency were mixed across participants and MLGPCs targeted. Martens et al. (2013) concluded that intervention targeting decoding words as an isolated task using modeling and feedback, and subsequent practice and reinforcement for generalisation may be a promising method to foster generalised reading skills.

Experiment 9 extends the findings of Martens et al. (2013) by targeting a different set of MLGPCs (/oo/, /ay/, and /ch/), with younger participants (i.e., kindergarten equivalent, as opposed to 2nd grade students), and used a multiple-baseline across participants design (as opposed to a multiple probe across MLGPCs

targeted). In addition, Martens and colleagues (2013) defined fluency as performance exceeding 50% of their known performance levels on another word list. The current thesis used the a performance standard that related to the number of words read correctly within the frequency building trials.

4.2.1. Methods

The purpose of Experiment 9 was to investigate if SAFMEDS could be used to build frequencies in decoding (2-4 phoneme) words with participants who had previously received the PT intervention programme for single letter sounds.

4.2.1.1. Participants and Setting

Participants and setting were identical to that described in Experiment 8; four participants (P1, P2, P3, and P4) who had completed the PT intervention for letter sounds were selected for intervention in decoding words. Participants were three males and one female (age range 6-6.3 years, $M = 6$ years). One participant spoke English as a first language (EL1) and three spoke English as a second language (EL2).

4.2.1.2. Dependent Variable and Data Analysis

Dependent variables and data analysis procedures employed were identical to those reported in Experiment 7 (Chapter 3). Data analysis pertained to words correctly blended by the end of the intervention. Due to time constraints maintenance data were not collected for this experiment.

4.2.1.3. Materials

Stimulus materials were SAFMEDS that included commercially available word cards (15 x 7cm; Miskin, 2011). Each word (approx. size 72 font) consisted of 2-4 phonemes, represented by single GPCs and MLGPCs. The specific MLGPCs selected for intervention were high frequency MLGPCs (Carnine et al., 1997; Solity & Vousden, 2009) that have already been targeted in general classroom instruction.

The MLGPCs were targeted in sets of 3-4, embedded in 2-5 different target words (see Appendix H). This resulted in four SAFMEDS decoding sets of 36-39 words. Set 1 targeted /oo/, /ay/, and /ch/; Set 2 targeted /ee/, /th/, /ou/, and /sh/; Set 3 targeted /ar/, /ir/, and /ew/; and, Set 4 targeted /igh/, /aw/, and /or/. Due to time constraints at the end of the school year, only SAFMEDS Set 1 was targeted during the intervention phase.

Other materials were identical to those used in Experiment 1.

4.2.1.4. Design

A multiple-baseline design across participants, with a multiple probe element across target skills, was employed to assess the effectiveness of the PT intervention programme on building frequencies in decoding words for SAFMEDS Set 1.

4.2.1.5. Procedure

Baseline. Baseline probes demonstrated current rates of responding in the absence of the PT intervention to compare to rates of responding following intervention. Baseline stimuli included a SAFMEDS word set that consisted of one example of each word, from SAFMEDS Sets 1 and 2, randomly presented to control for practice effects. Prior to the baseline probe, the experimenter modeled the correct response with a practice word from Set 3 (for example, the experimenter would say “my turn, /c/, /a/, /t/, /cat/”), and the participant was then given the opportunity to attempt decoding on another practice word from Set 4. The experimenter provided praise for a correct response, or error correction for an incorrect response (until the student emitted the correct response). Probes were timed during a 15 second duration. During the probe, each word-card was presented individually to the participant, who would attempt to identify each phoneme (segment), and then blend the phonemes back together to produce the word. The response was recorded as correct if the whole

word was decoded (i.e., all phonemes accurately blended), and incorrect if the phonemes were incompletely or inaccurately blended. Rates of responding were recorded on a daily data collection sheet, on the DPM SCC and on the TOY. Other baseline procedures were identical to those described in Experiment 1 (Chapter 2). Decoding was probed one or two times per week until a minimum of six probes were complete.

Intervention. The independent variable involved a Tier 2 PT intervention programme that included PT as an intervention to target fluency in decoding words, and a decision-making framework to monitor progress. Intervention procedures were identical to those described in Experiment 1, with the exception of the stimuli presentation. During timing trials in the intervention condition, based on the participant response to the word card it was placed in a ‘correct’ or an ‘incorrect’ pile. In this way performance feedback was immediate, but error correction was subsequent to the timing trial. Participants received the one-to-one intervention five days per week, daily sessions lasted on average five minutes, and the average total intervention duration was .5 hours.

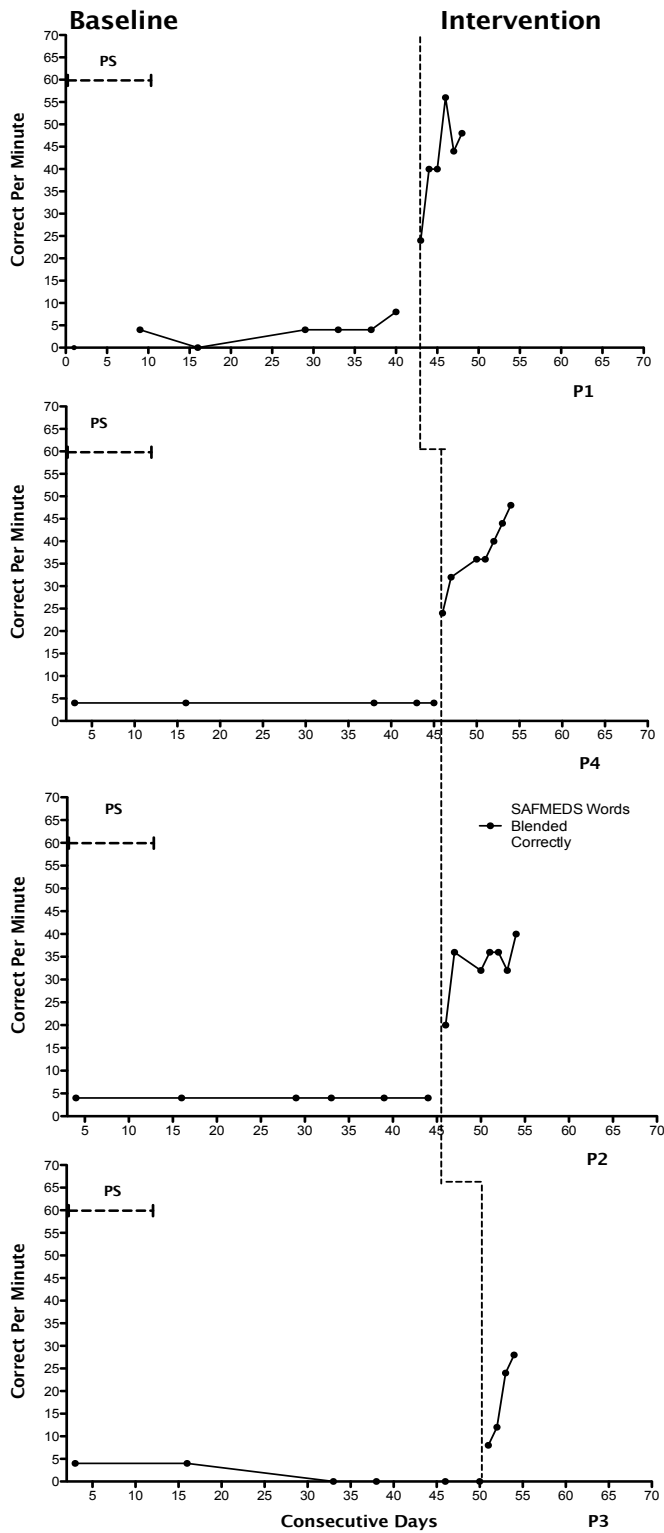
Decision Rules and Programme Modifications. Decision rules were identical to those employed in Experiment 1 (Chapter 3). Programme modifications involved two steps: Programme Modification 1 (PM1) involved reducing the number of unique words in the SAFMEDS set by 50%, and Programme Modification 2 (PM2) involved reducing the number of unique words in the SAFMEDS set by a further 20%.

4.2.1.6. Interobserver agreement (IOA). IOA data were calculated as described in Experiment 1 (Chapter 2). Exact Count-Per-Interval IOA data were collected for 19% of all baseline sessions and 36% of all intervention sessions demonstrating 100% agreement.

4.2.1.7. Procedural integrity. The procedural integrity checklists were identical to those used in Experiment 1 (see Appendix C for baseline checklist, and Appendix D for intervention checklist); PI data were calculated in the same manner as Experiment 1 (Chapter 2). Independent observers collected data on procedural integrity for 36% of baseline and 62.5% of intervention sessions. Overall adherence was 100% for these baseline and intervention sessions.

4.2.2. Results

Figure 26 depicts a multiple-baseline across participants ($n = 4$) design for the PT intervention programme in decoding (segmenting and blending) words. Figure 26 shows that three participants entered the baseline phase at the same time (P2, P3, and P4), while the remaining participant (P1) entered a week later (due to absenteeism). A minimum of five baseline probes were collected before P1 was entered into the intervention phase; he responded and gained 46 words blended CPM before intervention was terminated on the 5th session. Participants 4 and 2 were next entered and both responded to the intervention; P4 gained 36 words blended CPM, and P2 gained 43 words blended CPM before intervention was terminated on the 7th session. Finally, P3 was entered into the intervention, she responded and gained 46 words blended CPM before intervention was terminated on the 4th session (see Figure 26). All participants maintained or exceeded a x2 celeration for the duration of the intervention, and on average gained 38 words correctly blended after approximately six intervention sessions.



PS: Performance Standards in SAFMEDS

Figure 26. Experiment 9: multiple-baseline across participants design for the PT intervention in decoding words.

4.2.3. Experiment 9: Conclusions

The purpose of Experiment 9 was to investigate if SAFMEDS could be used to build frequencies in decoding (2-4 phoneme) words that comprised of single and multi-letter GPCs with participants who had previously received the PT intervention programme for single letter sounds. Mastery of MLGPCs is essential, as single GPCs permit decoding of only CVC words, thus limiting the students decoding repertoire. High frequency MLGPCs (e.g., Carnine et al., 1997) were targeted. In addition, these MLGPCs had been taught during regular classroom instruction, therefore all participants had previous exposure to such MLGPCs. Due to time constraints, only one SAFMEDS set was targeted- Set 1, containing the phonemes /oo/, /ay/, and /ch/. Results showed consistent performance patterns within baseline and intervention conditions, and three strong replications of effect in the intervention condition. This suggests that pre-intervention, participants could not decode whole words containing the target phonemes, and that the intervention resulted in accurate and speedy decoding of words containing the target phonemes.

This finding suggests that although decoding is a desired reading outcome, it may not emerge naturally. The strong replications of effect suggest that the PT intervention was a potent means to increase frequencies in decoding whole words. A caveat on the criteria for correct responding is noteworthy; a correct response involved all of the phonemes blended to make the entire word. Decoding behavior, however, can be observed as absent, partial, or complete. In partial decoding, some (but not all) of the phonemes are blended, to make part of the word. It was observed in the baseline condition that participants were partially blending (i.e., correctly blending two of the three phonemes), but such successive approximations were not recorded as criteria for correct responding required complete blending. This may

explain the near zero levels of correct responding in the baseline condition.

A number of limitations within Experiment 9 may be considered. For example, the four participants selected for intervention were selected based on availability i.e., they had completed the PT intervention for letter sounds and were available to take part in a short intervention (approximately six intervention sessions). Therefore, it remains unclear if the same outcomes would be achieved with students identified as “at risk” on a measure of decoding. In addition, as they had all previously received intervention in letter sounds, this may in part explain immediate and large magnitudes of the intervention effect, observed in the current experiment. Furthermore, only one set of SAFMEDS was targeted (Set 1). This may be considered a relatively small intervention stimulus set, although it remains unclear if the same outcomes would be obtained with additional SAFMEDS (Sets 2-4).

Due to time constraints, maintenance data were not collected for this experiment, thus it is uncertain if intervention effects were maintained over time. Another shortcoming of the current experiment may involve the lack of measurement of generalisation to untrained words that contained the same phonemes targeted in the intervention set. A measure of reading trained words in the context of a sentence was not administered and this may be considered an important indicator of generalization effects. Overall therefore, it is also unclear whether the current intervention effects generalise to novel words and/or sentence reading.

Finally, because a response was recorded as correct *only* if the whole word was blended, partial blending was not scored as correct. Using a measure of complete blending only resulted in a strict criterion for a correct response.

4.3. Experiment 10: Blending Sounds into Words

The purpose of Experiment 10 was to implement a PT intervention

programme as a Tier 2 intervention with Senior Infant readers, and to investigate its effectiveness in building frequencies in blending sounds to performance standards (10-15; Freeman & Haughton, 1997) at a learning rate of x2 celeration.

4.3.1. Method

4.3.1.1. Participants and Setting

Participants and setting were identical to those described in Experiment 7 (Chapter 3).

4.3.1.2. Universal Screening Phase

Participants' performance frequencies in blending sounds were probed three times over a 2-week period to establish a baseline level. Those participants who demonstrated a frequency range at 50% or lower than the performance standard in blending sounds were selected for subsequent intervention. Two participants demonstrated low frequencies (P1 and P2; age range 5.8 years, M = 6 years). Both participants were typically developing (two females); both spoke English as a second language (EL2).

4.3.1.3. Dependent Variables and Data Analysis.

Dependent variables and data analysis procedures employed were identical to those reported in Experiment 1 (Chapter 2) with two exceptions. First, in the multiple-baseline design, data were evaluated with respect to the students' response to the intervention, and performance at the end of the intervention. Second, maintenance and post-test outcomes were available for one participant and are therefore not reported.

4.3.1.4. Materials

Stimulus and other materials were identical to those described in Experiment 7 (Chapter 3).

4.3.1.5. Design

A multiple-baseline across participants design was employed to assess if the PT intervention programme would be effective in building frequencies in blending 2-3 phonemes into words.

4.3.1.6. Procedure

Baseline. Baseline procedures were identical to those in Experiment 7 (Chapter 3).

Intervention. The independent variable involved a Tier 2 PT intervention programme that included PT as an intervention to target fluency in blending sounds, and a decision-making framework to monitor progress. Intervention procedures were identical to those described in Experiment 7. Participants received the one-to-one intervention five days per week, daily sessions lasted on average 6.3 minutes, and the average total intervention duration was .8 hours.

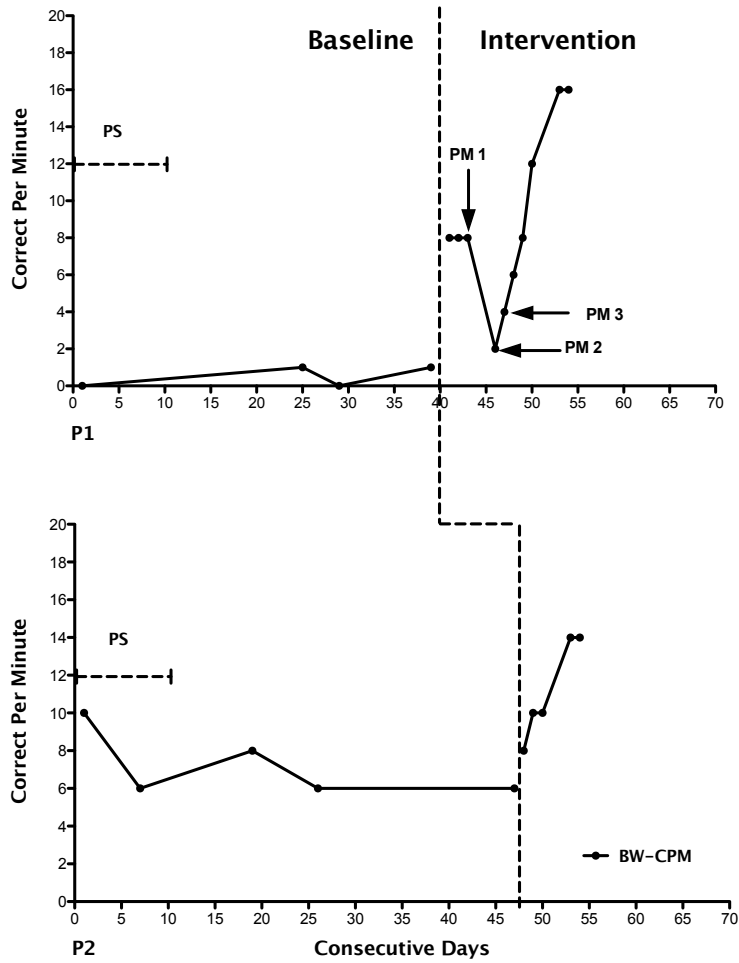
Decision Rules and Programme Modifications. Decision rules were identical to those employed in Experiment 7. The same hierarchy of programme changes was used, with one exception; instead of reducing the number of phonemes to two, CVC words were targeted to keep the number of phonemes at three. This change was made in recognition that some participants consistently made errors blending words that contained MLGPCs. Rather than reducing the number of phonemes in the word, CVC words were used as the 3rd programme modification (PM3) to keep the number of phonemes at three, yet reduce the complexity of the task.

4.3.1.7. Interobserver agreement (IOA). IOA data were calculated as described in Experiment 1 (Chapter 2). Exact Count-Per-Interval IOA data were collected for 50% of all baseline sessions demonstrating 90% agreement, and 38.7% of all intervention sessions achieving 90% agreement.

4.3.1.8. Procedural integrity. The procedural integrity checklists were identical to those used in Experiment 1 (see Appendix C for baseline checklist, and Appendix D for intervention checklist); PI data were calculated in the same manner as Experiment 1 (Chapter 2). Independent observers collected data on procedural integrity for 44% of baseline and 20% of intervention sessions. Overall adherence was 100% for baseline and intervention sessions.

4.3.2. Results

Figure 27 depicts a multiple-baseline across participants ($n = 2$) design for the PT intervention in blending phonemes into words. Figure 27 shows that both participants entered the baseline condition concurrently, and a minimum of four baseline probes were collected before P1 was entered into the intervention. Participant 1 responded to the intervention, and then experienced goal failure on the 3rd session, her programme was modified (PM1 - use of visual cue) but she did not respond to the modifications. The programme was modified again (PM2 – reducing the duration of the timing trials from 30 to 15 seconds), however P1 continued to have difficulty blending words that contained MLGPCs. Figure 27 shows that P1's programme was further modified (PM3 – targeting CVC words with single GPCs). Participant 1 responded to this modification and achieved performance standards of 16 words CPM by the 9th session. Participant 2 was the second participant to enter the intervention and responded, achieving performance standards (14 CPM) by the 4th session (see Figure 27).



BW-CPM: Blended Words – Corrects Per Minute
 PM 1: Programme modification 1 – Use of coloured blocks as visual cues
 PM 2: Programme modification 2 – Reduced duration of timing length 15 seconds
 PM 3: Programme modification 3– Targeted CVC words
 PS: Performance Standards Blending Sounds

Figure 27. Experiment 10: multiple-baseline across participants design for the PT intervention in blending sounds.

4.3.3. Experiment 10: Conclusions

The purpose of the Experiment 10 was to implement a PT intervention programme as a Tier 2 intervention with Senior Infant readers, and to investigate its effectiveness in building frequencies in blending sounds to performance standards. The current research demonstrated that PT intervention programme was effective for building frequencies in blending sounds to performance standards (12-16 words CPM) for both participants. Furthermore, the programme modifications used with P1 resulted in a response to the intervention.

A main limitation to this experiment involved the number of participants within the study. Typically, three replications of effect are necessary to demonstrate experimental control, however, the current experiment contained only two. This is an artifact of the universal screening outcomes (i.e., only two participants identified with difficulties in blending sounds), rather than an oversight. In addition, maintenance data within this experiment were limited with only two data points available limiting an evaluation of long-term intervention effects.

Similar to performance patterns observed in Experiment 7 (Chapter 3), multiple modifications were necessary for P1 to respond to the intervention. In this experiment a deviation of programme modifications was implemented in response to a limitation highlighted in the previous blending sounds experiment (Chapter 3). In Experiment 7 a programme modification involved reducing the number of phonemes in all target words to two, however, it was observed that participants were demonstrating difficulties blending three phoneme words that contained MLGPCs. For that reason, the current experiment targeted CVC words as a programme modification, keeping the number of phonemes at three, and omitting MLGPCs from target words. This modification was effective for Participant 1.

4.4. Results and Discussion

This section of the Chapter evaluates the effects of the PT intervention programme on global outcome measures rather than the specific measures used within each experiment and described above. As development of overall reading skills was the distal goal, it was therefore important to measure related outcomes in word reading, letter sound fluency, and nonsense word decoding in a pre- post-test context. To this end, all participants were administered three additional assessments (DIBELS, YARC and WIAT-II) as pre- post-tests to investigate the contribution of the PT intervention programme on foundational reading skills on overall reading development. These findings are presented as generalised outcomes in reading measures. In addition, a measure of social validity of the research was considered an important evaluation; participant views of the intervention are presented subsequent to the generalised reading outcomes.

Similar to previous chapters, participants were assigned a pseudo name to allow the reader to identify and evaluate participant performance across the Experiments. Table 14 shows participant (pseudo) names, gender, language status, age at pre-test, and participant identification numbers across the experiments within the current chapter.

Table 14

Participant Characteristics, and Participant Identification Across Experiments 8, 9, and 10.

Name	Gender	English Language Status	Age at Pre-test	Participant ID Exp. 8 <i>Letter Sounds</i>	Participant ID Exp. 9 <i>Decoding</i>	Participant ID Exp. 10 <i>Blending Sounds</i>
Guy	Male	EL1	6	P1	P1	
Joe	Male	EL1	6.3	P2	P2	
Dee	Female	EL2	6	P3	P4	
Eli	Male	EL2	6	P4	P3	
Bob	Male	EL2	6.9	P5		
Flo	Female	EL1	5.8	P6		
Ivy	Female	EL2	5.8	P7		P2
Kya	Female	EL2	5.8	P8		P1

EL1: English as a first language

EL2: English as a second language

4.4.1. Generalised Outcomes in Reading Measures for Participants in Experiments 8, 9, and 10.

Overall outcomes in reading are presented as pre- post-test gains in real and nonsense word measures, and summaries of generalisation checks implemented during baseline and intervention phases. Additional pre- post-measures were the recoded DIBELS NWF subtest, and the Real Word list (experimenter made), the YARC Early Word Reading (EWR) subtest, and the PseudoWord Decoding subtest from the Wechsler Individual Achievement Test Second UK addition (WIAT-II, Wechsler, 2005). The YARC EWR and CBM LSF subtests are described in detail in Chapter 2; the recoded DIBELS NWF, and Real Word list are described in Chapter 3.

The Pseudoword Decoding Subtest (PWD, WIAT-II) assesses the ability to apply phonetic decoding skills and contains 55 items of increasing difficulty. This subtest is reported to have a test-retest coefficient of 0.98 for 6 year olds, and concurrent criterion related validity of 0.75.

Table 15 displays the outcomes in the YARC Early Word Reading and WIAT-II PseudoWord Decoding subtests for participants in Experiments 8, 9, and 10. Table 15 depicts participants' total hours of intervention received; the total number of intervention sessions; EWR standard score gains (EWR-SSG); and EWR standard score ratio gains (EWR-SSRG); PWD standard score gains (PWD-SSG), and PWD standard score ratio gains (PWD-SSRG). Table 15 shows that participants gained an average of 2.9 (range -10-9) standard score gains in early word reading, and an average standard score ratio gain of 2.4 (range -10-6.7). Table 15 also shows that participants gained an average of 7.9 (range 3-18) standard score gains in pseudoword decoding, and an average standard score ratio gain of 7 (range 3-13). This means that participants gained approximately 2.4 standard scores in word reading, and seven standard scores in pseudoword decoding per hour of intervention received across Experiments 8, 9, and 10.

Table 15

Outcomes in YARC Early Word Reading (EWR) and WIAT-II Pseudoword Decoding (PWD) Subtests Across Experiments 8, 9, and 10.

	Total Hours Intervention	EWR-SSG	EWR- SSRG	PWD-SSG	PWD- SSRG
Participants Trained in Single Letter GPCs					
Bob	.5	2	4.2	6	12.5
Flo	1.1	7	6.4	--	--
Participants Trained in Single-letter GPCs and Segmenting/Blending Words					
Dee	1	-10	-10	3	3
Eli	1	9	9	13	13
Guy	1.4	1	.7	6	4.3
Joe	1.5	10	6.7	3	2
Participants Trained in Single-letter GPCs and Blending Words					
Ivy	1.2	0	0	6	5
Kya	1.8	4	2.2	18	10
Averages	1.2	2.9	3.7	7.9	7

Table 16 displays the outcomes in the recoded DIBELS NWF subtest, and the Real Word list for participants in Experiments 8, 9, and 10. Table 16 depicts participants' real words read correctly at pre-test (RW-WRC Pre-Test) and post-test (RW-WRC Post-Test); and, nonsense words read correctly at pre-test (NWF-WRC Pre-Test) and post-test (NWF-WRC Post-Test). Results are described as the average correct per minute (CPM) gains from pre- to post-test, and resulting fold change in average performance. Table 16 shows that the average participant gained 8.2 (range 8-16) real words segmented and blended CPM, and 9 (range 7-13) nonsense words segmented and blended CPM. These gains represent an average 2.3-fold increase in

number of real words correctly decoded, and a 2.8-fold increase in number of nonsense words correctly decoded.

Table 16

Outcomes in Words Read Correctly (WRC) on the Recoded DIBELS Nonsense Word Fluency Subtest (NWF) and Real Word (RW) Probe Across Experiments 8, 9, and 10.

	Total Hours Intervention	Total Intervention Sessions	RW- WRC Pre- Test	RW- WRC Post- Test	NWF- WRC Pre-Test	NWF- WRC Post- Test
Participants Trained in Single Letter GPCs						
Bob	.5	6	12	22	4	14
Flo	1.1	12	2	8	1	7
Participants Trained in Single-letter GPCs and Segmenting/Blending Words						
Dee	1	12	10	20	4	17
Eli	1	13	4	20	10	18
Guy	1.4	9	16	24	10	19
Joe	1.5	14	10	18	7	14
Participants Trained in Single-letter GPCs and Blending Sounds Into Words						
Ivy	1.2	13	8	18	4	12
Kya	1.8	16	2	18	0	12
Averages	1.2	12	8	18.5	5	14

These outcomes demonstrate that in approximately one hour (range .5-1.8) of intervention, over 12 (range 6-16) intervention sessions, the average participant gained 2.9 standard scores in word reading, seven standard scores in pseudoword decoding, 8.2 real words blended CPM, and nine nonsense words blended CPM.

4.4.2. Partial and Complete Decoding: Generalisation Probes Across Experiments 8, 9, and 10.

Participants in Experiments 8, 9, and 10 were also administered two types of generalisation checks (partial and complete decoding) across baseline and intervention conditions. The generalisation checks and form of reporting are identical to those described in Chapter 3. These checks were administered to evaluate the relation between increasing frequencies in letter sounds, blending sounds, and decoding words on partial decoding (i.e., the number of nonsense word GPCs correctly blended), and complete decoding (i.e., the number of real CVC words read correctly, WRC).

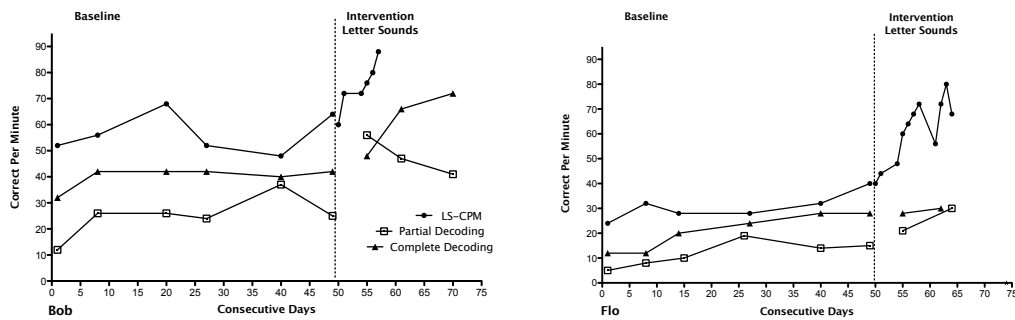
Figure 28 shows generalisation outcomes for participants that received the PT intervention in letter sounds only (Bob and Flo), and for participants that received the PT intervention in letter sounds and blending sounds (Ivy and Kya). Bob's generalisation checks suggest that increases in letter sound frequencies occasioned an increase in partial and complete decoding, and resulted in 10 WRC per minute (see Figure 28). For complete decoding, this increase continued post-intervention, however for partial decoding a decrease was observed post-intervention. This differential effect may be due to exposure to CVC words during regular classroom instruction and frequency effects of CVC words, in contrast to lack of exposure to nonsense words. Flo demonstrated a small increase in partial decoding, but not complete decoding. This performance pattern may indicate that Flo could benefit from the PT intervention programme in decoding words.

Figure 28 also shows generalisation checks for Ivy and Kya, trained in blending sounds first, and subsequently in letter sounds. Results are unclear for Ivy, as baseline data were not collected immediately prior to the blending sounds

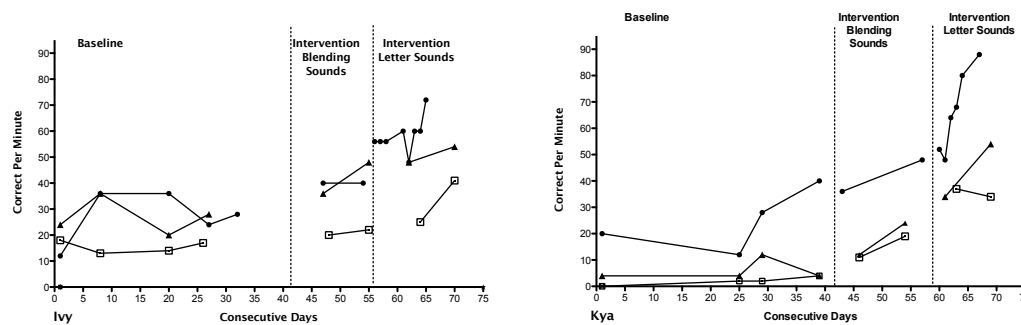
intervention. It appears that increased frequencies in letter sounds occasioned an increase in partial and complete decoding. Kya demonstrates at or near zero partial blended in baseline, and an increasing trend when provided with the PT intervention in blending sounds. When letter sound frequencies were increased through the PT intervention, generalisation effects were observed in partial and complete decoding resulting in a gain of 16 WRC per minute (see Figure 28).

These results suggest that to occasion generalisation to decoding for participants who demonstrate difficulties in letter sounds and blending, both foundational reading skills needed to be targeted with the PT intervention. This finding reflects the generalisation data reported for Alf, Cal, Dan, and Eve in the previous chapter. These participants also demonstrated difficulties in both of these foundational reading skills, and generalisation to decoding did not occur until both skills were targeted with the PT intervention.

Participants Trained in Letter Sounds



Participants Trained in Letter Sounds and Blending Sounds into Words



LS-CPM= Letter Sounds Correct Per Minute

Partial Decoding = Nonsense Word Fluency – Letter Sounds Blended

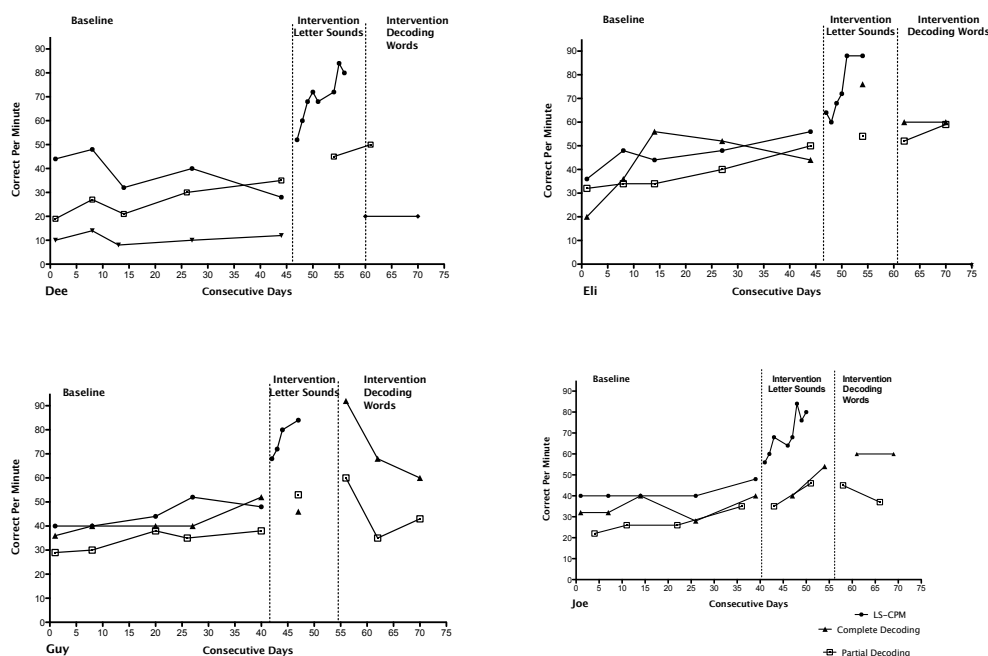
Complete Decoding= Real Words (CVC) – Letter Sounds Blended

Figure 28. Generalisation outcomes for participants that receive the PT intervention for letter sounds only, and participants that received the PT intervention in letter sounds and blending sounds. Generalisation checks for partial and complete decoding administered within and across baseline and intervention conditions.

Figure 29 shows generalisation data for participants trained in letter sounds and decoding words (Dee, Eli, Guy, and Joe). Dee's performance patterns suggest that increases in letter sound frequencies occasioned an increase in NW phonemes blended and resulted in gains of 10 WRC per minute (see Figure 29). No generalisation probes were collected in the decoding words condition, and thus generalisation effects of this intervention, and longer-term gains are unknown. Figure 29 shows that Eli demonstrated a moderate increase in NW phonemes blended in the intervention

conditions, but also showed an ascending performance in the baseline condition. Therefore, the effects of the intervention on decoding for this participant are unclear. Guy showed the largest increases in CVC and NW phonemes blended at the beginning of the decoding condition, but these gains subsequently decreased (see Figure 29). Finally, Joe’s performance pattern indicates that increasing frequencies in letter sounds produced gains in WRC and NW phonemes blended, this increase continued for RW in the decoding condition, but not for nonsense words.

Participants Trained in Letter Sounds and Decoding (Segmenting/Blending) Words



LS = Letter Sounds
 Partial Decoding= Nonsense Word Fluency – Letter Sounds Blended
 Complete Decoding= Real Words (CVC) – Letter Sounds Blended

Figure 29. Generalisation outcomes for participants that receive the PT intervention for letter sounds and decoding words. Generalisation checks for partial and complete decoding administered within and across baseline and intervention conditions.

Taken together, the generalisation outcomes suggest that increasing frequencies in foundational reading skills of blending sounds and letter sounds, results

in moderate generalisation to partial and complete decoding of CVC words. Generalisation gains however, are of less magnitude than those observed in the previous chapter. The higher levels of pre-intervention accuracy demonstrated by the current participants, and the shorter intervention time employed may explain these differences. In the previous chapter letter sound accuracy and fluency scores were low for participants, as were partial and complete decoding. Participants in the experiments reported in this chapter demonstrated at or near full accuracy in letter sounds, and were able to decode CVC words (albeit at a low frequency). Therefore, generalisation gains may be most apparent for those with larger skill deficits, and particularly with accuracy deficits.

Limited generalisation gains may also be an artifact of short intervention durations received by participants (approximately 1.2 hours). In addition, where decoding was directly targeted it was with a limited subset of MLGPCs (three targets from one set). The participants in this phase of the research may have benefited from additional intervention directly targeting a larger subset of MLGPCs, but did not receive such intervention due to time constraints.

4.5. Social Validity

Social validity was central to the research question to assess the acceptability and suitability of the Tier 2 intervention to Senior Infant students. Parental and teacher views were not solicited due to time constraints. The social validity measure (Musti-Rao, 2005) used with participants was identical to that described in Chapter 3 (see Appendix G). All students reported that they liked taking part in the intervention. When questioned on what they liked best five reported the treasure chest, one stated the stickers earned, and one reported using the SCC. In regard to

being part of the group, and earning rewards, all participants rated liking it “a lot”; in regard to learning new skills, six rated liking it “a lot”, one rated liking it “a little”, and one rated “not much”.

Seven participants reported that they had learned important things in the intervention (one did not). When probed specifically on the most important things learned, three commented on “letter sounds”, one on “listening for sounds”. All participants reported that what they learned helped them do better work in school (five rated “a lot”; three rated “a little”). Seven participants reported that what they learned helped them do better work at home (six rated “a lot”; one rated “a little”), one did not. Seven participants reported that they used the skills learned in the intervention, when queried exactly where they used the skills, three commented on reading words, one on homework, one on listening, one on spelling, and one on letter sounds. Six participants wanted the programme to have lasted longer, one did not, and one did not know.

Chapter 5: Introduction

Chapter 4 reported the second screening phase (January, 2013) in the Senior Infants grade across two classrooms in an urban DEIS school. Of the 36 participants eligible for the intervention, eight were identified with difficulties in letter sounds, and two with difficulties in blending sounds. Experiments 8 and 10 (Chapter 3) showed that the intervention was effective for participants in achievement of at or near performance standards in letter sounds and blending sounds. Experiment 9 demonstrated that the PT intervention programme was also effective in targeting decoding words comprised of multiple letter phoneme grapheme conversions (MLGPCs). Generalisation outcomes indicated that increasing frequencies in foundational reading skills resulted in gains regarding decoding real and nonsense words. The importance of fluency in sublexical skills of letter sounds and phonemic awareness for reading acquisition has been highlighted (e.g., Burke et al., 2010; Hudson et al., 2009; Hudson et al., 2010). Moreover, competence in decoding and automaticity in word reading are critical for proficiency in reading development (Snowling & Hulme, 2010; Torgesen, 2005).

Chapter 5 provides a focus on targeted intervention for high frequency MLGPCs and high frequency words (HFW). The importance of targeting these specific skill has been highlighted in Chapter 1 (e.g., Solity & Vousden, 2009; Vousden et al., 2011; Clarke, 1994, 2013). Some additional research studies add context to the introduction of the current chapter. In regard to teaching words, a number of studies have compared various teaching strategies in regards to effectiveness and efficiency (e.g., Joseph et al., 2012; Joseph & Nist, 2006; Nist & Joseph, 2008; Volpe, Mule, Briesch, Joseph, and Burns, 2011). These studies revealed

mixed findings in regard to the relative effectiveness of the different strategies, however, under many conditions the traditional drill and practice flashcard method (discrete trials) emerged as the most efficient (Joseph & Nist, 2006; Nist & Joseph, 2008; Volpe et al, 2011).

For example, Volpe and colleagues (2011) compared the effectiveness and efficiency of two flash card methods (traditional drill and practice i.e., discrete trials and incremental rehearsal) to improve word recognition with four 1st grade students identified as struggling readers. The authors held opportunities to respond and length of instructional time constant under two conditions. Results showed that differences in effectiveness were minimal when holding both time and opportunities to respond constant. When holding opportunities to respond constant (and not instructional time), the traditional drill and practice teaching strategy emerged as the most efficient method.

With regard to comparisons of targeting word level and phonic analysis, Schmidgall and Joseph (2007) evaluated the instructional effectiveness and efficiency of three word-reading interventions (interspersal drill, word boxes, and traditional drill and practice) with six 1st grade readers considered “at risk”. Results showed that the word boxes (which focused on phonics analysis) had the greatest impact on the cumulative number of words read for four participants. However, the traditional drill and practice method yielded the greatest cumulative learning rate for all participants.

Taken together, these results suggest that using the traditional drill and practice flashcard method is effective in teaching students to read words. None of the studies mentioned, however, use the 100 high frequency words as identified by Stuart and colleagues (2003) and Solity & Vousden (2009). Studies tend to use Dolch word lists, other HFW word lists, or teacher generated lists.

In regard to teaching the most frequent MLGPCs, one study was identified. Chen and Savage (2014) taught a set of frequently occurring MLGPCs to at risk 1st and 2nd Graders. These researchers compared the effects of two reading programmes in a randomised control group design. One programme taught MLGPCs by frequency of occurrence in children's text (intervention group), and the other taught whole word recognition (control group). The intervention group received instruction that specifically targeted the top 36 MLGPCs (omitting two from a total of 38). These researchers found that the intervention group demonstrated greater gains in post-test assessments of reading, spelling, and reading motivation. Chen and Savage (2014) are the only authors to date who have experimentally evaluated reading outcomes as a result of intervention with the most common MLGPCs.

The Experiment reported in the current chapter extends the findings of previous research in a number of ways. The current research targets Senior Infant students (kindergarten equivalent), whereas previous research has targeted 1st and 2nd graders. The current research directly targets a subset of the most frequently occurring MLGPCs embedded in real words, and the 100 high frequency words as identified by Solity and Vousden (2009). The current research provides the first demonstration of targeting these foundational reading skills using PT, and specifically as a Tier 2 intervention with at risk readers.

5.1. Experiment 11

Chapter 4 describes the PT treatment programme targeting fluency in letter sounds, blending sounds, and decoding words in an urban DEIS school. The current experiment attempts to address some of the limitations identified in the previous chapter, and to target an additional foundational reading skill – high frequency words

(HFW). This phase of the research targeted a DEIS school in a rural setting; a unique feature of the school is that it is situated in a Gaeltacht area. This means that Irish is spoken at home and in school as the first language, and therefore students speak English as a second language (EL2).

Similarities and differences are observed in the intervention focus between the current experiment and the previous experiments (Chapters 2, 3 and 4) in both the rural and urban DEIS schools. Three previous Experiments (1, 5, and 8) had demonstrated the effectiveness of the PT intervention programme in targeting letter sound fluency. Experiments 7 and 10 targeted blending sounds, and while the PT intervention programme was largely effective, outcomes were inconsistent across participants in that some required multiple programme modifications. Conversely, Experiment 9 directly targeted decoding, and no modifications were necessary.

This phase of the research focused on decoding words. Two limitations of Experiment 9 (Chapter 4) were addressed in the current experiment. Specifically, the participants in Experiment 9 were not directly identified as “at risk” using a measure of decoding. In addition, the scoring criteria for correct responding was defined as the whole word correctly blended, and therefore did not measure partial blending (i.e., some, but not all of the phonemes correctly blended). To address these shortcomings, Experiment 11 incorporated a universal screening instrument to measure decoding pseudo words, and redefined scoring criteria for correct responding to account for partial blending.

An additional intervention focus for Experiment 11 included HFW, as all previous experiments focused on sublexical fluency (i.e., below the word level). Automaticity in HFW is critical for reading development (Torgesen & Hudson, 2006). It was considered important to investigate if similar procedures used in Experiment 9

(Chapter 4; decoding words) were effective for targeting fluency in HFW.

The purpose of Experiment 11 was to identify a pool of students at risk of reading difficulties in decoding and word reading, and to implement a PT intervention programme as a Tier 2 intervention with these students. Experiment 11 aimed to investigate the effectiveness of the PT intervention programme in building frequencies in decoding and high frequency words, to evaluate maintenance of the skill, and to evaluate the effects of the PT intervention on other reading outcomes. The intervention goal was achievement of performance standards in high frequency words (60-65; Fabrizio & Moors, 2003), and in decoding words (80-120; K. Brooks Newsome, personal communication, January 15, 2013), at the learning rate of x2 celeration.

5.1.1. Method Condition 1: Decoding (Segmenting and Blending) Words

5.1.1.1. Participants and Setting

Two Senior Infant classes (equivalent to kindergarten; $n=35$) within a DEIS school in a rural Irish speaking school participated in the first phase of the study. Parents received information letters through the school with an option to consent to participation. Seven students were excluded as they were accessing a reading intervention or literacy support at the time of recruitment. Twenty-eight participants ($M = 6.6$, range 5.10-7.4 years) participated in the universal screening phase.

5.1.1.2. Universal Screening Phase

Students in need of additional instruction in decoding and sight word recognition were identified using three screening instruments. One screening instrument was identical to that used in Experiment 1, assessing letter sound knowledge (LSK), one assessed early word reading, and the other pseudo-word reading. The Early Word Recognition (EWR) subtest from the YARC assessment

battery (Hulme et al., 2009; as described in Chapter 1) and the Pseudo-Word Reading (PWD) subtest from the Wechsler Individual Achievement Test, 2nd UK edition (WIAT-II; Wechsler, 2005) as described in Chapter 4.

When all 27 participants had been screened, they were rank ordered in terms of scores across three outcome measures. The eight participants that scored lowest across the three screeners were selected for the PT intervention in decoding (Condition 1) or sight word reading (Condition 2). These eight participants were randomly assigned to the decoding condition (Condition 1: P1, P2, P3, and P4), or the sight word condition (Condition 2: P5, P6, P7, P8).

5.1.1.3. Participants

Four participants ($M = 7$, range 6.9-7.4 years) participated in Condition 1 of Experiment 11. Participants were typically developing (P1, P2, P3, and P4; two males, two females), and all spoke English as a second language (EL2).

5.1.1.4. Dependent Variables and Data Analysis

Dependent variables and data analysis were identical to those described in Experiment 1 (Chapter 2) with two exceptions. First, in the multiple-baseline design, data were evaluated with respect to the students' response to the intervention, and post-intervention performance. Second, post-intervention rates in decoding words were only available for three time points (one-, two-, and three-week checks).

5.1.1.5. Materials

Stimulus and other materials were identical to those utilized in Experiment 9 (Chapter 4).

5.1.1.6. Design

A multiple-baseline across participants with a multiple probe element across word sets design was employed to assess the effectiveness of the PT intervention

programme in building frequencies in decoding words.

5.1.1.7. Procedure

Baseline. Baseline procedures were identical to those described in Experiment 9 (Chapter 4).

Intervention. The independent variable involved a Tier 2 PT intervention programme that included PT as an intervention to target fluency in decoding words, and a decision-making framework to monitor progress. Intervention procedures were identical to those used in Experiment 8 (Chapter 4). Participants received the one-to-one intervention five days per week, daily sessions lasted approximately seven minutes, and the average total intervention duration was 2.3 hours.

Decision Rules and Programme Modifications. Decision rules were identical to those employed in Experiment 1 (Chapter 2) and programme modifications were identical to those described in Experiment 9 (Chapter 4).

5.1.1.8. Interobserver agreement (IOA). IOA data were calculated as described in Experiment 1 (Chapter 2). Exact Count-Per-Interval IOA data were collected for 24% of all baseline sessions demonstrating 95.7% agreement, and 27.8% of all intervention sessions achieving 97% agreement. Overall this is 96% agreement across 25.9% across all sessions.

5.1.1.9. Procedural integrity. The procedural integrity checklists were identical to those used in Experiment 1 (see Appendix C for baseline checklist, and Appendix D for intervention checklist); PI data were calculated in the same manner as Experiment 1 (Chapter 2). Independent observers collected data on procedural integrity for 17% of baseline and 20% of intervention sessions. Overall procedural adherence was 98.7% for intervention sessions and 100% for all baseline sessions.

5.1.2. Condition 1: Results and Discussion

Figure 30 depicts a multiple-baseline across participants ($n = 4$) with multiple probes across target skills design for the PT intervention programme in decoding words. Figure 30 shows that all participants entered the baseline condition concurrently, and a minimum of three probes were collected before P1 was entered into the intervention phase for Set 1 of decoding words. Participant 1 responded to the intervention, and P2 was entered into the intervention phase; P2 also responded to the intervention. Participant 3 was then entered and responded to the intervention followed by Participant 4 entering intervention for Set 1. Participant 4 also responded to the intervention (see Figure 30). Each participant received intervention on 2-4 word sets. Participants 1 and 2 achieved at (or near) performance standards on all four sets of SAFMEDS, i.e., Sets 1, 2, 3, and 4. Participant 3 reached performance standards on two of the four sets: 1 and 3, and reduced criteria on a third set (Set 2). Participant 4 achieved performance standards on two sets: 1 and 2, however performance on Set 1 subsequently decreased, and the intervention was terminated at reduced criteria due to time constraints (see Figure 30). It took approximately 2.3 hours (range 1.9-2.7) of intervention, and 19 intervention sessions (range 18-21) per participant to achieve performance standards (or reduced criteria) in these word sets.

Intervention time, cumulative words and number of MLGPCs targeted by each participant are described as follows. In 2.5 hours of intervention time over 19 intervention sessions, P1 reached performance standards with 48 unique words that comprised of 13 MLGPCs and 18 single letter sounds. In 2.7 hours of intervention time over 21 intervention sessions P2 reached performance standards (or reduced criteria) with 48 unique words that comprised of 13 MLGPCs and 18 single letter sounds. In 2.2 hours of intervention time over 18 intervention sessions, P3 reached

performance standards (or reduced criteria) with 36 unique words that comprised of 10 MLGPCs and 18 single letter sounds. In 1.9 hours of intervention time over 19 intervention sessions, P4 reached performance standards (or reduced criteria) with 24 unique words that comprised of seven MLGPCs and 15 single letter sounds.

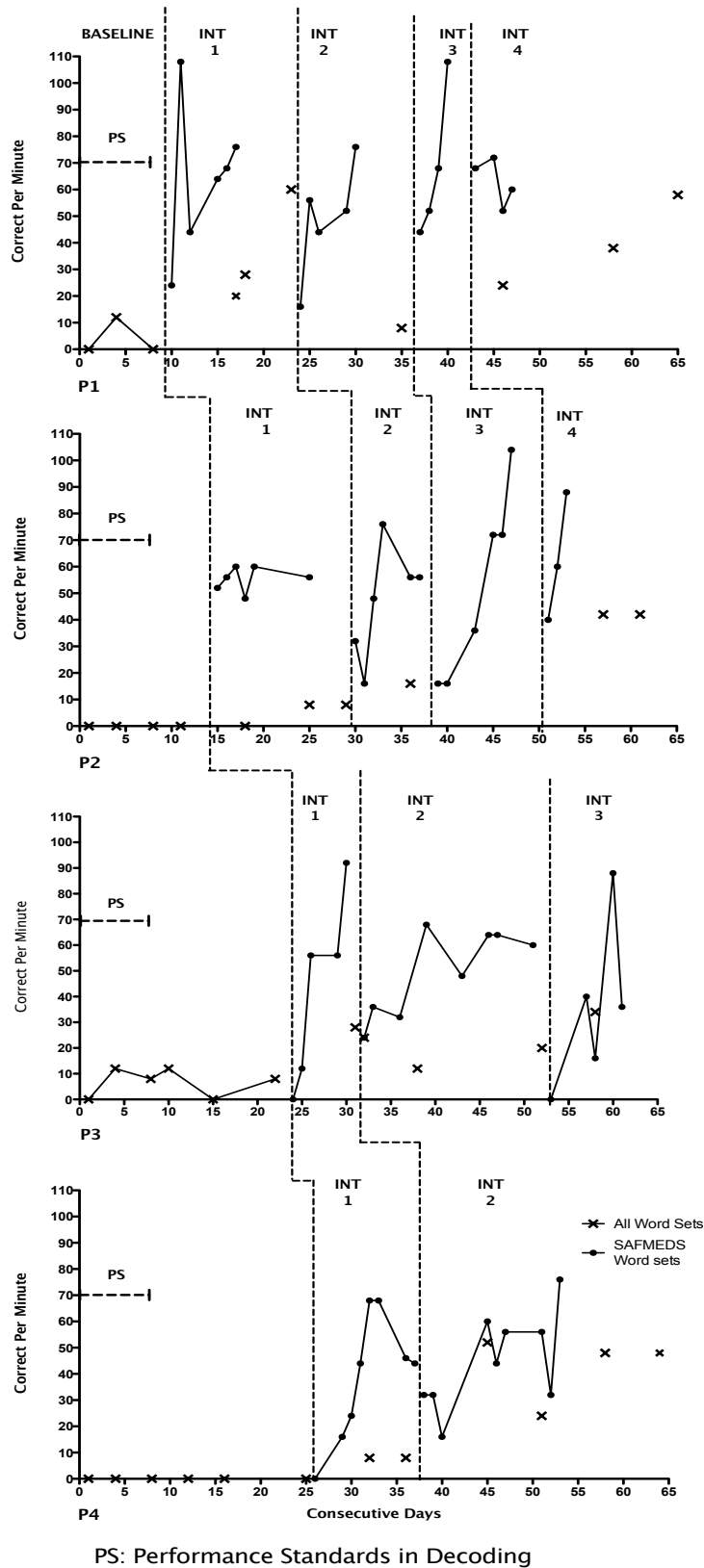


Figure 30. Experiment 11, Condition 1: multiple-baseline across participants with multiple probes across word sets design for the PT intervention in decoding words.

5.1.2.1. Maintenance checks

Figure 31 represents participants' performance at median baseline, end of intervention, one week, two weeks, and three weeks post-intervention. Figure 31 shows that participants' average performance in SAFMEDS word sets at median baseline was 8.5 GPCs blended CPM (range 4-16), and at the end of the intervention this had increased to 95 GPCs blended CPM (range 10.4); an 11-fold increase in correct responding. The maintenance checks implemented at Week 1, Week 2, and Week 3 post-intervention demonstrate average performance of 91 GPCs blended CPM (range 64-132). This suggests that the performance standards (or reduced criteria) achieved by participants in the intervention were largely maintained up to three weeks following intervention (see Figure 31).

Figure 31 shows that P1 achieved performance standards during the intervention phase and maintained these performance levels up to three weeks post-intervention. Figure 31 shows that P1's average performance across SAFMEDS word sets (1, 2, 3, and 4) at median baseline was 12 CPM (range 4-16), and at intervention end this had increased to 91 CPM (range 80-108); a 7.6-fold increase in correct responding. The maintenance checks implemented at Week 1, Week 2, and Week 3 post-intervention demonstrates average performance of 95 CPM (range 64-124; see Figure 25).

Figure 31 also shows that P2 achieved performance standards during the intervention phase and maintained these performance levels up to three weeks post-intervention. Figure 31 shows that P2's average performance across SAFMEDS word sets (1, 2, 3, and 4) at median baseline was seven CPM (range 4-16), and at intervention end this had increased to 96 CPM (range 80-108); a 13.7-fold increase in

correct responding. The maintenance checks implemented at Week 1, Week 2, and Week 3 post-intervention demonstrates average performance of 95 CPM (range 76-132; see Figure 31).

Figure 31 shows that P3 achieved performance standards (or reduced criteria) during the intervention phase and maintained these performance levels up to three weeks post-intervention. Figure 31 shows that P3's average performance across SAFMEDS word sets (1, 2, and 3) at median baseline was nine CPM (range 4-16), and at intervention end had increased to 105 CPM (range 80-108); a 11.7-fold increase in correct responding. The maintenance checks implemented at Week 1, Week 2, and Week 3 post-intervention demonstrates average performance of 94 CPM (range 80-116; see Figure 31).

Figure 31 shows that P4 achieved performance standards (or reduced criteria) during the intervention phase and maintained these performance levels up to three weeks post-intervention. Figure 31 shows that P3's average performance across SAFMEDS word sets (1 and 2) at median baseline was six CPM (range 4-8), and at intervention end this had increased to 86 CPM (range 80-108); a 14.3-fold increase in correct responding. The maintenance checks implemented at Week 1, Week 2, and Week 3 post-intervention demonstrates average performance of 86 CPM (range 76-108; see Figure 31).

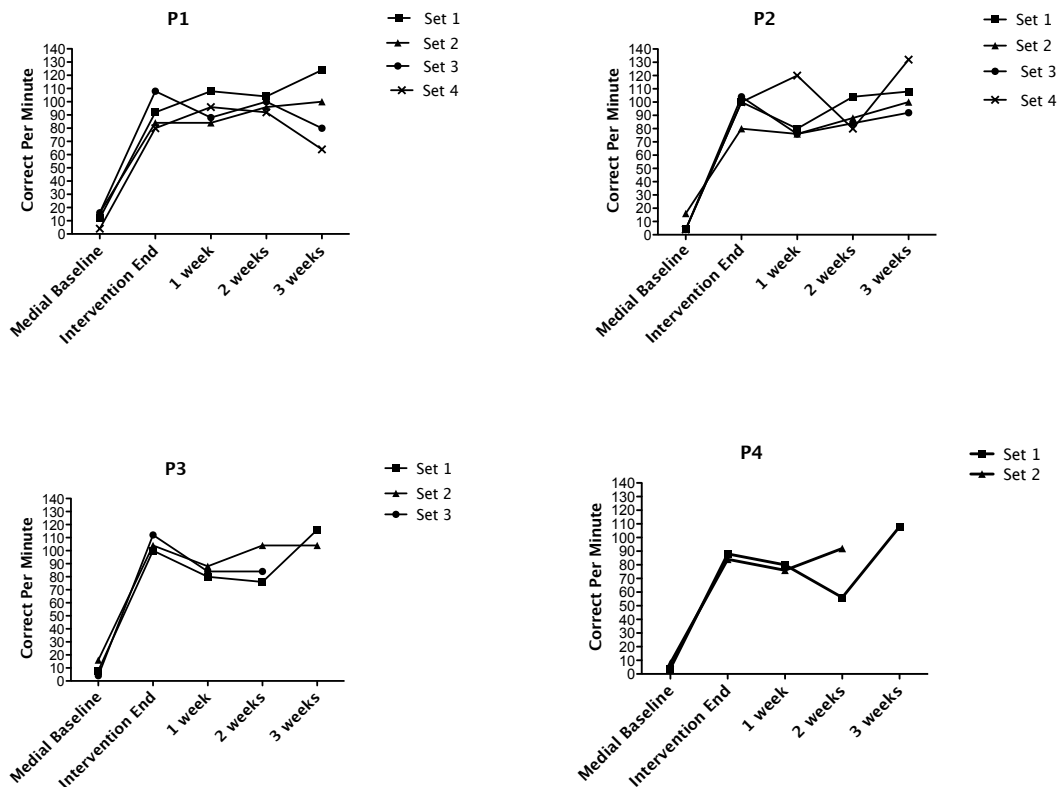


Figure 31. Experiment 11, Condition 1 - maintenance data displaying performance in decoding words at median baseline, intervention end and three post-intervention checks: 1 week; 2 weeks, and 3 weeks.

5.2.1. Method Condition 2: High Frequency Words

5.2.1.1. Participants and Setting

Two Senior Infant classes (equivalent to kindergarten; $n=35$) within a DEIS school in a rural Irish speaking school participated in the first phase of the study. Parents received information letters through the school with an option to consent to participation. Seven students were excluded as they were accessing a reading intervention or literacy support at the time of recruitment. Twenty-eight participants ($M = 6.6$, range 5.10-7.4 years) participated in the universal screening phase.

5.2.1.2. Universal Screening Phase

Students in need of additional instruction in decoding and sight word

recognition were identified using three screening instruments. One screening instrument was identical to that used in Experiment 1, assessing letter sound knowledge (LSK), one assessed early word reading, and the other pseudo-word reading. The Early Word Recognition (EWR) subtest from the YARC assessment battery (Hulme et al., 2009; is described in Chapter 1) and the Pseudo-Word Reading (PWD) subtest from the Wechsler Individual Achievement Test, 2nd UK edition (WIAT-II; Wechsler, 2005) is described in Chapter 4.

When all 27 participants had been screened, they were rank ordered in terms of scores across three outcome measures. The eight participants that scored lowest across the three screeners were selected for the PT intervention in decoding (Condition 1) or sight word reading (Condition 2). These eight participants were randomly assigned to the decoding condition (Condition 1: P1, P2, P3, and P4), or the sight word condition (Condition 2: P5, P6, P7, P8).

5.2.1.3. Participants

Four participants (M = 6.9, range 6.7-7 years) participated in Condition 2 of Experiment 11. Participants were typically developing (P5, P6, P7, and P8; 4 males), and all spoke English as a second language (EL2).

5.2.1.4. Dependent Variables and Data Analysis.

Dependent variables and data analysis were identical to those described in Experiment 1 (Chapter 2) with two exceptions. First, in the multiple-baseline design, data were evaluated with respect to the students' response to the intervention, and performance at the end of the intervention. Second, post-intervention rates in decoding words were only available for three time points (one-, two-, and three-week follow-up checks).

5.2.1.5. Materials

The experimenter manually constructed the stimulus materials, which constituted of word cards in the form of SAFMEDS, of the 100 most frequent words identified by Solity and Vousden (2009). The word cards were 10.5cm x 5 cm in size and the words were printed in size 46 font. There were 10 stimulus sets (Sets 1-10) in the High Frequency Words condition. Each set contained 30 stimuli (10 words, three examples of each word; this meant each word was repeated three times within each stimulus set). In order to control for stimulus presentation across the 10 stimulus sets the following measures were taken. Words beginning with the same letter were randomly divided between sets, as were regular words (i.e., decodable words), then the remaining words were randomly distributed to make up 10 sets comprising of 10 words each (see Appendix I).

Other materials were identical to those utilised in Experiment 1 (Chapter 2).

5.2.1.6. Design

A multiple-baseline across participants with a multiple probe element across word sets design was employed to assess the effectiveness of the PT intervention programme for building frequencies in high frequency words.

5.2.1.7. Procedure

Baseline. Baseline procedures were identical to those described in Experiment 11 Condition 1 with the exception that the participant would say the whole word as a sight word when presented with the SAFMEDS card (i.e., not segment and blend the word).

Intervention. The independent variable involved a Tier 2 PT intervention programme that included PT as an intervention to target fluency in high frequency words, and a decision-making framework to monitor progress. Intervention

procedures were identical to those described in Experiment 11 Condition 1 with the exception that the participant would say the whole word as a sight word when presented with the SAFMEDS card. Participants received the one-to-one intervention five days per week, daily sessions lasted approximately five minutes, and the average total intervention duration was 2.3 hours.

Decision Rules and Programme Modifications. Decision rules were identical to those employed in Experiment 1 (Chapter 2) and programme modifications utilised were identical to those described in Experiment 9 (Chapter 4).

5.2.2. Condition 2: Results

Figure 32 depicts a multiple-baseline across participants ($n = 4$) with multiple probes across word sets design for the PT intervention in high frequency words. Figure 32 shows that all participants entered the baseline condition concurrently, and a minimum of four probes were collected before P5 was entered into the intervention phase for HFW Set 1. Participant 5 responded to the intervention, and P6 was entered into the intervention phase (Set 1) and also responded to the intervention. Participant 7 was then entered into the intervention phase targeting HFW Set 5 and showed a response to intervention. Finally, participant 8 was entered into the intervention for HFW Set 6, and also responded to the intervention (see Figure 32).

Each participant received intervention on a number of different word sets ($M = 3.5$, range 2-5) in this manner. Participants 5 and 7 achieved performance standards on each of the SAFMEDS stimulus sets targeted (i.e., P5 Sets 1, 2, 3, 8, and 10; P7 Sets 5, 10, 9;). Participant 6 achieved performance standards on Set 7 and reduced criteria on Set 1. Participant 8 achieved performance standards on Sets 6 and 9 and reduced criteria on Set 4 and 8. It took on average 1.5 (range 0.9-1.9) hours of

intervention, and 17 intervention sessions per participant to performance standards (or reduced criteria) in the words sets targeted.

Intervention time, cumulative words targeted, and rate gains achieved by each participant are described as follows. In 1.9 hours of intervention time over 21 intervention sessions, P5 reached performance standards with 50 HFW, and demonstrated gains of 32 words CPM. In 1.9 hours of intervention time over 20 intervention sessions, P6 reached performance standards (or reduced criteria) with 30 HFW, and showed gains of 58 words CPM. In 1.1 hours of intervention time over 13 intervention sessions, P7 reached performance standards with 30 HFW, and demonstrated gains of 68 words CPM. In 0.9 hours of intervention time over 13 intervention sessions, P8 reached performance standards (or reduced criteria) with 40 HFW, and made gains of 56 words CPM.

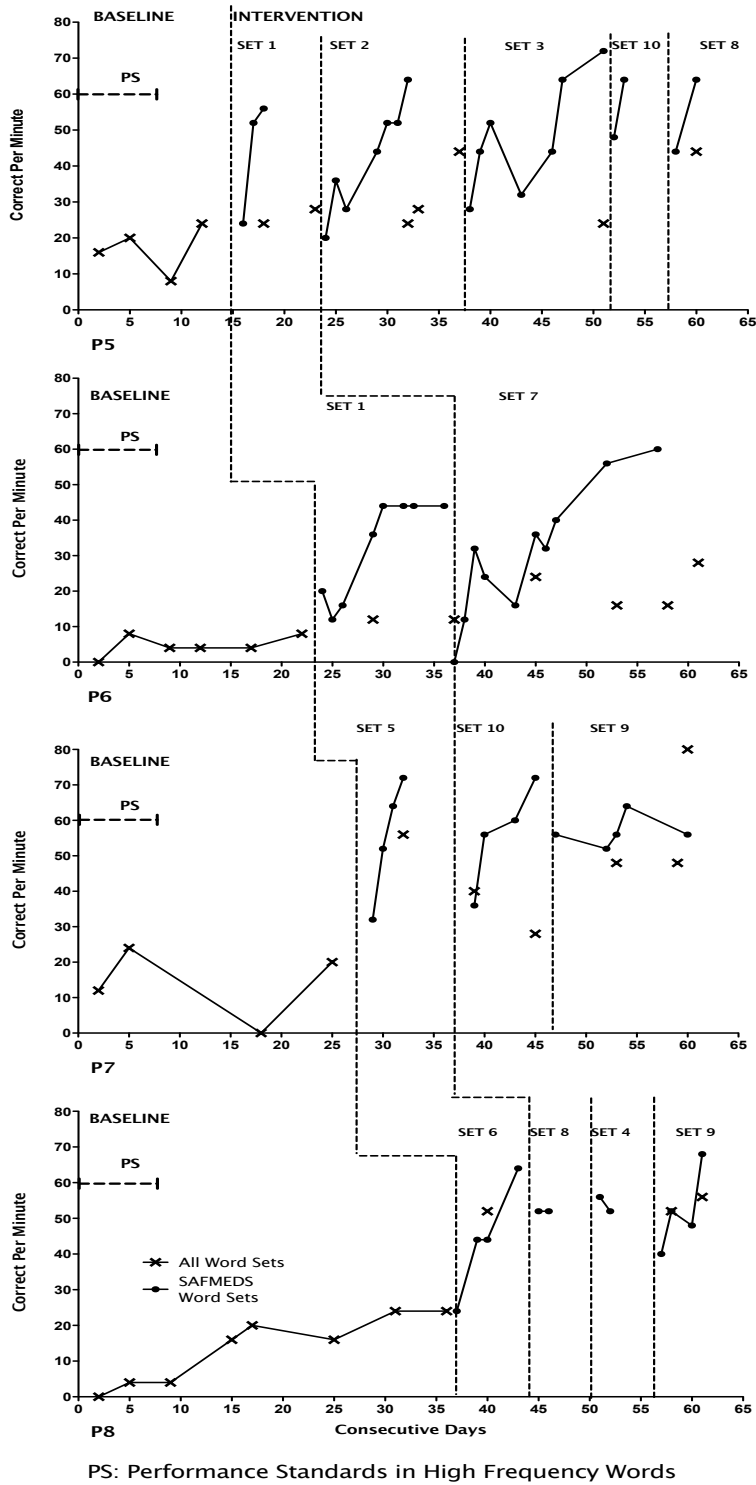


Figure 32. Experiment 11, Condition 2: multiple-baseline across participants with multiple probes across word sets design for the PT intervention in high frequency words.

5.2.2.1. Maintenance of high frequency words

Figure 33 represents participants' performance at median baseline, end of the intervention, one week, two weeks, and three weeks post-intervention. Figure 33 shows that participants' average performance in SAFMEDS word sets at median baseline was 15.5 words CPM (range 4-28), and at the end of the intervention this had increased to 67.5 CPM (range 60-72); a 4.4-fold increase in correct responding. The maintenance checks implemented at Week 1, Week 2, and Week 3 post-intervention demonstrate an average performance of 58 words CPM (range 32-76) a 0.9-fold decrease in correct responding.

Figure 33 shows that P5 achieved performance standards during the intervention phase and maintained these performance levels up to three weeks post-intervention. Figure 33 shows that P5's average performance across SAFMEDS high frequency word sets (1, 2, 3, 10, and 8) at median baseline was 15 CPM (range 4-20), and at intervention end increased to 67 CPM (range 60-72); a 4.5-fold increase in correct responding. The maintenance checks implemented at Week 1, Week 2, and Week 3 post-intervention demonstrate an average performance of 62 CPM (range 44-76; see Figure 33).

Figure 33 also shows that P6 achieved performance standards (or reduced criteria) during the intervention phase, but did not maintain these levels three weeks post-intervention. Figure 33 shows that P6's average performance across SAFMEDS word sets (1 and 7) at median baseline was four CPM (range 4-4), and at intervention end this had increased to 66 CPM (range 32-72); a 16.5-fold increase in correct responding. The maintenance checks implemented at Week 1, Week 2, and Week 3 post-intervention demonstrate an average performance of 40 CPM (range 32-64; see Figure 33).

Figure 33 shows that P7 achieved performance standards during the intervention phase, and largely maintained these performance levels up to three weeks post-intervention. Figure 33 shows that P7's average performance across SAFMEDS word sets (5, 10, and, 9) at median baseline was 23 CPM (range 16-28), and at intervention end this had increased to 71 CPM (range 68-72); a 3-fold increase in correct responding. The maintenance checks implemented at Week 1, Week 2, and Week 3 post-intervention demonstrate an average performance of 65 CPM (range 48-80; see Figure 33).

Figure 33 shows that P8 achieved performance standards (or reduced criteria) during the intervention phase and maintained these performance levels up to three weeks post-intervention. Figure 33 shows that P8's average performance across SAFMEDS word sets (4, 6, 8, and, 9) at median baseline was 20 CPM (range 16-24), and at intervention end this had increased to 66 CPM (range 60-72); a 3.3-fold increase in correct responding. The maintenance checks implemented at Week 1, Week 2, and Week 3 post-intervention demonstrate an average performance of 64 CPM (range 44-76; see Figure 33).

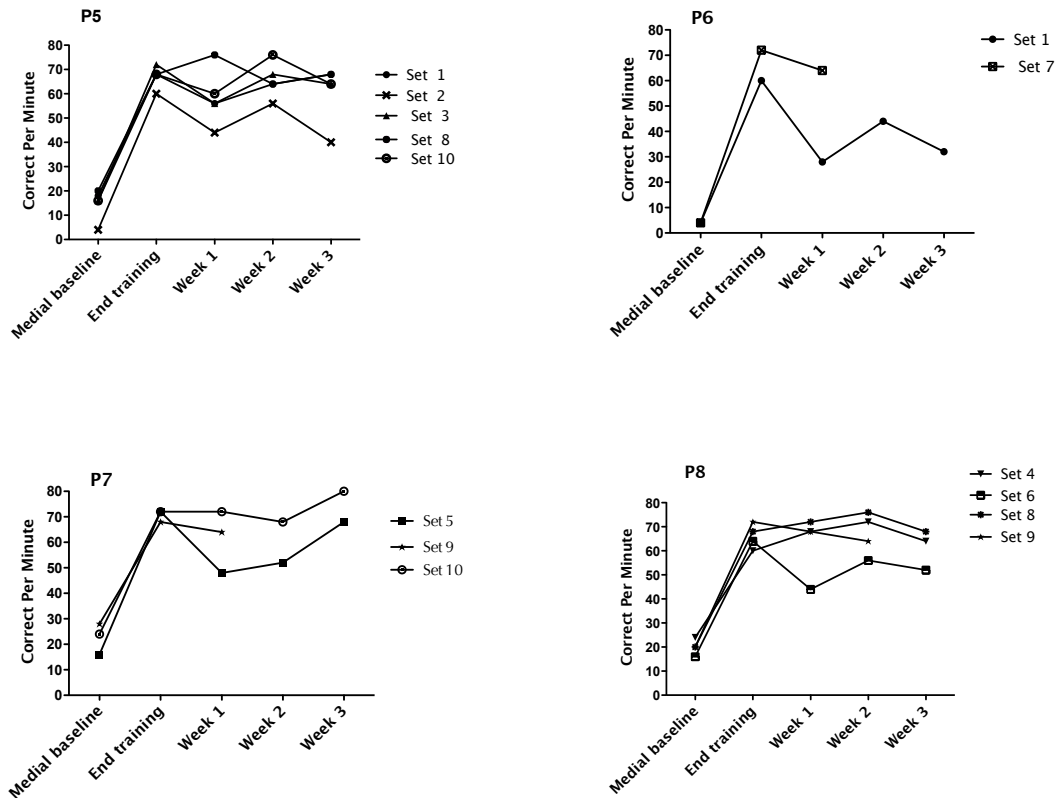


Figure 33. Experiment 11, Condition 2: maintenance data displaying performance in high frequency words at median baseline, intervention end and three post-intervention checks: 1 week; 2 weeks, and 3 weeks.

5.2.3. Results and Discussion

Experiment 11 was conducted with participants from two Senior Infant classes in a rural DEIS Gaeltacht school. In January 2013, all participants were administered three screening assessments (i.e., the YARC LSK and EWR subtests, and the WIAT-II PWD subtest). These screening instruments identified students in need of additional instruction in letter sounds, decoding and word reading; and participants were randomly assigned to receive the PT intervention in either decoding (segmenting and blending; Condition 1) or high frequency words (Condition 2). Although participants in the current experiment demonstrated difficulties across all assessments during the universal screening phase, they did not participate in more than one experiment.

Similar to Chapter 3, the current chapter also provides an examination of the effects of the PT intervention on measures of letter sound fluency (CBM LSF); nonsense word decoding (DIBELS NWF); and Real Word lists (RW). These findings are presented as generalised outcomes in reading measures in terms of standard score gains, and pre- post-test gains in letter sounds, and real and nonsense words read correctly.

5.2.3.1. Generalised Outcomes in Reading Measures

Table 17 shows the pre- to post-test standard score ratio outcomes in the YARC Letter Sound Knowledge (LSK-SSRG) subtest, the YARC Early Word Reading subtest (EWR-SSRG), and the WIAT-II PseudoWord Decoding (PWD-SRG) subtest. Participants trained in decoding words on average gained 10.4 SSRG in letter sounds (range 7-16.3); 1.9 SSRG in word reading (range.4-2.6); and 2.8 SSRG in pseudoword decoding (range .5-3.7). Table 17 also shows that participants trained in high frequency words on average gained 3.5 SSRG in letter sounds (range 0-8.3); 3.6 SSRG in word reading (range -3-10.6); and -1.3 SSRG in pseudoword decoding. Participants trained in decoding achieved these outcomes in approximately 2.3 hours of intervention time (range 1.9-2.7 hours); and, those trained in high frequency words achieved the observed outcomes in approximately 1.9 hours of intervention time (range .9-2.7 hours).

Table 17

Outcomes in Letter Sounds (LSK), Word Reading (EWR), and Pseudoword Decoding (PWD) for Participants That Received the PT Intervention in Decoding Words (Condition 1) and High Frequency Words (Condition 2).

	Total Hours Intervention	LSK-SSRG	EWR-SSRG	PWD-SSRG
Participants that Received the PT Intervention in Decoding Words				
P1	2.5	10.7	2.4	3.6
P2	2.7	7	.4	3.4
P3	2.2	7.4	2.3	.5
P4	1.9	16.3	2.6	3.7
Average Decoding	2.3	10.4	1.9	2.8
Participants that Received the PT Intervention in High Frequency Words				
P5	2	5.6	6.6	-2.5
P6	2	0	-3	-2.6
P7	1.1	0	10.6	-.9
P8	1	8.3	0	1
Average HFW	1.5	3.5	3.6	-1.3

The three additional fluency measures implemented at pre- and post-testing included the CBM Letter Sound Fluency (LSF) subtest; the recoded DIBELS NWF subtest; and, the manually constructed Real Word (RW) list (described in Chapter 3). Table 18 depicts participants' letter sounds frequencies at pre-test (LSF Pre-Test) and post-test (LSF Post-Test); real words read correctly at pre-test (RW-WRC Pre-Test) and post-test (RW-WRC Post-Test); and, nonsense words read correctly at pre-test

(NWF-WRC Pre-Test) and post-test (NWF-WRC Pre-Test). Results are described using the average correct per minute (CPM) gains from pre- to post-test, and resulting fold change in average performances. Table 18 shows that participants gained on average 22.5 (-2-39) letter sounds CPM; 1.5 (range -4-4) real words segmented and blended CPM; and, four (range 0-6) nonsense words segmented and blended CPM. These gains represent an average 1.5-fold increase in correct letter sounds, an average 1.2-fold increase in number of real words correctly decoded, and a 4-fold increase in number of nonsense words correctly decoded.

Table 18

Outcomes in Words Read Correctly (WRC) on the Recoded DIBELS Nonsense Word Fluency Subtest (NWF) and Real Word (RW) Probe for Participants That Received the PT Intervention in Decoding Words (Condition 1) and High Frequency Words (Condition 2).

	LSF Pre-Test	LSF Post- Test	RW- WRC Pre-Test	RW- WRC Post- Test	NWF- WRC Pre-Test	NWF- WRC Post- Test
Participants that Received the PT Intervention in Decoding (Segmenting and Blending) Words						
P1	60	58	8	12	7	11
P2	53	92	4	4	0	5
P3	29	55	12	8	1	6
P4	36	66	0	0	0	6
Participants that Received the PT Intervention in High Frequency Words						
P5	29	58	4	4	0	0
P6	53	55	8	12	2	6
P7	29	65	8	12	1	5
P8	45	65	8	12	1	7
Averages	41.8	64.3	6.5	8	2.9	5.8

Partial and Complete Decoding: Generalisation Probes Across Experiments 8, 9, and 10.

Participants in both conditions of Experiment 11 were administered two types of generalisation checks (partial and complete decoding) across baseline and

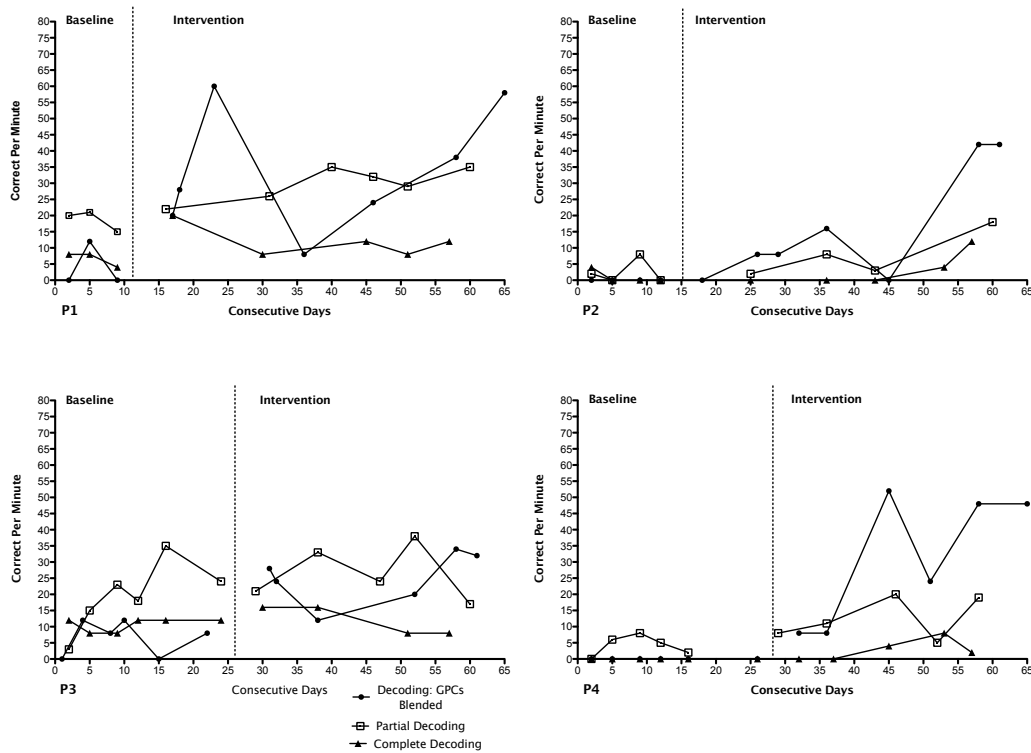
intervention phases The generalisation checks and form of reporting are identical to those described in Chapter 4. These probes were administered to evaluate the relation between increasing frequencies in decoding words/high frequency words on partial decoding (i.e., the number of nonsense word GPCs correctly blended), and complete decoding (i.e., the number of real words read correctly).

Figure 34 shows the generalisation probes for participants that received the PT intervention for decoding words (P1, P2, P3, and P4). Figure 34 shows that P3's performance in partial and complete decoding is similar across baseline and intervention conditions; therefore no generalisation gains are observed as a result of the intervention. Participants 2 and 4 showed zero complete decoding and limited partial decoding during the baseline phase, the intervention resulted in negligible complete decoding gains (see Figure 34). Gains in partial decoding were limited and delayed for both participants; this indicates that multiple intervention sessions were required to occasion generalisation to partial decoding. These gains for P4 are confounded, however, as there is a 10-day gap between the last baseline probe for nonsense words and the introduction of the intervention. Participant 1 demonstrated clearer generalisation gains in both partial and complete decoding (see Figure 34).

These results may be explained somewhat by the differential characteristics of the intervention stimuli and the generalisation stimuli. Specifically, the nonsense and real words used to measure partial and complete decoding contained the most frequent GPCs associated with the 26 letters of the alphabet, however the intervention stimuli contained only 18 of these letters. Therefore, encountering these untrained GPCs (eight in total) in the generalisation checks may have hindered decoding ability, or slowed the decoding response – resulting in labored decoding. In addition, the

MLGPCs targeted in the intervention did not feature in the generalisation checks; therefore generalisation gains may reasonably be limited.

Generalisation effects were evident on the pseudoword decoding subtest administered as a pre- post-test. Results showed that participants that received the PT intervention for decoding gained 2.8 SSRG on this subtest, compared to -1.3 SSRG in the HFW group. The pseudoword decoding subtest was untimed, and test stimuli contained words comprised of some of the MLGPCs targeted in the intervention. This meant that participants could take more time applying taught MLGPCs in an unfamiliar context (i.e., a pseudoword); indicating that the PT intervention in decoding does generalise to untaught words that contain the taught MLGPCs. Conversely, the HFW participants could not apply this skill and demonstrated a regression in standard scores.



Decoding GPCs Blended = Grapheme Phoneme Conversions (single and MLGPC)
 Blended Correctly
 Partial Decoding = Nonsense Word Letter Sounds Blended Correctly
 Complete Decoding = Real Words Read Correctly

Figure 34. Generalisation outcomes for participants that receive the PT intervention for decoding words. Generalisation checks for partial and complete decoding administered within and across baseline and intervention conditions.

Figure 35 shows the generalisation probes for participants that received the PT intervention for high frequency words (Condition 2; P5, P6, P7, and P8). Participants 5 and 7 demonstrated at or near zero complete decoding during the baseline phase. Figure 35 shows that performance in partial and complete decoding did not grow for P5, and increased moderately for P7 during the intervention phase. Figure 35 shows that P6 showed an ascending trend in complete decoding during baseline and this performance did not escalate during the intervention. However, for P6 partial decoding does increase in the intervention condition. Participant 8 did not show

significant gains in partial and complete decoding. However, a limited number of generalisation checks may make these results difficult to interpret (see Figure 35). Furthermore, because intervention was targeted at the level of the word (and not the phoneme), gains in decoding performance, were not anticipated. Decoding performances shown by participants during the generalisation phase may be explained by intervention at the word level.

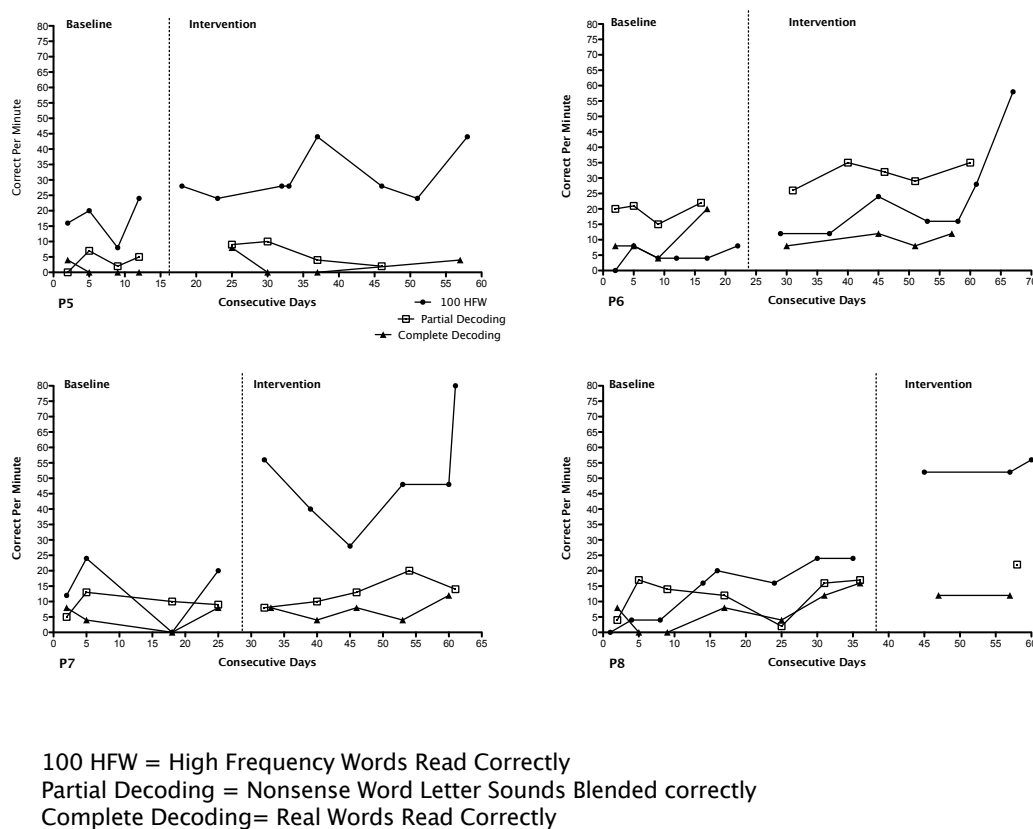


Figure 35. Generalisation outcomes for participants that receive the PT intervention for high frequency words. Generalisation checks for partial and complete decoding administered within and across baseline and intervention conditions.

Taken together, these results suggest that in order to achieve the greatest gains in complete decoding, the PT intervention should involve targeting the most frequent sounds from the 26 letters of the alphabet, before targeting decoding words with

MLGPCs. Therefore, decoding may be maximised by targeting CVC words before words that contain MLGPCs. These results also show that building frequencies in words did not result in generalisation to partial decoding, therefore both levels (word and phoneme) should be targeted for with the PT intervention.

5.3.1. Experiment 11: Conclusions

The purpose of Experiment 11 was to implement the PT intervention programme as a Tier 2 intervention for building frequencies in decoding words consisting of high frequency MLGPCs and the most high frequency words to performance standards with at risk Senior Infant readers. Additional objectives were to evaluate outcomes on reading measures and generalisation checks between participants that received intervention in either decoding words (Condition 1) or high frequency words (Condition 2).

The current results show that the PT intervention programme was an effective intervention for building frequencies in high frequency MLGPCs and high frequency words to performance standards (or reduced criteria). In regards to high frequency words, based on the treatment effects observed in the current experiment, it should be possible to teach 100 HFW in approximately 5 hours of intervention. The maintenance data presented show a small decrease by week 3 post-intervention, this may be due to lack of subsequent exposure/practice in other contexts such as reading in the classroom or independent reading. In regard to the high frequency MLGPCs, the 29 most frequent of these mappings could be taught in 4.9 hours of instruction. Furthermore, because the PT intervention programme targets the MLGPCs embedded in real words, it also provides a level of decoding practice that may facilitate proficiency in reading. This may explain the maintenance data differences between the two conditions. The HFW decrease contrasts with the increase in decoding that is

demonstrated three weeks post intervention. This finding suggests that that the decoding skill acquired may have been practiced in other contexts, and as such may have strengthened decoding behaviour.

No programme modifications were necessary across the participants in both conditions implemented in Experiment 11, despite the significant difficulties demonstrated by participants across a number of reading assessments during the universal screening phase. This suggests that the PT intervention may be easily applied in applied settings with same age students, i.e. the intervention may not need to be modified with other at risk readers.

One possible limitation may involve the lack of probes administered to assess generalisation to untaught words that contained the MLGPCs targeted in intervention, and to reading taught words in connected text. Generalisation checks were identical to those in previous experiments for purposes of comparison. While this yielded valuable information in terms of future intervention focus, it did not permit a comprehensive evaluation of generalisation effects within the current experiment.

A further limitation of the current experiment may be that all eight participants during the universal screening phase demonstrated deficits in decoding skills. However, due to time constraints 50% ($n=4$) of the participants received intervention in decoding. In addition, during Condition 1, all high frequency MLGPCs were not targeted during the intervention phase. Only 13 of the 29 high frequency MLGPCs were targeted as a result of time constraints within the setting.

This research did not compare the effectiveness of the PT intervention programme with other strategies such as incremental rehearsal or traditional flashcard drill methods. This was beyond the scope of the current research programme and

therefore its relative effectiveness to these methods is unknown. Future research is warranted in directly comparing the PT intervention with these or other methods.

5.4. Social Validity

Social validity was central to the research question to assess the acceptability and suitability of the Tier 2 intervention to Senior Infant students. Parental and teacher views were not solicited due to time constraints. The social validity measure (Musti-Rao, 2005) used with participants was identical to that described in Chapter 3 (see Appendix G). All participants reported that they liked taking part in the intervention. When questioned on what they liked best, two reported the treasure chest, two stated saying the sounds in words, one liked *“learning things you didn’t know before”*, one liked *“getting really big scores”* and one reported to like using the SCC. In regard to being part of the group, earning rewards, and learning new skills, seven participants rated liking it “a lot”, and one “a little”.

Seven participants reported that they had learned important things in the intervention (one did not). When probed specifically on the most important things learned, three commented on learning words, one on reading, one on spelling, and one on listening for sounds. Seven participants reported that what they learned helped them perform better work at school (six rated “a lot”; one rated “a little”); one did not. All participants reported that they used the skills learned in the intervention. When queried on the application of such skills, four commented on reading in school, and two on reading at home, two comment on reading books, one stated *“it helps to remember words I learned”*, and another that *“it helps you in your brain”*. Seven participants wanted the programme to have lasted longer, and one did not. Of the seven participants that wanted the programme to last longer, one wanted it to last for

“10 more years”, one said “lots and lots and lots longer”, another for a “really long time”, one was more specific “until 2nd class”, and the remaining three did not know how much longer they wanted it to last for.

Chapter 6: General Discussion

The primary aim of the current thesis was the implementation and evaluation of a Precision Teaching (PT) intervention programme as a Tier 2 Response to Intervention, targeting fluency in foundational reading skills with at risk Senior Infant readers. The PT intervention programme combined: (a) fluency intervention with; (b) progress monitoring, and (c) decision rules for intervention modifications.

The PT intervention programme described in the current thesis provides an account of the first application of PT as a Tier 2 intervention. The programme displays a hybrid model of Tier 2 support as it utilises a standard intervention protocol to target fluency in foundational reading skills, combined with performance data to systematically modify subsequent intervention where decision rules indicated this was necessary. The PT intervention programme employed across each of the experiments reported, was grounded in a causal model of reading development based on the stimulus control paradigm and IH, and used behaviour analytic strategies for effective instruction to promote core reading skills.

PT was selected as the intervention as it directly targets fluency and accelerates rates of learning; identified as well founded intervention foci for at risk kindergarten readers (Burke et al., 2010; Skinner, 2008). In addition, the strategies employed in the PT intervention (e.g., goal specificity, feedback) are effective instructional ingredients for targeting at risk and struggling readers (Bramlett et al., 2010; Konrad et al, 2011; Mellard et al., 2010). The foundational reading skills targeted for intervention across the experiments, are recognised as critical skills for struggling readers (phonemic awareness and letter sounds; Snowling & Hulme, 2010),

and to optimise early reading instruction (decoding and high frequency words; Solity & Vousden, 2009; Vousden, 2008).

The current research aimed to investigate the effectiveness and efficiency of the PT intervention programme using SCED and pre- post-test standard score changes. SCED can demonstrate students' response to intervention (Riley-Tilman & Burns, 2010; Barnett et al., 2004), and are an effective methodology for establishing educational interventions (Stoiber & Kratochwill, 2000). In addition, because the goal of Tier 2 is to help students close the gap with average performing peers, norm-referenced standardised assessments were administered as pre- and post-tests. Gains in standard scores on these assessments indicate that a student's performance approximates that of his or her average performing peers, or that the student is "closing the gap" (Torgesen, 2005). Moreover, SSRG (ratio gains) were computed to provide a measure of instructional efficiency, and to permit comparisons with interventions of differing durations.

The current thesis aimed to precisely report the instructional intensity of the PT intervention programme (i.e., total intervention duration, number of intervention sessions, and average duration of intervention sessions). Reporting instructional intensity in this way permitted exact quantification of performance data, by providing a measure of instructional efficiency. In addition, because intervention effectiveness and efficiency should be considered in terms of its effect on the IH (Cates et al., 2010; Daly et al, 2007), the research aimed to document maintenance and generalisation effects. Maintenance data ranged from three weeks to nine months post-intervention, thus revealing the long-term effects of building frequencies in foundational reading skills. Generalisation data were collected within and across experimental conditions,

and permitted evaluation of global reading outcomes that resulted from the PT intervention.

6.1. Summary of Experiments

6.1.1. Experiment 1 (Chapter 2): Letter Sounds

The combined findings of the multiple-baseline design, the pre- post-test outcomes, and post-intervention maintenance effects reported in Experiment 1, suggest that the PT intervention programme was both an effective and efficient form of intervention. One-to-one daily intervention sessions lasted on average nine minutes. In an average of 2.4 hours of intervention time (range 1.3-3.9 hours) over approximately 18 intervention sessions (range 9-24) participants achieved performance standards or reduced criterion in letter sounds (M=62; range 40-76 CPM). These intervention outcomes were largely maintained until the end of the school year and at 38 weeks post-intervention showed a 0.9-fold decrease (i.e., 57 CPM (range 40-72)).

Participants also on average achieved rate gains of 31 letter sounds (range 24-44) on the CBM letter sound fluency subtest, and an average standard score ratio gain of 7.3 (range 6.3-9.2) in the YARC letter sound knowledge subtest. Post-testing revealed that performance across all seven participants was in the average range of scores in letter sounds (at pre-testing five were shown to have “severe difficulty” and two “below average”).

6.1.2. Experiment 2 (Chapter 2): Letter Names

Experimental control was less clear for the multiple-baseline design described in Experiment 2. Results of this experiment showed differential effects of the intervention across participants. The intervention was terminated for two participants

at reduced criterion rates due to limited growth in performance and despite a number of programme modifications. One-to-one daily intervention sessions lasted on average nine minutes. In an average of 2.3 hours of intervention time (range 0.4-3 hours) over approximately 10 intervention sessions (range 3-15) three participants achieved performance standards ($M = 77$), and two achieved reduced criterion ($M = 48$) in letter names ($M = 66$; range 48-80 CPM). Differential outcomes of maintenance of intervention effects were also shown across participants. However, findings from Experiment 2 showed a 1.2-fold increase in correct responding at the 38 weeks post-intervention check.

Participants on average achieved rate gains of 25 letter sounds CPM (range 7-34) on the DIBELS letter name fluency subtest. At the post-test phase four of the five participants were considered “low risk” (at pre-test three were shown to be “at risk”, and two “some risk”) in letter names.

6.1.3. Experiment 3 (Chapter 2): Sound Isolation

The multiple-baseline with multiple probes design employed in Experiment 3, demonstrated experimental control through multiple replications of effect in the intervention conditions, for initial and final phoneme. This experiment showed that the intervention was effective for all five participants. The combined findings of the multiple-baseline design, the pre- post-test outcomes, and post-intervention maintenance effects suggest that the PT intervention programme was an effective and efficient intervention for building frequencies in initial and final phonemes.

One-to-one daily intervention sessions lasted on average eight minutes. In an average of 1.9 hours of intervention time (range 1.7-2 hours) over approximately 14 intervention sessions (range 9-18) participants achieved performance standards in isolation of initial phoneme ($M = 25$; range 24-26 CPM), and in isolation of final

phoneme ($M = 24$; range 20-26 CPM). These intervention outcomes were largely maintained until the end of the school year, and at 38 weeks post-intervention a 1.1-fold decrease in correct responding was shown. However, frequencies remained within performance standards. Experiment 3 also showed clear generalisation effects for two participants and a delayed effect for two additional participants. Importantly, this suggests that the PT intervention targeting initial phoneme isolation may generalise to isolation of the final phoneme in the absence of direct intervention.

Participants also achieved an average standard score ratio gain of 15.6 (range 12.8-17) in the YARC Sound Isolation subtest. At the post-test phase, performance across all five participants were shown to be in the average range of scores in sound isolation (at pre-test three were shown to have “severe difficulty” and two “below average”).

6.1.4. Experiment 4 (Chapter 2): Sound Deletion

The multiple-baseline with multiple probes design employed in Experiment 4 demonstrated experimental control through multiple replications of effect in the intervention conditions for compound words, and initial and final phoneme. This demonstrates that the intervention was effective for all participants. One-to-one daily intervention sessions lasted on average eight minutes. In an average of two hours of intervention time (range 1.8-2.3 hours) over approximately 16 intervention sessions (range 15-24) participants achieved performance standards in: sound deletion of compound words ($M = 20$; range 18-20 CPM); deletion of initial phoneme ($M = 25$; range 24-26 CPM), and deletion of final phoneme ($M = 26$; range 26-26 CPM). These intervention outcomes were largely maintained until the end of the school year, and at 38 weeks post-intervention a 4-fold decrease in correct responding was shown.

Pre-to-post test results from Experiment 4 were less consistent across participants. Despite evidencing large standard score ratio gains for most participants two of the four did not move into the “average” range at post-testing. All four participants had reached criterion on frequency aims during the intervention, notwithstanding zero or near zero levels of correct responding pre-intervention. However, it is possible that the sound deletion subtest may have placed a floor effect on performance due to a limited number of items (12 items with three devoted to the medial phoneme, an untargeted skill in the current study).

Experiment 4 showed that targeting compound words did not readily generalise to increases in correct deletion of syllables or initial phonemes. However, for the two participants that received the intervention in compound words and multisyllabic words, generalisation to final and/or initial phoneme was demonstrated. Failure to demonstrate generalisation from compound words to initial/final phoneme suggests that the smaller unit of the syllable sets the occasion for generalisation to the level of the phoneme. However, while this is an interesting theoretical finding the current results do not support targeting multisyllabic words prior to targeting phonemes.

6.1.5. Experiment 5 (Chapter 3): Letter Sounds

The combined findings of the multiple-baseline design, the pre- post-test outcomes, and post-intervention maintenance effects reported in Experiment 5 suggest that the PT intervention programme was an effective and efficient intervention. One-to-one daily intervention sessions lasted on average 6.6 minutes. In an average of 1.5 hours of intervention time (range 1.5-2.4 hours) over approximately 12 intervention sessions (range 5-18) participants achieved near performance standards in letter sounds ($M = 67$; range 64-68 CPM). These intervention outcomes

were largely maintained until the end of the school year, and at 28 weeks post-intervention, a 0.9-fold decrease in correct responding was shown.

Participants who received the intervention in single letter sounds on average achieved rate gains of 35 CPM (range 20-56) on the CBM letter sound fluency subtest. An average SSRG of 15.8 (range 4.1-35.6) per hour of intervention in the YARC letter sound knowledge subtest was also shown. At the post-test phase, three participants were categorised as “average”, one as having a “severe difficulty”, and one as “below average” in letter sounds (at pre-test all five participants were shown to have a “severe difficulty”).

6.1.6. Experiment 6 (Chapter 3): Multi Letter Grapheme Phoneme Conversions (MLGPCs). Experiment 6 showed that all participants maintained or exceeded a x2 celeration for the duration of the intervention. Due to time constraints within the school setting, the intervention was terminated at reduced criteria for all participants after approximately six intervention sessions (range 4-7 sessions). Therefore, it is unknown whether higher frequencies would have been achieved if the intervention continued. However, the frequency aims achieved by participants during the intervention were largely maintained up to 28 weeks later. While such effects appear to be strong, the functional use of targeting MLGPCs in isolation for decoding skills was questioned and not targeted in subsequent experiments.

For those participants trained in single letter sounds and MLGPCs ($n = 4$), an average SSRG of 20.6 (range 17.1-25.8) per hour of intervention in the YARC letter sound knowledge subtest was also shown. At the post-test phase, three participants were categorised as “average”, and one was “excellent” in letter sounds (at pre-test three were shown to have a “severe difficulty”, and one was “below average”).

6.1.7. Experiment 7 (Chapter 3): Blending Sounds

The combined findings of the multiple-baseline design, the pre- post-test outcomes, and post-intervention maintenance effects reported in Experiment 7 suggest that the PT intervention programme was an effective and efficient intervention. One-to-one daily intervention sessions lasted on average 7.5 minutes. In an average of 0.8 hours of intervention time (range .6-1.3 hours) over approximately seven intervention sessions (range 5-9) participants achieved performance standards in blending sounds (M = 12; range 12-12 CPM). These intervention outcomes were largely maintained until the end of the school year, and at 28 weeks post-intervention showed a 1.4-fold increase in correct responding.

Participants also on average achieved a standard score ratio gain of 8.5 (range 5.9-11.8) in the CTOPP Blending Words subtest. At the post-test phase, three participants were categorised as “average” and one as “above average” in blending sounds (at pre-test one was identified as “very poor”, one as “poor”, and two as “below average”).

6.1.8. Experiment 8 (Chapter 4): Letter Sounds

The combined findings of the multiple-baseline design, the pre- post-test outcomes, and post-intervention maintenance effects suggest that the PT intervention programme was an effective and efficient intervention. One-to-one daily intervention sessions lasted on average 5.6 minutes. In an average of 1.2 hours of intervention time (range .5-1.5 hours) over approximately seven intervention sessions (range 4-12) participants achieved performance standards in letter sounds (M = 82; range 72-88 CPM). These intervention outcomes were largely maintained 14 to 15 weeks post-intervention.

Participants in this experiment also achieved an average rate gain of 51 letter

sounds CPM (range 27-77) on the CBM letter sound fluency subtest, and an average standard score ratio gain of 14 (range 0-41.7) in the YARC letter sound knowledge subtest.

6.1.9. Experiment 9 (Chapter 4): Decoding Words

All participants in Experiment 9 maintained or exceeded a x2 celeration for the duration of the intervention. Due to time constraints, only one set of MLGPCs (/oo/, /ay/, and /ch/) was targeted, and subsequently, intervention was terminated after approximately six intervention sessions (range 4-7 sessions). In addition, maintenance data were not collected for this experiment.

Results showed consistent performance patterns within baseline and intervention conditions across all four participants, and three strong replications of effect in the intervention condition (two participants were added concurrently). This suggests that pre-intervention, participants could not decode whole words containing the target phonemes, and that the intervention resulted in accurate and speedy decoding of words containing the target phonemes.

6.1.10. Experiment 10 (Chapter 4): Blending Sounds

The multiple-baseline design employed in Experiment 10 demonstrated that the PT intervention programme was effective for building frequencies in blending sounds to performance standards (12-16 words CPM) for both participants. One-to-one daily intervention sessions lasted on average 6.3 minutes. In an average of .8 hours of intervention time (range .5-1 hours) over approximately 7.5 intervention sessions (range 5-10) participants achieved performance standards in blending sounds (M = 15; range 14-16 CPM). Multiple programme modifications were required for one participant in this experiment and resulted in a response to the intervention. Maintenance data and post-test outcomes were only available for one participant and

were therefore not reported.

6.1.11. Experiment 11 (Chapter 5): Decoding Words and High Frequency Words

Experiment 11 included two conditions. Condition 1 examined the effects of the PT intervention programme on decoding words. The combined findings of the multiple-baseline design, the pre- post-test outcomes, and post-intervention maintenance effects reported in Condition 1 of Experiment 11, suggest that the PT intervention programme was an effective and efficient intervention. One-to-one daily intervention sessions lasted on average seven minutes. In an average of 2.3 hours of intervention time (range 1.9-2.7 hours) over approximately 19 intervention sessions (range 18-21) participants achieved performance standards (or reduced criteria) in decoding 2-4 phoneme words ($M = 92$; range 76-120 CPM). These outcomes were largely maintained three weeks post-intervention.

Condition 2 examined the effects of the PT intervention programme on high frequency words. The combined findings of the multiple-baseline design, the pre-post-test outcomes, and post-intervention maintenance effects reported in Condition 2 of Experiment 11 suggest that the PT intervention programme was an effective intervention. One-to-one daily intervention sessions lasted on average five minutes. In an average of 1.5 hours of intervention time (range 1-2 hours) over approximately 17 intervention sessions (range 13-21) participants achieved performance standards (or reduced criteria) in high frequency words ($M = 67$; range 76-120 CPM). These outcomes showed a slight decrease however, three weeks post-intervention.

The primary focus of each of the eleven experiments was to test the effectiveness and efficiency of the intervention programme on building frequencies in foundational reading skills to performance standards. Experiments 1-11 demonstrate

that universal screening was successful in identifying students in need of additional instruction in foundational reading skills, and that the PT intervention provided was effective in building frequencies in those skills to performance standards (or reduced criteria in some instances). Effective interventions for struggling readers should accelerate rates of learning (Skinner, 2010). Experiments 1-11 established that the PT intervention programme was successful in accelerating rates of learning in short timeframes, highlighting its efficiency as a potential Tier 2 intervention programme.

The experiments in the current thesis aimed to accelerate rates of learning regardless of initial performance levels in each of the foundational skills targeted for intervention. The participants identified for intervention demonstrated differing accuracy levels at universal screening points. For example, Experiment 1 (letter sounds) showed that despite varying levels of accuracy (pre-test average 44% accuracy; range 0%-69% for the 26 letter sounds) performance standards were attained by five of the seven participants (P1 and P4 demonstrated zero accuracy at pre-test and the intervention was terminated at 52 CPM due to time constraints). Large rate gains were observed across the three letter sounds experiments (Experiments 1, 5, and 8). In Experiment 1, the average total pre- post-test gain on the LSF CBM was 31 CPM (range 24-44); in Experiment 5 the gain was 42 CPM (range 29-56), and Experiment 8 this was 51 CPM (range 27-77). According to the CBM LSF technical manual the standard rate of growth is one letter sound per week at kindergarten; the current study demonstrated rates of growth on the LSF measure of on average 5.7 letter sounds per week of intervention. Effectiveness and efficiency were also demonstrated with large SSRG in outcome measures of the skills targeted for the majority of participants as a result of the intervention.

The experiments also sought to evaluate if the PT intervention programme was effective in establishing sublexical fluency. This was primarily evaluated through evaluation of the overall effects of the intervention on maintenance and generalisation outcomes. Maintenance data revealed long-lasting effects of the PT intervention over significant periods of time (up to nine months post-intervention). Such findings provide a behavioural marker of fluency (maintenance), and also provide further evidence of the effectiveness and efficiency of the intervention. Evaluation of maintenance data was a central feature in the overall analysis of the PT intervention programme, as the effectiveness of an intervention must be evaluated in terms of its effect on the entire IH (Cates et al., 2010; Daly et al, 2007). Generalisation is considered the final stage of the IH, and should be the distal goal of an intervention targeting specific core skills (Snowling & Hulme, 2010) - this was the case with the current research.

The current thesis was the first application and evaluation of targeting fluency in foundational reading skills within a Tier 2 intervention. Therefore, to precisely evaluate the effects of the intervention, generalisation to reading words and connected text was not directly programmed. Such adjustments may have confounded outcomes on the global reading measures employed during the post-testing phase. However, as generalisation to word reading was the distal goal of the PT intervention, additional reading measures were administered as pre- and post-tests and as generalisation checks within and across experiments. These results are summarised in the following section.

6.2. Summary of Generalised Reading Outcomes Across Experiments

The current research provides an account of a Tier 2 intervention programme that directly matched the focus of Tier 2 instructional provision to universal screening outcomes. This meant that participants received the PT intervention for foundational reading skills on an identified need basis, rather than each participant receiving the same intervention concurrently across skills. It also meant that some participants took part in more than one experiment; i.e., 21 of the 36 participants (i.e., 58.33 %) took part in two or three experiments. This can be considered both a strength and a weakness of the current research. Because the majority of participants were part of multiple experiments, it was challenging to disentangle the relative influence of the foundational skills targeted on overall reading development. Although this presented as a challenge, it also provided the opportunity to pinpoint concurrent gains in partial and complete decoding through performance observations within and across multiple experimental conditions.

The generalisation checks of partial and complete decoding as a result of the intervention, also permit greater confidence in interpreting the pre- post-test standard score gains on word reading measures, that were observed across all experiments. The standardised assessments functioned to measure overall reading gains; and the generalisation checks provided evidence that decoding emerged as a direct result of building frequencies in foundational reading skills.

6.2.1. Experiments 1, 2, 3, and 4 (Letter Sounds, Letter Names, Sound Isolation, and Sound Deletion)

Participants ($n=10$) who received intervention in the areas of letter sounds, letter names, sound isolation and sound deletion across Experiments 1-4 showed an average of 4.4 (range 0-22) standard score gains in the YARC Early Word Reading

measure and an average standard score ratio gain of one (range -1.35-4.8). Specifically, these participants gained approximately one standard score in word reading, as a ratio to hours of intervention received across Experiments 1, 2, 3, and 4. Four participants, however, showed no standard score gains (Abe, Con, Gus, and Ike) in word reading following intervention in these four foundational reading skills. Five participants also gained an average of 1.2 correctly decoded nonsense words, however, the remaining five showed no gains in this area (Jay, Kim Ella, Gus, and Con).

The inconsistency in gains observed across participants in Experiments 1-4, on the measures of word reading and nonsense word decoding, led to a re-evaluation of the target skills for the next phase of the research (Experiments 5-7). Letter names and the phonemic awareness skills of sound isolation and sound deletion were not subsequently targeted as they were considered not to be aligned to the core school curriculum. Therefore, limited generalisation to word reading may be observed. The focus of the intervention in the next phase was on letter sounds, multi-letter grapheme phoneme conversions (MLGPCs), and blending sounds as these skills were considered more closely aligned to the core curriculum (e.g., *Jolly Phonics* – a synthetic phonics approach) and thus to decoding words.

6.2.2. Experiments 5, 6, and 7 (Letter Sounds, MLGPCs, and Blending Sounds)

Participants ($n=10$) who received intervention in the areas letter sounds, MLGPC's and blending sounds across Experiments 5, 6 and 7, showed an average of 11.2 (range 0-27) standard score gains in the YARC Early Word Reading Measure, and an average standard score ratio gain of 6.9 (range 0-14.3). Overall, participants gained approximately 6.9 standard scores in word reading, per hour of intervention

received across Experiments 5, 6, and 7. At the post-test phase, five participants were categorised within the “average” range of scores in word reading; however, three remained in the “severe difficulty” category (at pre-test seven were identified in the “severe difficulty” category, and two as “below average”).

In terms of other outcomes, the average participant gained 42 (range 29-56) letter sounds CPM; 11.3 (range 2-18) real words segmented and blended CPM; and nine (range 0-19) nonsense words segmented and blended CPM. These outcomes show that in approximately two hours (range 1-2.7) of intervention, over 18 (range 12-23) intervention sessions, the average participant gained 6.9 standard scores in word reading, 42 letter sounds CPM, 11.3 real words blended CPM, and nine nonsense words blended CPM.

6.2.3. Experiments 8, 9, and 10 (Letter Sounds, Decoding Words, and Blending Sounds)

Participants ($n=8$) who received intervention in the areas letter sounds, decoding words and blending sounds across Experiments 8, 9 and 10, gained an average of 2.9 (range -10-9) standard score gains in the YARC Early word Reading subtest and an average standard score ratio gain of 2.4 (range -10-6.7). In terms of pseudo-word reading, participants gained an average of 7.9 (range 3-18) standard score gains in the WIAT PseudoWord Decoding subtest; and an average standard score ratio gain of seven (range 3-13). The average participant also gained 8.2 (range 8-16) real words segmented and blended CPM on the experimenter generated word list; and nine (range 7-13) nonsense words segmented and blended CPM on the recoded DIBELS NWF subtest. These outcomes show that in approximately one hour (range 0.5-1.8) of intervention, over 12 (range 6-16) intervention sessions, the average participant gained 2.9 standard scores in word reading, seven standard scores in

pseudoword decoding, 8.2 real words blended CPM, and nine nonsense words blended CPM across Experiments 8, 9, and 10.

6.2.4. Experiment 11 (Decoding Words and High Frequency Words)

Participants ($n=4$) who received intervention in decoding words on average gained 10.4 SSRG in letter sounds (range 7-16.3); 1.9 SSRG in word reading (range .4-2.6); and 2.8 SSRG in pseudoword decoding (range .5-3.7). The average participant also gained 23.3 (-2-39) letter sounds CPM; 0 (range -4-4) real words segmented and blended CPM; and 2.2 (range -2-5) nonsense words segmented and blended CPM.

Participants ($n=4$) who received intervention in high frequency words on average gained 3.5 SSRG in letter sounds (range 0-8.3); 3.6 SSRG in word reading (range -3-10.6); and -1.3 SSRG in pseudoword decoding. Participants also gained on average 22.5 (-2-36) letter sounds CPM; 1.5 (range -4-4) real words segmented and blended CPM; and 4 (range 0-6) nonsense words segmented and blended CPM.

6.3. Implications of the Generalised Reading Outcomes Resulting from the PT Intervention Programme

The combined generalisation outcomes show that the PT intervention programme resulted in large gains in word reading and pseudo-word decoding, within and across experiments, and in a pre- post-test context. These findings have important theoretical and practical implications relating to the role of sublexical fluency in reading development, and consequently, why this needs to be a focus in early reading intervention with Senior Infant (kindergarten equivalent) students.

6.3.1. Pre- Post-Test Outcomes – At Risk Readers “Closing the Gap” with Average Peers

The PT intervention was designed for at risk Senior Infant participants to catch up with their average performing peers on standardised measures of reading. Standard score gains have been described as an excellent metric to evaluate if students are closing this gap (Torgesen, 2005). Moreover, standards score ratio gains provide an efficiency metric as they account for the amount of intervention time that the observed gains were achieved in. The current thesis reports data from 13 multiple baseline SCED for at risk readers in the age range of 5.5-7.3 years ($M = 6.3$). The average intervention time was 2.3 hours, and SSRG for word reading on the YARC EWR subtest averaged at 3.4 SSRG (range -10-14.3). Large SSRG in word reading were achieved for most participants; this demonstrates both the effectiveness and the efficiency of the intervention. In very short timeframes the PT intervention programme was shown to have helped at risk readers “close the gap” with average performing peers. These outcomes highlight the potency of the current PT intervention programme as an effective and efficient early intervention for at risk readers.

In addition, because the SSRG describes standard score gains as a function of intervention time, this statistic permits outcome comparisons between interventions of differing length (Torgesen, 2005). The effectiveness of the PT intervention is further demonstrated through a comparison of the SSRG in word reading found in the current research to other prevention research (i.e., typically developing children) and to remediation research (i.e., students with dyslexia). A summary of SSRG for typically developing children across a range of high quality studies is currently not available (Savage et al., 2009). However, a small number of studies have been identified that

report gains in the range of -.17-2.6 (Hatcher et al., 1994; McGuinness et al., 1996; Hatcher et al., 2004; Savage et al., 2009). Of these studies, three report SSRG in decimal figures; McGuinness and colleagues (1996) provide the only study reporting SSRG comparable to the current research. The large SSRG reported by these authors, however, have not been replicated since. Remediation studies report even lower SSRG outcomes. For example, Torgesen (2005) reviewed 14 intervention studies for students with dyslexia in the age range of 8.9-12.10 years. Average intervention time was 65.5 hours, and SSRG for word reading averaged at 0.2.

The results of the current research show that participants demonstrated large SSRG ($M = 3.4$, range -10-14.3) on the YARC Early Word Reading measure in very short intervention duration ($M = 2.3$ hours). Large SSRG were yielded across a number of different settings with a diverse student population, demonstrating a consistent replication of intervention effectiveness. This adds to a body of literature demonstrating that prevention, or early intervention, can yield greater gains than later remediation. Moreover, it demonstrates that the PT intervention specifically leads to greater gains than those generally obtained in prevention research (i.e., targeting typically developing students).

The gains observed in the current thesis however, may not be generalised to the wider population, as there was no control group with which to compare scores. Therefore it could be argued that these gains were the result of general classroom instruction and not the supplemental PT intervention programme. For this reason, generalisation checks administered within and across experiments helped elucidate the precise contribution of the PT intervention to overall reading development.

6.3.2. Generalisation Checks: Sublexical Fluency and Dysfluency Effects on Decoding Words

The generalisation outcomes observed across participants in the current thesis provides evidence that sublexical fluency, and conversely, dysfluency, has the potential to affect proficiency in decoding. The baseline conditions across experiments represent participants' performance in foundational reading skills as a result of general classroom instruction in reading (i.e., pre-intervention performance levels). The baseline conditions for the majority of participants illustrate that where frequencies in letter sounds and/or blending sounds are low, partial and complete decoding were also low (in many cases at or near zero levels). This shows that dysfluency (i.e., low levels of accuracy and speed) in the foundational skills of letter sounds and blending sounds limits decoding ability. These results provide direct behavioural evidence corroborating findings from statistical modeling research (e.g., Burke et al., 2010; Hudson et al., 2010); specifically, that dysfluency in sublexical skills will impede decoding and ultimately competence in word reading.

Conversely, in the intervention conditions across experiments, generalisation outcomes illustrated that increasing frequencies in these core skills fosters partial and complete decoding for some, but not all at risk readers. For some participants, merely increasing frequencies in letter sounds to performance standards occasioned generalisation to partial and complete decoding. This pattern, however, was inconsistent across participants, and may be due to differential skill deficit patterns exhibited by at risk Senior Infant/kindergarten readers. For example, outcomes for participants with difficulties in both letter sounds and blending sounds indicated that both foundational skills required targeting with the PT intervention to observe generalisation to decoding.

These findings indicate the importance of accurate identification of students' individualised instructional needs and the focus of Tier 2 intervention. The universal screeners must identify *all* core reading skill deficits, and subsequent Tier 2 provision must provide a match to these screening outcomes. Across all settings participants demonstrated varied universal screening outcomes, suggesting differential skill deficits that may impede progression in regular classroom instruction. If students' idiosyncratic skill deficits are not addressed, it is likely that progress will be limited. The generalisation outcomes demonstrated by participants exhibiting differential foundational skill deficits, illustrate that skill specific fluency training in the form of the current PT intervention programme, is a worthy focus with at risk Senior Infant/kindergarten readers.

A further observation with regard to generalisation outcomes was the limited generalisation effects to partial and complete decoding of real and nonsense words for participants who received the intervention for decoding and high frequency words (Experiment 11, Chapter 5). Limited generalisation effects were hypothesised to result from the linguistic unit targeted with the PT intervention. That is, targeting decoding words that contained MLGPCs before letter sounds in isolation resulted in a number of single letter sounds (eight of the 26 most frequent sounds in the alphabet) being omitted from the intervention stimulus materials. Thus, the eight untargeted letters were not recognised automatically during the generalisation checks, and may have hindered the emergence of decoding behaviour.

To address this issue, the PT intervention could perhaps first target single letter sounds in isolation, followed by decoding words with MLGPCs. Indeed, targeting the most functional foundational reading skills (letter sounds, blending sounds, decoding words comprised of the critical high frequency MLGPCs, and the

100 most frequent words) as an overall early intervention programme may best equip at risk Senior Infant/kindergarten readers for reading development. Such a programme would be delivered as a standard protocol, using performance data to systematically adjust the intervention to meet students' instructional needs.

In recognition that letter sounds and blending sounds are pre-requisites for decoding, a recommendation for early intervention may involve all students considered at risk, receiving the PT intervention. Students who do not show difficulty with blending sounds should reach the performance standard for this skill very quickly, and subsequent skills could be introduced systematically. For example, a subsequent skill to blending sounds could involve decoding CVC words, followed by progression to decoding words comprised of the critical MLGPCs. Next the 100 most frequent words could be targeted with the PT intervention. In addition, as generalisation to connected text reading is the distal goal; daily practice of decodable connected text would be a significant addition to the intervention package.

The current research consistently demonstrated that the PT intervention was effective in targeting these specific foundational skills for fluency in very short time frames, and despite low pre-intervention accuracy levels with some participants. Based on the current findings, the average participant reached performance standards in letter sounds ($M = 70$ CPM, range 40-88) in on average 1.5 hours of intervention time (range .5-3.9), and in blending sounds ($M = 13$ CPM, range 12-16) in on average .8 hours of intervention time (range .5-1.3). These outcomes suggest that fluency in foundational reading skills of letter sounds and blending could be established with at risk Senior Infant readers in an average of 2.3 hours of intervention time.

Furthermore, Experiment 11 demonstrated that the PT intervention resulted in participants gaining on average, 10 high frequency words CPM per week, in

approximately 30 minutes of intervention time. Based on this outcome, using the PT intervention programme daily to target words, automaticity in the 100 most frequent words could be achieved in 10 weeks, in approximately five hours of intervention time. Experiment 11 also showed that 11 of the most frequent MLGPCs could be effectively targeted for fluency within an average of two hours. Based on these findings, all 29 of the most frequent MLGPCs could be targeted in approximately 5 hours of intervention.

Using the proposed PT intervention package, early readers could achieve fluency in four core foundational reading skills: the 26 single letter sounds, blending phonemes into words, decoding words embedded with the 29 most frequent MLGPCs, and 100 high frequency words. These skills should subsequently enable students to automatically access over 50% of words encountered in text, and to decode the majority of decodable words in children's texts (Solity & Vousden, 2009; Stuart et al., 2003; Carnine et al., 1997). Importantly, this proposed PT intervention package targeting these specific four crucial foundational skills would take an estimated average of 12.5 hours intervention time to implement. Clearly, this represents an effective and efficient intervention for at risk Senior Infant/kindergarten readers. Considering the importance of success in reading by the first grade (Cunningham & Stanovich, 1997) and the importance of this for long-term reading achievement (Cunningham & Stanovich, 2001; Sparks et al., 2014), it is clear that the PT intervention programme holds significant potential to provide a good start in early reading.

6.4. The PT Intervention Programme as an Effective Means for Establishing Sublexical Fluency (Mastery) in Foundational Reading Skills

The proximal aim of the experiments reported in the current thesis was to increase frequencies in foundational reading skills; the distal aim was to establish sublexical fluency. The literature frequently refers to “fluency” in relation to word and connected text reading, and to “mastery” of foundational reading skills. Indeed, mastery of core skills is frequently described as a pre-requisite for reading acquisition, and conversely, failure to master core skills as a justification for reading intervention (Bramlett et al., 2010). Definitions of mastery vary, however, and can be dictated by curriculums of differing pedagogical persuasions. For example, in the traditional drill and practice for flashcards, mastery can be defined as four correct responses, and in incremental rehearsal it can be eight correct responses (Burns & Sterling-Turner, 2010).

Mastery is commonly understood as demonstration of the full command of a skill, or comprehensive knowledge in a subject matter. As a concept, masterful performance should be quick, smooth, accurate, and effortless performance (Binder et al., 2002). Therefore accuracy is a component of mastery; it is necessary - but not sufficient - to occasion mastery (Binder, 1988). For example, a student may accurately discriminate letter sounds, but do so in a slow manner, or may not be able to discriminate the same sounds under distracting conditions, or for prolonged periods of time. Consequently, measuring mastery with just accuracy (i.e., can the student perform the skill) captures just one aspect of this concept. Mastery is therefore more synonymous with fluency than accuracy.

The current research provides evidence that the PT intervention programme was effective in establishing fluency in foundational skills of reading. The

behavioural outcomes of fluency (and thus the true test for fluency) are Retention, Endurance, Application, and Performance Standards (Lindsley, 1992). Evaluation of maintenance data collected across prolonged periods of time (up to 38 weeks for some participants in the current research) indicates that the performance levels achieved in the intervention were largely retained. Generalisation outcomes described, provide evidence that the performance standards achieved through the PT intervention were sufficient to occasion application of foundational reading skills (letter sounds and blending sounds), to compound reading skills such as partial and complete decoding. The performance standards achieved across experiments suggest that the PT intervention programme was effective in occasioning accurate and fast responding in the foundational reading skills targeted. These performance characteristics indicate that the PT intervention programme was an effective Tier 2 intervention to target sublexical fluency with at risk readers.

Frequency criteria for interventions are often based on normative samples or the individual's current level of performance. The difficulty with normative approaches is that they do not guarantee important outcomes of practice, such as retention or generalisation (Johnson & Street, 2013). Martens and colleagues (2013) highlight the need for establishing functional fluency aims for decoding words to occasion generalisation to untrained words. The current research demonstrated that the performance standards employed were functional for maintenance outcomes; however their effects on generalisation outcomes were not systematically measured in terms of untrained words and connected text reading. The magnitude of generalisation outcomes may therefore have been greater if higher performance standards were achieved.

The performance standards adhered to in the current thesis were determined by a number of factors. For the phonemic awareness tasks (blending sounds, sound isolation, and sound deletion) the performance standards suggested by the commercial programme used in the intervention (Freeman & Hauhgton, 1997) were employed as intervention criteria. For high frequency words the performance standards recommended by Fabrizio and Moors (2003) were employed. Reduced performance standards were selected for letter sounds and names and for decoding words. This was due to a balance of practical and experimental considerations. The research was conducted in a school setting with very young participants under time constraints (i.e., the research focus was to target a number of foundational reading skills). In addition, participants demonstrated differing pre-existing accuracy levels, with some demonstrating at or near zero levels. Despite differential pre-intervention performance levels, all participants demonstrated large celerations as a result of the intervention, which translated to significant gains in letter sounds correct per minute evident by the 3rd training session. When performance levels approximated 70 CPM celeration was shown to maintain. Rather than keeping the focus of the PT intervention on letter sounds to achieve 80 CPM, additional foundational skills were subsequently targeted.

The intervention for letter sounds was terminated within the range of 68-90; the relation between frequency levels achieved and decoding suggest that the intervention criteria used in the current thesis may have a functional status (i.e., a fluency threshold). This implies that achieving such frequency levels may result in additional gains in related skills, such as decoding. This was evident in the generalisation outcomes. However, differential outcomes may have been observed if higher performance standards had been employed. For example, Evans and Evans (1985) demonstrated that training letter sounds to a criterion of 90 CPM resulted in

progress on decoding real and nonsense CVC words. Although such generalisation was observed in the current research at the lower threshold of 70 CPM, the functionality of this fluency criterion for Senior Infant/kindergarten students is the subject for future analyses.

6.5. The PT Intervention Programme as a Tier 2 Screening and Intervention

Protocol for At Risk Senior Infant/Kindergarten Readers

The current research aimed to identify students at risk for reading difficulties in foundational reading skills. From an initial overall sample ($n=111$) across the three settings in which the experiments were conducted, a total risk pool of 36 participants (32% of overall sample) was identified as presenting with reading difficulties. Of this sample, 18 participants spoke English as a first language (EL1), and 19 as a second language (EL2). Two EL2 participants had significant speech difficulties in their native language, and two EL1 participants had speech difficulties. All four attended speech and language therapy. Six participants originated from the Traveller Community (an indigenous minority group considered to be at risk for early school drop-out: Department of Education and Science; DES, 2005).

The PT intervention appeared to be similarly effective across participants regardless of language status, gender, or minority group membership. The risk pools identified within settings, however, were disproportionate to the initial sample size. For example, in Experiment 1 (rural DEIS school) 58% of participants screened for difficulties in letter sounds were selected for the PT intervention. Using the same criteria in Experiment 5 (urban DEIS school) 15% of participants screened for difficulties in letter sounds were selected for the PT intervention. This indicates that pre-existing levels of performance may have been dependent on setting characteristics

(access to resources for example) and consequently, the PT intervention may have demonstrated differential outcomes across settings. To further compound this issue, participant characteristics varied across urban and rural settings. Experiment 1 had a greater proportion of children from the travelling community (57%) and one EL2 participant (14%), whereas Experiment 5 accounted for less children from the travelling community (20%) and more EL2 participants (50%).

The research designs used, however, limit the disaggregation of such data to provide a meaningful analysis of any potential differential intervention effects across participant groups and/or settings. Such analyses would contribute to future development of the PT intervention programme, as participant characteristics play a role in special education identification and service delivery. EL2 students generally score lower on academic achievement tests than their EL1 counterparts (Hemphill & Vanneman, 2011), this must be taken into consideration by those implementing and evaluating reading intervention research. For interventions to be adopted with EL2 participants, the research must clearly demonstrate success with EL2s and explicitly state whom the findings can be applied to (Moore & Klingner, 2014). Members of the Traveller Community are three times more likely to have special educational needs than the general population, and in the general learning disability category the proportion is approximately six to seven times greater than the expected occurrence of this disability in the general population (Department of Education and Science; DES, 2005).

6.6. The PT Intervention Programme for Progress Monitoring and Decision

Making

Progress monitoring is a core component of Tier 2 interventions (Gersten et al., 2009). The PT intervention programme in the current research used the SCC as the progress-monitoring tool for the intervention. The SCC was used to graphically display performance, to set goals and to illustrate the mastery criterion for the participants. The SCC permitted level and rate of learning to be recorded and quantified for each participant. This allowed measureable performance data to form the basis for making decisions about adequate response to the intervention. Barnett and colleagues (2004) propose that RTI necessitates implementing a cohesive and sequenced set of procedures with the precise application of decision rules. The current thesis demonstrates how the PT intervention programme can provide the mechanism to blend a standard protocol with a problem solving approach to providing a hybrid Tier 2 intervention.

Programme modifications were based on celeration – a quantifiable learning trajectory. Using a x2 celeration as the parameter for decision making enabled the consistent application of decision rules within a highly structured intervention. Multiple reports and the practice guide for RTI (e.g., The report on the Presidents Commission on Excellence in Special Education, 2002; Gersten et al., 2009) recommend the use of scientifically validated progress monitoring tools to make instructional decisions regarding special education, and that continuation of such services incorporate the finding of dynamic assessment methods. Importantly, the SCC has the potential to meet these requirements. The current research is the first demonstration of utility of the SCC within Tier 2 intervention as a progress-monitoring tool.

Using the decision rules of the PT intervention programme in 67% of cases participants reached performance standards with no programme modifications. The remaining 33% required antecedent changes to accelerate learning rates on approximately the 8th day of intervention, and subsequent to this over half of those received consequence changes (in the form of the personal best component). However, the incidence of programme modification was disproportionately represented across the experiments as a function of the foundational skill targeted. Programme modifications were most commonly applied in relation to phonological/phonemic awareness skills: Experiments 3, 4, 7, and 10 evidenced modifications in 67% of cases. Letter sounds/names experiments (Experiments 1, 2, 5, 6, and 8) evidenced modifications in 30% of cases. Conversely, decoding/high frequency word experiments did not require any programme modifications (Experiments 9 and 11). The programme modifications implemented resulted in all participants achieving performance standards, or in a small number of cases, minimally reduced criteria that were determined to be satisfactorily functional.

The performance standards used in the current research functioned as CEMs (curriculum embedded mastery checks) that indicated when intervention should be terminated. This provided focus for the participant and functioned as the ultimate intervention goal upon which the “treasure chest” (reinforcement system) was accessed and an item selected. In addition, CBMs such as the DIBELS Nonsense Word Fluency subtest facilitated evaluation of the effect of the intervention on decoding. The combination of CEMs and CBMs is recommended for comprehensive progress monitoring as both reveal different intervention outcomes (Coyne et al., 2013; Gersten et al., 2009). The current thesis demonstrates how PT performance standards can be used as CEMs in conjunction with the SCC for accurate progress

monitoring, in combination with CBMs as application checks, to evaluate the overall effect of the PT intervention on untargeted reading skills. The current data suggest that increasing frequencies in letter sounds and blending sounds fosters partial and complete decoding. Limited generalisation gains on the CBM measures may quickly indicate a student in need of more intensive intervention.

While CBM as a progress monitoring procedure has a robust research base, the same cannot be said for the decision rules that accompany its use (Ardoin et al., 2013; Reed et al., 2014). In addition, the number of data points required to make decisions necessitates prolonged periods of time before data-based decision making can occur. The current thesis illustrates the utility and sensitivity of the SCC as a sensitive progress monitoring tool. The daily PT intervention yielded performance data on every school day. A precise celeration can be calculated with just five data points. The PT intervention programme used both paper and electronic versions of the SCC. The electronic version of the chart was used to quickly (and accurately) calculate celeration values within conditions (baseline/intervention), and to estimate variability of the data. In addition, daily performance data allowed programme modifications to be made in a timely fashion, therefore, precious instructional time was not wasted where a participant's learning rates were not accelerating at a 100% increase in performance (i.e., a x2 celeration). This ensured that participants' instructional needs were continually monitored with immediate modifications to the instructional process to ensure accelerated learning rates.

Moreover, a x2 celeration value is not a static value, but based on the student's own performance. This means that the trajectory is unique to the student, and takes into account the student's pre-existing levels of performance. This makes the learning goals achievable for individual students, while keeping a consistent intervention

performance criterion among students. For example, a student who shows a median baseline performance of two responses per minute would be expected to demonstrate a minimum of eight responses per minute after one week using a x2 celeration. Similarly, a student with 20 responses per minute would be expected to demonstrate a minimum of 40 responses per minute after one week using a x2 celeration. Therefore, the programme modification based on celeration outcomes adopted in the current thesis was always uniquely related to the student's idiosyncratic performance, yet was applied consistently across all of the participants within the intervention.

6.6.1. The PT Intervention Programme as a Dynamic Assessment

The report on the Presidents Commission on Excellence in Special Education (2002) report recommended the use of dynamic assessment to support the continuation of special education services. The current research illustrates the use of PT as a dynamic assessment of use within a Tier 2 intervention. Because PT incorporates error correction and feedback in-between each timed trial, performance on the subsequent trials can be considered as the students "response" to the intervention. Moreover, when participants did not respond to systematic error correction and feedback, and experienced goal failure over two days this indicated the need for programme modification. The programme modifications included a range of instructional supports used to strengthen the relation of the participant's response with the relevant academic stimuli. The participant's response to these modifications may also be considered part of the dynamic assessment. Participants who do not respond to modifications may be considered for Tier 3 intervention; however, this was beyond the remit of the current research.

6.8. Qualitative Outcomes of the PT Intervention Programme

It is likely that children find activities that they are good at naturally reinforcing - and will pursue them. Conversely, the struggling reader is less likely to seek out books or initiate homework and prone to avoid such activities. Cunningham and Stanovich (2001) reported the magnitude of reading deficits exhibited by struggling and proficient readers: by 5th class, a student at the 90th percentile of reading might read the same number of words in two days as a child in the 10th percentile reads in an entire year outside of school. A parental interview was incorporated in the current research and the majority of parents commented on an improved attitude towards school and homework, and an increased interest in books and reading for participants since taking part in the intervention. This provides a qualitative context to quantitative findings. Employing a measure of social validity showed results beyond quantitative changes in foundational reading skills and generalisation to word reading, allowing parents to report a change in attitude towards school and homework with their children, as well as an increase in reading books. The results of the social validity measure suggest that when struggling readers are provided with some success in reading skills, they actively seek out opportunities to use these skills.

The participant interview showed that participants favourably viewed the intervention and provided anecdotal support for the effectiveness of the strategies employed in the PT intervention. Participants typically provided positive comments on specific elements of the intervention such as the treasure chest, stickers earned, and “chart work” (i.e., the SCC). Such reinforcement is critical when attempting to increase reading achievement with at risk readers (Bramlett et al., 2010). Interestingly, many participants referred to liking the “chart work”, the chart readily

communicated goal attainment, and graphing data from the timed trials illustrated the participant's current and previous performance. This permitted increased goal specificity and allowed the participants to self-monitor progress, important features of effective interventions. Immediate graphing of performance allowed participants to make adjustments to their performance; this is an important feature for the struggling reader (Chan et al., 2014).

6.9. Limitations and Directions for Future Research

Limitations to the current thesis pertain to the characteristics of the PT intervention and its implementation. Additional methodological limitations were identified and a proposal for subsequent analyses is presented to address these issues. Based on these limitations, recommendations for future research are provided. Limitations concerning the PT intervention concern alignment to the core curriculum, targeting the foundational skills in isolation, programming for generalisation, the one-to-one instructional format for implementation, and the multi-components of the programme. Methodological issues relate to the general limitations associated with Single Case Experimental Designs (SCED), specifically, the generalisability of findings and the multi-layered nature of the data. A criticism of the current research may be that the majority of participants took part in more than one experiment, and the PT intervention programme used decision rules and applied programme modifications based on student performance data. To address these limitations, a multi-level model of analysis is proposed that aggregates the data from the SCED implemented in the current research.

A major advantage of the multilevel model is that it provides information about the average treatment effects across studies, as well as differences between

participants and studies in these treatment effects (Shadish & Rindskopf, 2007; Van den Noortgate & Onghena, 2003a, 2003b, 2008). The multi-level model could also estimate magnitude of treatment effect *before* any programme modifications were made, as well as estimating the maintenance of treatment effect across time. In addition, the flexibility of multilevel modelling can explain the variation between participants and studies by means of predictors at the different levels (e.g., age, gender, school type). Such models can also account for cross-classified data (i.e., when participants are included in multiple experiments). Therefore, multi-level modelling multiple-baseline across participants designs results in greater externally valid conclusions (Van den Noortgate & Onghena, 2003a, 2003b, 2008). This contributes to evidence-based research as general conclusions can be made about the effectiveness of a specific intervention that might be of interest to researchers, practitioners, and policy-makers.

The current thesis used SCED as the primary methodology to investigate the effect of the PT intervention programme. The daily data yielded through the PT intervention programme across the current research were easily incorporated into SCED, which are suitable for indexing a student's "response" to intervention (Daly et al., 2007). For these and other reasons, SCED are highlighted as suitable for use in educational settings (Riley-Tilman & Burns, 2010), and specifically multiple-baseline designs for evaluating literacy research (Axelrod, 1983; Kucera & Axelrod, 1995).

No research design, however, is a panacea for all research goals. Despite their benefits, SCED are not without their limitations; chief among these is that findings cannot be readily generalised to the wider population (i.e., external validity). The current thesis demonstrates multiple replications of intervention effect across a number of settings with different participants, thus providing some degree of

generalisability of findings (Kucera & Axelrod, 1995). The extent of generalisation to the wider population, however, is tentative and limited. To address this issue, Van den Noortgate and Onghena (2003a, 2003b, 2008) suggest combining several multiple-baseline across multiple participants to facilitate a multi-level analysis, thus increasing external validity. Extended simulation studies have validated this multi-level modelling method (Moeyaert, Ugille, Ferron, Beretvas, & Van den Noortgate, 2013a, 2013b). Future research could combine the multiple-baselines presented in the current thesis as a multi-level model, thus addressing many of the limitations identified.

For example, the multi-level method of analysis is favourable because it takes the multi-layered nature of the SSED data into account (Van den Noortgate & Onghena, 2003a, 2003b). Specifically within such a method, measurements are nested within participants, and participants in turn are nested within studies. This would have particular relevance with regard to the current research, as the studies were setting-specific and such a model would permit an evaluation that takes setting into account. An additional limitation identified is that many participants took part in a number of experiments, the multi-level method of analysis can account for cross-classified data i.e., when participants are included in multiple studies, as is the case in the current research.

The multiple-baseline designs implemented yielded a wealth of data on several students' response to the PT intervention programme, permitting assessment of the effects of this instructional approach across groups of students, and in identifying individual variation in response (Axelrod; 1983). A limitation of SCED, however, is that these differences cannot be precisely evaluated or disaggregated, leaving important questions concerning differential intervention effectiveness across

settings and/or student populations. A multi-level model could identify predictors such as gender, school, language status, and foundational skill targeted that may explain between study and within study differences. These findings would elucidate whether the intervention was more effective with certain student populations and/or specific foundational reading skills.

The staggered entrance of participants into the intervention condition in some of the multiple-baseline designs may also be considered a limitation. Because the research was carried out within the confines of the school week (Monday to Friday) the entrance of participants into the intervention was a balance of practical and experimental considerations. To demonstrate experimental control a minimum of three data points are required, similarly, students' response to the intervention was evaluated over two days. Once a participant had demonstrated response to the intervention over three days the next participant was entered into the intervention, this was to maximise the intervention received by participants, in consideration of the time constraints faced in each study. A prolonged demonstration of experimental effect was not the focus of the research. Repeated demonstrations of effect across each multiple-baseline design in future studies may address this issue.

In addition, a possible limitation of the current research is the multiple component nature of the PT intervention programme, and the decision rules that guided the intervention. The multiple components of the PT intervention programme can be considered a strength of the intervention, but it also makes it difficult to link the observed effects to specific ingredients of the PT intervention programme. Future research could employ a component analysis to determine the relative contribution of each element to intervention outcomes.

The PT intervention programme used decision rules to make data-based judgments about adequate progress. The decision rules in the PT intervention programme were based on student performance data, specifically, maintaining a $\times 2$ acceleration while building rates to a performance standard. Decision rules based on these performance data are objective, quantitative, and can be systematically applied across all learners. Using the decision rules the intervention was modified in varying degrees across foundational reading skills. Such changes (although systematic) are potentially a confounding variable in evaluating the effect of the PT intervention.

The multi-level model proposed could also provide a means of addressing this issue. Many of the experiments reported large gains in frequencies by the 3rd intervention session; programme modifications were applied on average eight days into the intervention. The multi-level model could estimate the immediate effect of the intervention, and thus demonstrate its effectiveness across participants, prior to any programme modifications. Additional modelling could investigate any influence of intervention modifications on treatment effect over time.

A possible limitation with the decision rules in the current research is the parameters used to operationalise “response” (i.e., $\times 2$ acceleration), and the stringency and focus of the procedures used to accelerate learning rates. Future research could investigate differing operationalisations “response” and evaluate the effects of other alterable variables that can be adjusted to increase or decrease the intensity of the supports within the PT intervention programme. The programme modifications used in the current research may not be the most effective in accelerating rates of learning, future research could investigate the effectiveness of a range of different modifications that could be applied within a problem solving hierarchy. The personal best component was the only consequence modification that was applied in the

current research; similarly, future research could explore the effectiveness of other consequence modifications for the PT intervention programme.

The stringency of the performance standards employed in the current research is another variable that could be investigated in future research. Reduced performance standards were used for letter sounds and names, and for decoding words. This reduction made in light of the young age of participants, and the large celerations observed in the intervention conditions (i.e., the performance levels achieved were viewed as satisfactory in relation to baseline performance). It may be that achievement of higher performance standards would occasion increased generalisation to partial and complete decoding. The overall effect of the PT intervention as a function of varied performance standards is a worthy focus for future research.

Ascending baselines demonstrated by some participants were in part an artifact of the graphic display used in the presentation of the data for the current thesis. Add/subtract graphs were chosen to display the multiple-baseline designs to accommodate the sometimes-large numbers of participants in the designs. Thus although these graphs were chosen for ease of data presentation, they are however subject to limitations in visual inspection of the data as described in the introduction. In the day to day implementation of the intervention, the SCC was employed and celerations within these charts provided quantitative metrics to make data-based decisions. The increase in trend observed in the current analysis, however, was not clinically significant on the level of the individual, as these trends were quantifiable on the SCC. Use of the SCC in the baseline condition ensured that even where a small increase was observed, this could be quantitatively compared to the effect of the intervention when it was introduced. The exact celeration values were not presented

in the current theses as a matter of interpretability for the reader. Magnitudes of acceleration values are not readily understood in the wider population, for that reason results were presented as a participants “response” in relation to the x2 acceleration.

Future research could further increase the generalisability of findings by first calculating an effect size measure per participant and subsequently combining these effect sizes across participants and across studies using the multilevel model (as an alternative to combining raw data) as suggested by Van den Noortgate and Onghena (2008). This entails the advantage that this effect size can be transformed to an effect size comparable to group comparison designs (i.e., standardised mean difference). If effect sizes from both types of designs can be combined, more data is available for synthesis, which will increase external validity.

The current research focused on one outcome of behavioral fluency – maintenance. Another outcome of fluency is generalisation of skills (Binder, 1996; Johnson & Street, 2004, 2013). Considering the importance of generalisation, future research could further examine this outcome of the intervention. Because this was the first evaluation of the PT intervention programme, generalisation was measured, but not specifically programmed for. Future research could programme for generalisation and measure generalisation effects in untrained stimuli and reading in connected text. In addition, real books could be incorporated into the intervention. An interesting avenue of investigation would be measuring any concurrent increases in independent reading behaviour e.g., increases in book loans from the school library.

One consideration in the interpretation of findings from the current research involves the standard score gains reported from the YARC letter sound and CTOPP blending words subtest. These gains specifically need to be interpreted with caution as letter sounds and blending sounds were targeted with the PT intervention programme,

thus possibly inflating the score gains. It has been argued that the specificity of the measures and the young age of participants may overstate the standard score gains on these measures (J. Torgesen, personal communication, May 13, 2014).

The participants received the intervention on a one-to-one basis, not the typical format for Tier 2 instruction, and often a mode of instruction criticised for being expensive. However, one-to-one instruction has been identified as the “gold standard” (Slavin et al., 2011), and as the intervention took on average 10 minutes to administer it can be compared to delivering small group (2-3 participants) for 20-30 minutes a day as recommended in the RTI practice guide (Gersten et al., 2009).

This research was the first to implement and evaluate the PT intervention package developed; as such foundational skills were targeted in isolation e.g., letter sounds only, or two skills sequentially e.g., letter sounds followed by decoding. Some of the skills yielded more benefits than others; their combination into one treatment package would potentially yield more meaningful outcomes for overall reading development. Specifically, letter sounds, blending sounds, decoding words (with high frequency MLGPCs), and high frequency words could be combined into a complete PT intervention programme, targeting each one with the same group of at risk readers. In addition, the outcomes for overall reading development could be evaluated over a longer period of time.

Future research could more closely align this Tier 2 intervention to the core-reading programme, and indeed it could be implemented within an overall RTI framework, increasing the ecological validity of the intervention programme. In addition, the fidelity of the core reading intervention could be measured and presented. Other behavioural measurements such as increases in classroom responding could be measured as a pre- post intervention effect.

While the results demonstrated robust intervention effects over general classroom instruction, the intervention was not compared to other methods for targeting foundational reading skills such as incremental rehearsal and traditional drill and practice methods. Future research could systematically compare this intervention programme against other methods and commercial reading interventions, disentangling relative effectiveness and efficiency outcomes. An example of this may be a groups design comparison to the *Reading Recovery* programme, which is the main early reading intervention in DEIS schools in Ireland. *Reading Recovery* is an intensive intervention (60-100 hours) that stipulates admission criteria such as age of entry, and expected success rates (i.e. 70% of population). Considering the widespread implementation of this programme as the sole early reading intervention protocol, a comparative analysis would prove beneficial.

6.10. Summary of Conclusions

Education is plagued by logical fallacies to describe reading difficulties, for example, “Susie has a reading problem because she has a learning disability” (Cooper et al., 2007; Joseph, 2008). For students experiencing reading difficulties (and then potentially diagnosed as reading disabled), use of such tautological arguments do not address the actual reading skill areas that are in need of intervention. For many students, reading difficulties are a result of insufficient types and amount of reading instruction (Joseph, 2008). Lindsley’s assertion that “the child is always right” provides a welcome reminder that a learning disability is really a learning rate difficulty (Skinner, 2008). It is the responsibility of the educator to systematically observe academic responding and match instruction to the child’s learning needs. Lindsley’s other dictum to “try, try again” when the desired results are not obtained

demonstrates PTs commitment to the systematic use of performance data to find such a match.

Konrad and colleagues (2011) assert that evidence-based instructional methods are necessary but perhaps not sufficient for closing achievement gaps. Moreover, if instructional time is not maximised, these achievement gaps are likely to widen over time. Due to this cumulative widening, educators must provide the maximally efficient and effective instruction. The current research demonstrated that the PT intervention programme provides an effective and efficient intervention. Intervention effects reported as a function of the amount of intervention received through performance standards achieved, rate gains in foundational skills targeted and generalisation effects to word reading, as well as standard score gains provide the practical standard that can be understood and interpreted by educators who may wish to implement such strategies. The PT intervention accelerated learning, and in many cases brought participants into the average range of performance in very short timeframes. The instructional intensity of the intervention is also fully reported.

RTI frameworks require treatment packages that include an evidence-based targeted intervention, a system for progress monitoring, decision rules, and data-based choices about adequate progress. The PT intervention programme contained all of these elements. RTI focuses on prevention rather than remediation, as a good start in reading is essential (Slavin et al., 2011), and success in school depends largely on success in reading (Brooks, 2007). The current research successfully implemented an early reading intervention programme with at risk readers in schools where reading failure is disproportionately represented.

Struggling readers consistently demonstrate fluency difficulties that persist throughout formal schooling. This paper shows that fluency in foundational reading

skills can be successfully targeted using the PT intervention, and that these effects can remain up to 9 months post-intervention. Maintenance of treatment effect across time gives evidence that rate building to performance standards in foundational reading skills results in fluent performance. Excellent maintenance of academic skills is desirable for the at risk reader, and demonstrates the efficiency of the intervention. The PT intervention is shown to be intensive and of short duration, and demonstrates long-term effects – critical features of effective interventions for at risk readers.

REFERENCES

- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Massachusetts: Massachusetts Inst. Of Tech., Cambridge.
- Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2007). Lasting consequences of the summer learning gap. *American Sociological Review*, *72*, 167–180.
- Ardoin, S. P., & Daly, E. J., III. (2007). Introduction to the special series: Close encounters of the instructional kind – how the instructional hierarchy is shaping instructional research 30 years later. *Journal of Behavioral Education*, *16*, 1-6.
- Ardoin, S. P., Christ, T. J., Morena, L. S., Cormier, D. C., & Klingbeil, D. A. (2013). A systematic review and summarization of the recommendations and research surrounding Curriculum-Based Measurement of oral reading fluency (CBM-R) decision rules. *Journal of School Psychology*, *51*, 1–18.
- Axelrod, S. (1983). *Behavior modification for the classroom teacher* (2nd ed.). New York, NY: McGraw-Hill.
- Barbetta, P. M., Heron, T. E., & Heward, W. L. (1993). Effects of active student response during error correction on the acquisition, maintenance, and generalization of sight words by students with developmental disabilities. *Journal of Applied Behavior Analysis*, *26*, 111-119.
- Barger-Anderson, R., Domaracki, J. W., Kearney-Vakulick, N., Kubina, R. M., Jr. (2004). Multiple baseline designs: the use of a single-case experimental design in literacy research. *Reading Improvement*, *41*, 217-228.
- Barnett, D., W., Daly III, E., J., Jones, K., M., & Lentz, Jr., F., E. (2004). Response to intervention: Empirically based special service decisions from single-case

- designs of increasing and decreasing intensity. *The Journal of Special Education*, 38, 66–79.
- Berens, K. N., Boyce, T. E., Berens, N. M., Doney, J. K., & Kenzer, A. L. (2003). A technology for evaluating relations between response frequency and academic performance outcomes. *Journal of Precision Teaching*, 19, 20-34.
- Berens, N. M. (2010). *Using a vocabulary building program to increase the rate of arbitrarily applicable derived relational responding*. Unpublished doctoral dissertation, University of Nevada-Reno.
- Binder, C. (1988). Precision Teaching: Measuring and attaining exemplary academic achievement. *Youth Policy*, 10, 12-15.
- Binder, C. (1990). Precision teaching and curriculum based measurement. *Journal of Precision Teaching*, 7, 33-35.
- Binder, C. (1996). Behavioral Fluency: Evolution of a new paradigm. *The Behavior Analyst*, 19, 163–197.
- Binder, C., Haughton, E., & Bateman, B. (2002). Fluency: achieving true mastery in the learning process. Retrieved from:
http://special.edschool.virginia.edu/papers/Binder-et_al_Fluency.pdf
- Binder, C., & Watkins, C. L. (1990). Precision teaching and direct instruction: measurably superior instructional technology in schools. *Performance Improvement Quarterly*, 3, 74-96.
- Black, P., & Wiliam, D (1998). Assessment and Classroom Learning. *Assessment in Education*, 5, 7-71.
- Blaiklock, K. E. (2004). The importance of letter knowledge in the relationship between phonological awareness and reading. *Journal of Research in Reading*, 27, 36-57.

- Bliss, S. L., Skinner, C. H., Hautau, B., & Carroll, E. E. (2008). Articles published in four school psychology journals from 2000 to 2005: an analysis of experimental/intervention research. *Psychology in the Schools, 45*, 483-499.
- Bollman, K. A., Silberglitt, B., & Gibbons, K. A. (2007). The St. Croix education district model: Incorporating systems-level organization and a multi-tiered problem-solving process for intervention delivery. In S. R. Jimerson, M. K. Burns, and A. M. VanDerHeyden (Eds.), *Handbook of Response to Intervention* (pp. 319 - 330). New York, NY: Springer.
- Bramlett, R., Cates, G. L., Savina, E., & Lauinger, B. (2010). Assessing effectiveness and efficiency of academic interventions in school psychology journals: 1995-2005. *Psychology in the Schools, 47*, 114-125.
- Brooks, G. (2007). *What Works for Children with Literacy Difficulties? The effectiveness of intervention schemes*. London, UK: Department for Children, Schools and Families.
- Brooks Rickard, K. (2010). *An investigation of language-building procedures on derived relations of coordination and distinction: Implications for comprehension*. Unpublished doctoral dissertation, University of Nevada-Reno.
- Bullara, D.T, Kimball, J.W., & Cooper, J.O. (1993). An assessment of beginning addition skills without instruction or practice. *Journal of Precision Teaching, 1*, 11-16.
- Burke, M. D., Hagan-Burke, S., Zou, Y., & Kwok, O. (2010). A Structural Equation Model Using Fluency- Based Early Literacy Measures to Predict Emerging Reading Ability in Kindergarten. *Remedial and Special Education, 31*, 385-399.

- Burns, M. K., Dean, V. J., & Foley, S. (2004). Preteaching unknown key words with incremental rehearsal to improve reading fluency and comprehension with children identified as reading disabled. *Journal of School Psychology, 42*, 303–314.
- Burns, M. K., Parker, D. C. (2010). Using the instructional level as a criterion to target reading interventions. *Reading & Writing Quarterly: Overcoming Learning Difficulties, 30*, 79-94.
- Burns, M. K., & Sterling-Turner, H. H. (2010). Comparison of efficiency measures for academic interventions based on acquisition and maintenance. *Psychology in the Schools, 47*, 26-134.
- Bursock, B., & Blanks, B. (2010). Evidenced-based early reading practices within a response to intervention system. *Psychology in the Schools, 47*, 421–430.
- Calkin, A. B. (2005). Precision teaching: the standard celeration charts. *The Behavior Analyst Today, 6*, 207–213.
- Carnine, D. W., Silbert, J., & Kameenui, E. J. (1997). *Direct instruction reading*. Upper Saddle River, NJ: Prentice Hall.
- Carnine, D. W., Silbert, J., Kame'enui, E. J., Tarver, S. G., & Jungjohann, K. (2006). *Teaching struggling and at-risk readers: A direct instruction approach*. Upper Saddle River, NJ: Pearson.
- Case, L., Harris, K.R., & Graham, S. (1992). Improving the mathematical problem solving skills of students with learning disabilities: Self-regulated strategy development. *Journal of Special Education, 26*, 1-19.
- Cates, G. L., Burns, M. K., & Joseph, L. M. (2010). Introduction to the special issue: Instructional efficiency and the impact on data-based decision making. *Psychology in the Schools, 47*, 111-113.

- Cates, G. L., Dunne, M., Erkfritz, K. N., Kivisto, A., Lee, N., & Wierzbicki, J. (2006). Differential effects of two spelling interventions on acquisition and maintenance of spelling words and adaptation to reading. *Journal of Behavioral Education, 16*, 71 – 82.
- Chan, P. E., Konrad, M., Gonzalez, V., Peters, M. T., & Ressa, V. A. (2014). The critical role of feedback in formative instructional practices. *Intervention in School and Clinic, 50*, 96-104.
- Chen, V., & Savage, R. S. (2014). Evidence for a simplicity principle: teaching common complex grapheme-to-phonemes improves reading and motivation in at-risk readers. *Journal of Research in Reading, 37*, 196-214.
- Clark, M. M. (1994). Young literacy learners: how we can help them. Learnington Spa: Scholastic.
- Clark, M. M. (2013). High frequency words: A neglected resource for young literacy learners. *Reading News, Autumn*, 15-18.
- Codding, R. S., & Lane, K. L. (2015). A spotlight on treatment intensity: an important and often overlooked component of intervention enquiry. *Journal of Behavioral Education, 24*, 1-10.
- Connor, C. M., Alberto, P. A., Compton, D. L., O'Connor, R. E. (2014). Improving Reading Outcomes for Students with or at Risk for Reading Disabilities: A Synthesis of the Contributions from the Institute of Education Sciences Research Centers (NCSER 2014-3000). Washington, DC: National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education. Retrieved from: <http://ies.ed.gov/>.
- Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis*. Columbus, OH: Merrill.

- Cortiella, C. (2011). *The state of learning disabilities*. New York, NY: National Center for Learning Disabilities. Retrieved from:
http://www.nclld.org/images/stories/OnCapitolHill/PolicyRelatedPublications/stateofld/2011_state_of_ld_final.pdf
- Coyne, M., Kamefenui, E., & Simmons, D. (2001). Prevention and intervention in beginning reading: Two complex systems. *Learning Disabilities Research and Practice, 16*(2), 62- 73.
- Coyne, M. D., Little, M., Rawlinson, D., Simmons, D., Kwok, O., Kim, M., & Civetelli, C. (2013). Replicating the impact of a supplemental beginning reading intervention: The role of instructional context. *Journal of Research on Educational Effectiveness, 6*, 1–23.
- Coyne, M. D., Simmons, D. C., Hagan-Burke, S., Simmons, L. E., Kwok, O., Kim, M., & Rawlinson, D. M. (2013). Adjusting beginning reading intervention based on student performance: An experimental evaluation. *Exceptional Children, 80*, 25-44.
- Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology, 33*, 934–945.
- Cunningham, A. E., & Stanovich, K. E. (2001). What reading does for the mind. *Journal of Direct Instruction, 1*, 137-149.
- Daly, E. J., III, Lentz, F. E., & Boyer, J. (1996). The instructional hierarchy: A conceptual model for understanding the effective components of reading interventions. *School Psychology Quarterly, 11*, 369-386.
- Daly III, E. J., Martens, B. K., Barnett, D., Witt, J. C., & Olson, S. C. (2007). Varying intervention delivery in response to intervention: confronting and resolving

- challenges with measurement, instruction, and intensity. *School Psychology Review*, 36, 562-581.
- Daly, E. J., III, Witt, J. C., Martens, B. K., & Dool, E. J. (1997). A model for conducting a functional analysis of academic performance problems. *School Psychology Review*, 26, 554–574.
- Datchuk, S. M., & Kubina, Jr., R. M. (2011). Communicating experimental findings in single case design research: How to use celeration values and celeration multipliers to measure direction, magnitude, and change of slope. *Journal of Precision Teaching and Celeration*, 27, 3-17.
- de Graaff, S., Bosman, A. M. T., Hasselman, F., & Verhoeven, L. (2009). Benefits of systematic phonics instruction. *Scientific Studies of Reading*, 13, 318-333.
- Denton, C. A. (2012). Response to intervention for reading difficulties in the primary grades: some answers and lingering questions. *Journal of Learning Disabilities*, 45, 232.
- Denton, C. A., Nimon, K., Mathes, P. G., Swanson, E. A., Kethley, C., Kurz., T. B., et al. (2010). Effectiveness of a supplemental early reading intervention scaled up in multiple schools. *Exceptional Children*, 76, 394–416.
- Denton, C. A., Tolar, T. D., Fletcher, J. M., Barth, A.E., Vaughn, S., & Francis D. J. (2013). Effects of Tier 3 intervention for students with persistent reading difficulties and characteristics of inadequate responders. *Journal of Educational Psychology*, 105, 633-648.
- Department of Education and Science, DES. (2005). Survey of traveller education provision. Inspectorate of the Department of Education and Science (DES), Evaluation, Support, and Research Unit. Dublin: DES. Retrieved from: <https://www.education.ie/en/Publications/Inspection->

[ReportsPublications/Evaluation-Reports-](#)

[Guidelines/insp_survey_traveller_education_provision_pdf.pdf](#)

- Downey, D. B., von Hippel, P. T., & Beckett, B. (2004). Are schools the great equalizer? Cognitive inequality during the summer months and the school year. *American Sociological Review*, *69*, 613–635.
- Duhon, G. J., House, S. E., Poncy, B. C., Hastings, K. W., & McClurg, S. C. (2010). An examination of Two Techniques of Promoting response generalisation of early literacy skills. *The Journal of Behavioral Education*, *19*, 62-75.
- Eivers, E., Shiel, G., & Shortt, F. (2004). *Reading literacy in disadvantaged primary schools*. Dublin, Ireland: Educational Research Centre.
- Evans, M. A., Bell, M., Shaw, D., Moretti, S., & Page, J. (2006). Letter names, letter sounds, and phonological awareness: an examination of kindergarten children across letters and of letters across children. *Reading and Writing*, *19*, 959-989.
- Evans, S. S., & Evans, W. H. (1985). Frequencies that ensure skill competencies. *Journal of Precision Teaching VI*, 25–30.
- Fabrizio, M. A. (February and March 2003). A parent’s introduction to fluency: Parts I and II. The OARacle, Alexandria, VA: The organisation for Autism research. Retrieved from: <http://plea-agency.org/wp-content/uploads/2014/10/A-Parents-Intro-to-Fluency.pdf>
- Fabrizio, M. A., & Moors, A. L. (2003). Evaluating mastery: Measuring instructional outcomes for children with Autism. *European Journal of Behavioral Analysis*, *4*, 23-36.
- Fien, H., Smith, J. L. M., Smolkowski, K., Baker, S. K., Nelson, N. J., & Chaparro, E. (in press). An examination of the efficacy of a multi-tiered intervention on

- early reading outcomes for first grade students at risk for reading difficulties. *Journal of Learning Disabilities*, advanced online publication.
- Fergusson, D. M., & Lynskey, M. T. (1997). Early reading difficulties and later conduct problems. *Journal of Child Psychology and Psychiatry*, 38, 899-907.
- Freeman, G., & Haughton, E. (1997a). *Phonological Coding: Word and Syllable Awareness, Teachers Manual*. Napa, CA: Haughton Learning Center.
- Freeman, G., & Haughton, E. (1997b). *Phonological Coding: Phonemic Awareness, Teachers Manual*. Napa, CA: Haughton Learning Center.
- Fuchs, D. & Deshler, D. D. (2007). What we need to know about responsiveness to intervention (and shouldn't be afraid to ask). *Learning Disabilities Research and Practice*, 22, 129-136.
- Fuchs, L.S., & Fuchs, D. (2003). Curriculum-based measurement Letter Sound Fluency. Measure available from lynn.a.davies@vanderbilt.edu.
- Fuchs, D., Fuchs, L. S., Compton, D. L., Bouton, B., Caffrey, E., & Hill, L. (2007). Dynamic assessment as responsiveness to intervention. *Teaching Exceptional Children*, 39, 58-63.
- Fuchs, L. S., & Vaughn, S. (2012). Responsiveness-to-intervention: a decade later. *Journal of Learning Disabilities*, 45, 195-203.
- Gast, D. L. (2010). *Single subject research methodology in behavioral sciences*. New York, NY: Routledge.
- Gellert, A., & Elbro, C. (1999). Reading disabilities, behaviour problems and delinquency: a review. *Scandinavian Journal of Educational Research*, 43, 131-155.
- Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., et al. (2009). *Assisting students struggling with reading: Response to*

Intervention and multi-tier intervention for reading in the primary grades. A practice guide. (NCEE 2009-4045). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from:
http://ies.ed.gov/ncee/wwc/publications_reviews.aspx://ies.ed.gov/ncee/wwc/publications/practice.

Gilbert, J. K., Compton, D. L., Fuchs, D. F., & Fuchs, L. S. (2012). Early screening for risk of reading disabilities: recommendations for a four-step screening system. *Assessment for Effective Intervention, 38*, 6-14.

Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the development of Educational Achievement.

Gontijo, P. F. D., Gojntijo, I., & Shilcock, R. (2003). Grapheme-phoneme probabilities in British English. *Behavior Research Method, Instruments, and Computers, 35*, 136-157.

Graf, S., Auman, J., & Lindsley, O. (2007). *The one-year standard celeration chart*. Youngstown, OH: Graf Implements.

Gresham, F.M., & Witt, J.C. (1997). Utility of intelligence tests for treatment planning, classification, and placement decisions: Recent empirical findings and future directions. *School Psychology Quarterly, 12*, 249-267.

Grigorenko, E. L. (2009). Dynamic assessment and Response to Intervention: Two sides of one coin. *Journal of Learning Disabilities, 42*(2), 111-132.

Gross, J., Hudson, J., Price, D. (2009). *The long term costs of numeracy difficulties*. Every Child a Chance Trust (KMPG).

Harn, B. A., Stoolmiller, M., & Chard, D. J. (2008). The role of automaticity and

- unitization. *Journal of Learning Disabilities*, 41, 143-157.
- Haring, N. G., Lovitt, T. C., Eaton, M. D., & Hanson, C. L. (1978). *The fourth R: Research in the classroom*. Columbus, OH: Charles E. Merrill Publishing.
- Hatcher, P. J., Hulme, C., & Ellis, A.W. (1994). Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The phonological linkage hypothesis. *Child Development*, 65, 41-57.
- Hatcher, P., J., Hulme, C., & Snowling, M. J. (2004). Explicit phoneme training combined with phonic reading instruction helps young children at risk of reading failure. *Journal of Child Psychology and Psychiatry*, 45, 338-358.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Taylor & Francis.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81-112.
- Hecht, S. A., Burgess, S. R., Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2000). Explaining social class differences in growth of reading skills from beginning kindergarten through fourth-grade: The role of phonological awareness, rate of access, and print knowledge. *Reading and Writing: An Interdisciplinary Journal*, 19, 99-127.
- Hemphill, F. C., & Vanneman, A. (2011). *Achievement gaps: How Hispanic and White students in public schools perform in mathematics and reading on the National Assessment of Educational Progress* (NCES 2011-459). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Retrieved from: <http://nces.ed.gov/nationsreportcard/pdf/studies/2011459.pdf>
- Hill, D. R., King, S. A., Lemons, C. J., & Partanen, J. N. (2012). Fidelity of

- implementation and instructional alignment in response to intervention research. *Learning Disabilities Research & Practice*, 27, 116–124.
- Hogan, T. P., Catts, H. W., & Little, T. D. (2005). The relationship between phonological awareness and reading: Implications for the assessment of phonological awareness. *Language, Speech, and Hearing Services in Schools*, 36, 285–293.
- Hoover, J. J. (2011). Making informed instructional adjustments in RTI models: essentials for practitioners. *Intervention in School and Clinic*, 47, 82-90.
- Hosp, J. L. (2008). Best practices in aligning academic assessment with instruction. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology V* (Vol. 2, pp. 363–376). Bethesda, MD: National Association of School Psychologists.
- Huang, F. L., Tortorelli, L. S., & Invernizzi, M. (2013). An investigation of factors associated with letter sound knowledge at kindergarten entry. *Early Childhood Research Quarterly*, 29, 182-192.
- Hudson, R. F., Pullen, P., C., Lane, H. B., & Torgesen, J. K. (2009). The complex nature of reading fluency: A multidimensional view. *Reading & Writing Quarterly*, 25, 4–32.
- Hudson, R. F., Torgesen, J. K., Lane, H. B., & Turner, S. J. (2010). Relations among reading skills and subskills and text-level reading proficiency in developing readers. *Reading and Writing*, 25, 483–507.
- Hughes, C. A., & Dexter, D. D. (2011). Response to Intervention: A research-based summary. *Theory Into Practice*, 50, 4–11.
- Hulme, C., Stothard, S. E, Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., & Snowling, M. J. (2009). *York Assessment of Reading for Comprehension – Early Reading*. GL Assessment.

- Ivarie, J. J. (1986). Effects of proficiency rates on later performance of a recall and writing behavior. *Remedial and Special Education, 7*, 25-30.
- Jenkins, J. R., Hudson, R. F., & Johnson, E. S. (2007). Screening for at-risk readers in a response to intervention framework. *School Psychology Review, 36*, 582-600.
- Johnson, K., & Street, E. M. (2004). *The Morningside model of generative instruction*. Concord, MA: Cambridge Center for Behavioral Studies.
- Johnson, K., & Street, E. M. (2013). *Response to Intervention and Precision Teaching: creating synergy in the classroom*. New York, NY: The Guilford Press.
- Joseph, L. M. (2008). Best practices on interventions for students with reading problems. In A. Thomas and J. Grimes, (Eds.), *Best Practices in School Psychology*, (Vol. 5, pp. 1163–1309).
- Joseph, L., Eveleigh, E., Konrad, M., Neef, N., & Volpe, R. (2012). Comparison of the efficiency of drill methods on children's reading performance. *Journal of Applied School Psychology, 28*, 317-337.
- Joseph, L. M., Kastein, L. A., Konrad, M., Chan, P. E., Peters, M. T., & Ressa, V. A. (2014). Collecting and documenting evidence: Methods for helping teachers improve instruction and promote academic success. *Intervention in the School and Clinic, 50*, 86-95.
- Joseph, L. M., & Nist, L. (2006). Comparing the effects of unknown-known ratios on word reading learning versus learning rates. *Journal of Behavioral Education, 15*, 69 – 79.
- Kamps, D., Abbott, M., Greenwood, C., Wills, H., Veerkamp, M., & Kaufman, J. (2008). Effects of small-group reading instruction and curriculum differences

- for students most at risk in kindergarten: Two-year results for secondary- and tertiary-level interventions. *Journal of Learning Disabilities, 41*, 101–114.
- Kairaluoma, L., Ahonen, T., Mikko, A., Holopainen, L. (2007). Boosting reading fluency: An intervention study at the subword level. *Scandinavian Journal of Educational Research, 51*, 253-274.
- Katzir, T., Kim, Y., Wolf, M., O'Brien, B., Kennedy, B., Lovett, M., & Morris, R. (2006). Reading fluency: The whole is more than the parts. *Annals of Dyslexia, 56*, 51–82.
- Kavale, K. A., & Spalding, L. S. (2008). Is response to intervention good policy for specific learning disability? *Learning Disabilities Research & Practice 23*, 169-179.
- Konrad, M., Helf, S., & Joseph, L. M. (2011). Evidence-based instruction is not enough: strategies for increasing instructional efficiency. *Intervention in the School and Clinic, 47*, 67-74.
- Konrad, M., Keeseey, S., Ressa, V. A., Alexeeff, M., Chan, P. E., & Peters, M. T. (2014). Setting clear learning targets to guide instruction for all students. *Intervention in the School and Clinic, 50*, 76-85.
- Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). *Single-case designs technical documentation*. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf
- Kubina, Jr., R. M., & Lin, F. Y. (2007). Defining frequency: A natural scientific term. *The Behavior Analyst Today, 9*, 125–129.
- Kubina, Jr., R. M., & Yurich, K. L. (2012). *The precision teaching book*. Lemont, PA: Greatness Achieved Publishing Company.

- Kucera, J., & Axelrod, S. (1995). Multiple baseline designs. In S. B. Neuman and S. McCormick (Eds.), *Single subject experimental research: applications for literacy*. Newark, D: International Reading Association.
- LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6, 293–323.
- Leppänen, U., Aunola, K., Niemi, P., & Nurmi, J-E. (2008). Letter knowledge predicts Grade 4 reading fluency and reading comprehension. *Learning and Instruction*, 18, 548-564.
- Lidz, C. S. (1991). *Practitioners guide to dynamic assessment*. New York, NY: The Guilford Press.
- Lin, F. Y., & Kubina, R. M. (2005). Learning channels and verbal behavior. *The Behavior Analyst Today*, 5, 1-14.
- Lindsley, O. R. (1964). Direct measurement and prosthesis of retarded behavior. *Journal of Education*, 147, 62-81.
- Lindsley, O. R. (1972). From Skinner to Precision Teaching: The child knows best. In J. B. Jordan & L. S. Robbins (Eds.), *Let's try doing something else kind of thing*. Arlington, VA: Council on Exceptional Children.
- Lindsley, O. R. (1992). Precision teaching: Discoveries and effects. *Journal of Applied Behavior Analysis*, 25, 51-57.
- Little, M. E., Rawlinson, D., Simmons, D. C., Kim, M., Kwok, O., Hagan-Burke, S., et al. (2012). A comparison of responsive interventions on kindergartners early reading achievement. *Learning Disabilities Research & Practice*, 27, 189–202.
- Macmann, G. M., Barnett, D. W., Lombard, T. J., Belton-Kocher, E., & Sharpe, M. N. (1989). On the actuarial classification of children: Fundamental studies of

- classification agreement. *The Journal of Special Education*, 23, 127-149.
- MacMillan, D. L., Gresham, F. M., & Bocian, K. M. (1998). Discrepancy between definitions of learning disabilities and school practices: An empirical investigation. *Journal of Learning Disabilities*, 31, 314-326.
- Malmquist, S. (2004). Using a multi-level system of assessment to inform instructional decisions and determine program effectiveness. In K. Johnson and E. M. Street, *The Morningside Model of Generative Instruction*. Concord, MA: Cambridge.
- Marston, D., Muyskens, P., Lau, M., & Canter, A. (2003). Problem-solving model for decision making with high-incidence disabilities: The Minneapolis experience. *Learning Disabilities Research & Practice*, 18, 187-200.
- Martens, B. L., Werder, C. S., Hier, B. O., & Koenig, E. A. (2013). Fluency training in phoneme blending: A preliminary study of generalized effects. *Journal of Behavioral Education*, 22, 16-36.
- Mathes, P. G., & Denton, C. A. (2002). The prevention and identification of reading disability. *Seminars in Paediatric Neurology*, 9, 185-191.
- McBride-Chang, C. (1999). The ABC's of the ABC's: The development of letter-name and letter-sound knowledge. *Merrill-Palmer Quarterly*, 45, 285-308.
- McGuinness, C., McGuinness, D., & McGuinness, G. (1996). Phono-Graphix: A new method for remediating reading difficulties. *Annals of Dyslexia*, 46, 73-96.
- McMaster, K. L., Fuchs, D., Fuchs, L. S., & Compton, D. L. (2005). Responding to non-responders: An experimental field trial of identification and intervention methods. *Exceptional Children*, 71(4), 445-464.
- Melby-Lervåg, M. M., Halaas Lyster, S. A., & Hulme, C. (2012). Phonological skills and their role in learning to read: a meta-analytic review. *Psychological*

Bulletin, 138, 322-353.

- Mellard, D., McKnight, M., & Jordan, J. (2010). RTI tier structures and instructional intensity. *Learning Disabilities Research & Practice*, 25, 217–225.
- Miller, A. D., Hall, S. W., & Heward, W. L. (1995). Effects of sequential 1-minute time trials with and without inter-trial feedback and self-correction on general and special education students' fluency with math facts. *Journal of Behavioral Education*, 5, 319-345.
- Miskin, R. (2011). Read Write Inc. Phonics, Green word cards. Oxford: Oxford University Press
- Moeyaert, M., Ugille, M., Ferron, J., Beretvas, S. N., & Van den Noortgate, W. (2013a). Three-level analysis of single-case experimental data: Empirical validation. *The Journal of Experimental Education*, 82, 1–21.
- Moeyaert, M., Ugille, M., Ferron, J., Beretvas, S. N., & Van den Noortgate, W. (2013b). Three-level analysis of standardized single-case experimental data: Empirical validation. *Multivariate Behavior Research*, 48, 719–748.
- Mol, S. E., & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin*, 137, 267–296.
- Moore, B. A., & Klinger, J. K. (2014). Considering the needs of English language learner populations: An examination of the population validity of reading intervention research. *Journal of Learning Disabilities*, 47, 391-408.
- Moore Partin, T. C., Robertson, R. E., Maggin, D. M., Oliver, R. M., Wehby, J. H. (2010). Using teacher praise and opportunities to respond to promote appropriate student behavior. *Preventing School Failure*, 54, 172-178.
- Musti-Rao, S. (2005). *The effects of a supplemental early reading intervention with*

urban kindergarten and first grade students: a preventative approach.

Unpublished doctoral dissertation, Ohio State University.

National Center on Response to Intervention (NCRTI, January 2013). *Progress Monitoring Brief #1 Common Progress Monitoring Omissions: Planning and Practice*. Washington, DC: U.S. Department of Education, Office of Special Education Programs, National Center on Response to Intervention. Retrieved from:
<http://www.rti4success.org/sites/default/files/RTI%20ProgressMonitoringBrief1-Planning%20and%20Practice.pdf>

National Center on Response to Intervention (NCRTI, January 2013b). *Progress Monitoring Brief #3: Common Progress Monitoring Graph Omissions: Making Instructional Decisions*. Washington, DC: U.S. Department of Education, Office of Special Education Programs, National Center on Response to Intervention. Retrieved from:
<http://www.rti4success.org/sites/default/files/RTI%20ProgressMonitoringBrief1-Planning%20and%20Practice.pdf>

National Reading Panel (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature in reading and its implications for reading instruction*. Retrieved from:
<http://www.nichd.nih.gov/publications/pubs/nrp/documents/report.pdf>

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

Organisation for Economic Co-operation and Development. (2014). *PISA 2012 Results: What students know and can do – student performance in*

mathematics, reading and science (Volume I, Revised edition, February 2014). Paris, France: OECD. Retrieved from:

<http://www.oecd.org/pisa/keyfindings/pisa-2012-results-volume-I.pdf>

- Patrick, H., Mantzicopoulos, P., Samarapungavan, A., & French, B. F. (2008). Patterns of young children's motivation for science and teacher-child relationships. *The Journal of Experimental Education*, 76, 121-144.
- Pennington, B. F., & Lefly, D. L. (2001). Early reading development in children at risk for dyslexia. *Child Development*, 72, 816-833.
- Perfetti, C. A., & Marron, M. A. (1998). Learning to read: Literacy acquisition by children and adults. In D. A. Wagner (Ed.). *Advances in adult literacy research and development*. New York, NY: Hampton Press.
- Peterson, M., Brandes, D., Kunkel, A., Wilson, J., Rahn, N., Egan, A., & McComas, J. (2014) Teaching letter sounds to kindergarten English language learners using incremental rehearsal, *Journal of School Psychology*, 52, 97-107.
- Philips, B. M., Clancy-Menchetti, J., & Lonigan, C. J. (2008). Successful phonological awareness instruction with preschool children, lessons from the classroom. *Topic in Early Childhood Special Education*, 28, 3-17.
- Piasta, S., B., Purpura, D. J., & Wagner, R. K. (2010). Fostering alphabet knowledge development: a comparison of two instructional approaches. *Reading and Writing*, 23, 607-626.
- Power, T. J. (2008). Editorial note: improving intervention effectiveness and efficiency. *School Psychology Review*, 37, 291-293.
- President's Commission on Excellence in Special Education. (2002). *A new era: Revitalizing special education for children and their families*. Washington, DC: Author. Retrieved from:

<https://education.ucf.edu/mirc/Research/President%27s%20Commission%20on%20Excellence%20in%20Special%20Education.pdf>

Progress in International Leading Literacy Study. (2011). *International results in reading*. Chestnut Hill, MA: TIMSS & PIRLS International Study Centre.

Retrieved from:

http://timssandpirls.bc.edu/pirls2011/reports/downloads/P11_IR_FullBook.pdf

Pufpaff, L. A. (2009). A developmental continuum of phonological sensitivity skills. *Psychology in the Schools, 46*, 679-691.

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest, 2*, 31-74.

Reason R., & Morfidi E. (2001) Literacy difficulties and single - case experimental design. *Educational Psychology in Practice, 17*, 227–244.

Reed, D. K., Cummings, K. D., Schaper, A., & Biancarosa, G. (2014). Assessment fidelity in reading intervention research: a synthesis of the literature. *Review of Educational Research, 84*, 275-321.

Reynolds, C. R., & Shaywitz, S. E. (2009). Response to intervention: ready or not? Or, from wait-to-fail to watch-them-fail. *School Psychology Quarterly, 24*, 130-142.

Riley-Tilman, T., C., & Burns, M. K. (2010). *Evaluating educational interventions: Single-case design for measuring response to intervention*. New York, NY: Guilford Press.

Ritchey, K. D., & Speece, D. L. (2006). From letter names to word reading: The nascent role of sublexical fluency. *Contemporary Educational Psychology, 31*, 301–327.

- Roberts, W. R., & Norwich, B. (2010). Using precision teaching to enhance the word reading skills and academic self-concept of secondary school students: a role for professional educational psychologists. *Educational Psychology in Practice, 26*, 279-298.
- Rose, J. (2006). Independent review of the teaching of early reading. Department for Education and Skills. Retrieved from: www.standards.dfes.gov.uk/rosereview/
- Ross, S. G., & Begeny, J. C. (2014). Single-case effect size calculation: comparing regression and non-parametric approaches across previously published reading intervention data sets. *Journal of School Psychology, 52*, 419-431.
- Savage, R. S., Abrami, P., Hipps, G., & Deault, L. (2009). A randomized controlled trial study of the ABRACADABRA reading intervention programme in Grade 1. *Journal of Educational Psychology, 101*, 590-604.
- Savage, R., & Carless, S. (2005). Phoneme manipulation not onset-rime manipulation ability is a unique predictor of early reading. *Journal of Child Psychology and Psychiatry, 46*, 1297-1308.
- Schmidgall, M., & Joseph, L. M. (2007). Comparison of phonic analysis and whole word reading on first graders' cumulative words read and cumulative reading rate: An extension in examining instructional effectiveness and efficiency. *Psychology in the Schools, 44*, 319 – 332.
- Scholin, S. E., & Burns, M. K. (2012). Relationship between pre-intervention data and post-intervention reading fluency and growth: A meta-analysis of assessment data for individual students. *Psychology in the Schools, 49*, 385-398.
- Shapiro, E. S. (2009). *Response to intervention, the two models of RTI: standard protocol and problem solving*. Virginia: Virginia Department of Education.

Retrieved from:

http://www.doe.virginia.gov/instruction/virginia_tiered_system_supports/responsive_intervention/two_models.pdf

- Shirley, M. J., & Pennypacker, H. S. (1994). The effects of performance criteria on learning and retention of spelling words. *Journal of Precision Teaching, 12*, 73-86.
- Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York, NY: McGraw-Hill.
- Simmons, D. C., Coyne, M. D., Hagan-Burke, S., Kwok, O., Simmons, L., Johnson, C., & Crevecoeur, Y. C. (2011). Effects of supplemental reading interventions in authentic contexts: A comparison of kindergarteners' response. *Exceptional Children, 77*, 207–228.
- Skinner, C. H. (2008). Theoretical and applied implications of precisely measuring learning rates. *School Psychology Review, 37*, 309-314.
- Skinner, C. H. (2010). Applied comparative effectiveness researchers must measure learning rates: a commentary on efficiency articles. *Psychology in the Schools, 47*, 166-172.
- Skinner, C. H., Belfiore, P. J., Mace, H. W., Williams-Wilson, S., & Johns, G. A. (1997). Altering response topography to increase response efficiency and learning rates. *School Psychology Quarterly, 12*, 54 – 64.
- Skinner, C. H., Belfiore, P. B., & Watson, T. S. (1995/2002). Assessing the relative effects of interventions in students with mild disabilities: Assessing instructional time. *Journal of Psychoeducational Assessment, 20*, 345-356
(Reprinted from *Assessment in Rehabilitation and Exceptionality, 2*, 207-220.)
- Slate, J. R., & Jones, C. H. (2000). Assessment issues in special education. In M. A.

- Windsor & K. Mazurek, (Eds.), *Special Education in the 21st Century: Issues of Inclusion and Reform* (pp. 66-82) Washington, DC: Gallaudet University Press.
- Slavin, R. E., Lake, C., Davis, S., & Madden, N. A. (2011). Effective programs for struggling readers: a best evidence synthesis. *Educational Research Review*, 6, 1–26.
- Snowling, M. J., & Hulme, C. (2010). Evidence-based interventions for reading and language difficulties: Creating a virtuous circle. *British Journal of Educational Psychology*, 81, 1-23.
- Snowling, M. J., & Hulme, C. (2012). Interventions for children’s language and literacy difficulties. *International Journal of Language Communication Disorders*, 47, 27-34.
- Solity, J., & Vousden, J. (2009). Real books vs. reading schemes: a new perspective from instructional psychology. *Educational Psychology*, 29(4), 469–511.
- Sparks, R. L., Patton, J., & Murdoch, A. (2014). Early reading success and its relationship to reading and reading volume: replication of ‘10 years later’. *Reading and Writing*, 27, 189-211.
- Speece, D. L., & Case, L. P. (2001). Classification in context: an alternative approach to identifying early reading disability. *Journal of Educational Psychology*, 93, 735-749.
- Stoiber, K. C., & Kratochwill, T. R. (2000). Empirically supported interventions in schools: Rationale and methodological issues: Part I. *School Psychology Quarterly*, 15, 75–115.
- Stuart, M., Masterson, J., & Dixon, M. (2000). Inferring sublexical correspondence from sight vocabulary: Evidence from 6- and 7-year olds. *Quarterly Journal*

of Experimental Psychology, 52, 353-366.

- Stuart, M., Dixon, M., Masterson, J., & Gray, B. (2003). Children's early reading vocabulary: Description and word frequency lists. *British Journal of Educational Psychology*, 73, 585-598.
- Suggate, S. P. (2014). A meta-analysis of the long-term effects of phonemic awareness, phonics, fluency, and reading comprehension interventions. *Journal of Learning Disabilities*, 47, 1-20.
- Swanson, H. L., & Sachse-Lee, C. (2000). A meta-analysis of single-subject-design intervention research for students with LD. *Journal of Learning Disabilities*, 33, 114-136.
- Torgesen, J. K. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research & Practice*, 15, 55-64.
- Torgesen, J. K. (2004). Lessons learned from research on interventions for students who have difficulty learning to read. In P. McCardle & V. Chhabra (Eds.), *The voice of evidence in reading research* (pp. 355-382). Baltimore, MD: Brookes.
- Torgesen, J. K. (2005). Remedial interventions for students with dyslexia: national goals and current accomplishments. S. Richardson and J. Gilger, (Eds.), *Research-based education and intervention: what we need to know* (pp. 103-124). Boston: International Dyslexia Association.
- Torgesen, J. K., & Hudson, R. F. (2006). Reading Fluency: Critical issues for struggling readers. In S.J. Samuels and A. Farstrup (Eds.), *Reading fluency: The forgotten dimension of reading success*. Newark, DE: International Reading Association.

- Treptow, M. A., Burns, M. K., & McComas, J. M. (2007). Reading at the frustration, instructional, and independent levels: the effects on students' reading comprehension and time on task. *School Psychology Review, 36*, 159-166.
- VanDerHeyden, A. M., Witt, J. C., & Gilbertson, D. (2007). A multi-year evaluation of the effects of a response to intervention (RTI) model on identification of Children for special education. *Journal of School Psychology, 45*, 225-256.
- Van den Noortgate, W., & Onghena, P. (2003a). Combining single-case experimental data using hierarchical linear models. *School Psychology Quarterly, 18*, 325–346.
- Van den Noortgate, W., & Onghena, P. (2003b). Hierarchical linear models for the quantitative integration of effect sizes in single-case research. *Behavior Research Methods, Instruments, & Computers, 35*, 1–10.
- Van den Noortgate, W., & Onghena, P. (2008). A multilevel meta-analysis of single-subject experimental design studies. *Evidence Based Communication Assessment and Intervention, 2*, 142–151.
- Vaughn, S., Denton, C. A., & Fletcher, J. M. (2010). Why intensive interventions are necessary for students with severe reading difficulties. *Psychology in the Schools, 47*, 432-444.
- Vaughn, S., & Linan-Thompson, S. (2003). What is special about special education for students with learning disabilities? *The Journal of Special Education, 37*, 140-205.
- Vaughn, S., & Wanzek, J. (2014). Intensive interventions in reading for students with reading disabilities: meaningful impacts. *Learning Disabilities Research & Practice, 29*, 46–53.
- 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, 223-238.
- Verhoeven, L., & van Leeuwe, J. (2009). Modeling the growth of word-decoding skills: evidence from Dutch. *Scientific Studies of Reading*, 13, 205-223.
- Volpe, R. J., Burns, M. K., DuBois, M., & Zaslofsky, A. F. (2011). Computer-assisted tutoring: Teaching letter sounds to kindergarten students using incremental rehearsal. *Psychology in the Schools*, 48, 332–342.
- Volpe, R. J., Mulé, C. M., Briesch, A. M., Joseph, L. M., & Burns, M. K. (2011). A comparison of two flashcard drill methods targeting word recognition. *Journal of Behavioral Education*, 20, 117-137.
- Vousden, J. I. (2008). Units of English spelling-to-sound mapping: A rational approach to reading instruction. *Applied Cognitive Psychology*, 22, 247-272.
- Vousden, J. I., Ellefson, M. R., Solity, J., & Chater, N. (2011). Simplifying reading: applying the simplicity principle to reading. *Cognitive Science*, 35, 34-78.
- Wanzek, J., Roberts, G., & Al Otaiba, S. (2014). Academic responding during instruction and reading outcomes for kindergarten students at-risk for reading difficulties. *Reading and Writing: An Interdisciplinary Journal*, 27, 55-78.
- Wanzek, J., & Vaughn, S. (2007). Research-based implications from extensive early reading interventions. *School Psychology Review*, 4, 541.
- Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive test of phonological processing*. Austin, TX: PRO-ED.
- Warren, S. F., Fey, M. E., & Yoder, P. J. (2007). Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Mental Retardation and Developmental Disabilities Research*

Reviews, 13, 70–77.

- Wechsler, D. (2005). *Wechsler individual achievement test, 2nd UK ed.* London, Harcourt Assessment.
- Weinstein, G., & Cooke, N. L. (1992). The effects of two repeated reading interventions on generalisation of fluency. *Learning Disability Quarterly, 15*, 21-28.
- Wiliam, D. (2010). An integrative summary of the research literature and implications for a new theory of formative assessment. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 18-40). New York, NY: Taylor & Francis.
- Whalen, K. P., Willis, R. J., & Sweeney, W. J. (1993). Using 1-minute time trials and 4-minute practice sessions to improve a student's performance of fraction problems. *Journal of Precision Teaching, 5*, 55-64.
- White, O. (2009). A focus on the individual: Single-subject evaluations of response to intervention. In S. Rosenfeld and V. Beringer (Eds.), *Implementing evidenced-based interventions in school settings* (pp. 531–558). New York, NY: Oxford University Press.
- Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology, 91*(3), 415–438.
- Wolf, M., & Katzir-Cohen, T. (2001). Reading fluency and its intervention. *Scientific Studies of Reading, 5*, 211–239.
- Yell, M. (2004, February). *Understanding the three-tier model*. Presentation at the Colorado State Directors of Special Education Meeting, Denver, CO.
- Yoder, P. J. & Woynaroski, T. (2015). How to study the influence of intensity of

treatment on generalised skill and knowledge acquisition in students with disabilities. *Journal of Behavioral Education*, 24, 152-166.

APPENDIX A

Participant Information Sheet**Participant Information Sheet for the Multiple Baseline Studies.****Participant Information Sheet**

Title of study: Supporting the Development of Senior Infant Students' Early Reading Skills.

Experimenter(s): Julie Brosnan, BA, MSc.

Supervisor(s): Dr. Olive Healy & Dr. Kendra Brooks Newsome.

Objective of the study: The objective of this study is to develop a framework that supports students' early reading development. Students learn to read at different rates, and some students might find certain reading skills harder to learn. This study attempts to support those students who may need some additional teaching in early reading skills. By this the study hopes to prevent the development of reading difficulties experienced by some students.

Invitation to participate: Your child is being invited to take part in a research study. Before you decide, it is important for you to understand why the research is being done and what it will involve. This participant information sheet will tell you about the purpose, risks, and benefits of this research study. If you do **NOT** wish for your child to take part you can sign this information sheet and return it to the school. Otherwise, we will contact you by phone to answer any questions about the research. If you agree for your child to take part, and your child agrees to take part, we will ask you to sign a consent form. If there is anything you are not clear about, we will be happy to explain it to you. Please take as much time as you need to read this information sheet. You should only consent for your child to participate in this research study when you feel that you understand what is being asked of you and your child, and you have had enough time to think about your decision. Thank you for reading this.

Purpose of the study: Our study is looking at the assessment of early reading skills, and supporting the development of such skills, to prevent reading difficulties happening for some students. In particular we are interested in developing a framework to accomplish this in senior infant classrooms. Your child is being asked to participate as they are a Senior Infant student who may directly benefit from the assessments and intervention we will deliver as a part of the study. We will measure your child's reading level before, and after the study. This will tell us how effective our framework is in supporting students' early reading skills.

Does my child have to take part?

Your child is under no obligation to take part. It is up to you to decide if you want your child to take part. If you agree to them taking part we will request that you explain the study to them in a developmentally appropriate manner, and if your child then agrees to take part we will give you this information leaflet to keep and ask you to sign a consent form. We will explain the study also to your child and we will request verbal assent from her/him (this means that he/she agrees with taking part). In all cases, if you decide to take part you, or your child, are still free to withdraw at any point, and without giving a

reason. A decision to withdraw at any time, or a decision not to take part, will not affect your rights in any way.

What will happen to my child if she/he takes part?

Initially your child will be given a computer administered assessment (Children's Progress Academic Assessment or CPAA). This assessment indicates if your child is near, or below, the early reading skill level expected for students at this grade level. Should your child be in either of these categories they will be given 3 additional reading assessments (but never more than one on any day), and these will be given again at 1 or 2 more time points later on in the study. These assessments give us a better idea of the reading level your child is at, and their performance in relation to other students. After the 1st assessment your child will be randomly assigned a start point for the intervention. This means that some children will start straight away, while other others will start at a later date. This allows us to see if the intervention is working for your child specifically.

What it means for your child on a day to day basis is that from the assigned start point he or she will receive additional instruction delivered by the researcher every day during the school week. The results from CPAA are specific in the skills that your child would benefit from additional support in. For example, if your child is having difficulty in recognising the sounds at the beginning or end of words CPAA will specify this. Based on these findings the researcher will use Precision Teaching (PT) to work with your child every day for 5-20 minutes, teaching and practicing that skill until your child is fluent in that skill (i.e. she or he performs the skill accurately, automatically, and with ease). This will happen in the classroom so your child should not miss out on valuable class time. The PT session will involve making sure your child learns the skill through repeated practice and error correction. Your child's performance in the skill will be measured and graphed daily on special charts called Standard Celeration Charts (SCC). This graph will show that your child is progressing through and learning the skills being taught. While it is important that our approach is scientific, it is of equal importance that your child is learning, and having fun! Students tend to find PT enjoyable and motivating as they can see their progress every day and they get frequent reinforcement for progress made.

How long will my child's part in the study last?

Your child's part in the study will last 1, 2, or 3 terms during the school year from either January 2012-May 2012 or September 2012-May 2013. The length of your child's part in the study will depend on assessment findings: should CPAA show that your child would continue to benefit from further additional support, that support will continue to be provided to your child for 1, 2, or 3 terms as necessary. All of the research will take place during normal school hours and within your child's classroom.

Who can take part?

Most students who are in general education Senior Infant classrooms save those groups outlined below.

Who cannot take part?

Those who cannot take part are: Students who already receive additional learning support; students with a diagnosed speech or language impairment; students with developmental disabilities.

What is the educational intervention being investigated?

The intervention being investigated is the use of CPAA to identify students that could benefit from additional educational support in early reading skills. This educational support is called Precision Teaching (PT). PT involves teaching the student one skill at a

time and then practicing that skill until it is fluent i.e. automatic and requiring no effort. Then further skills are taught in the same way. PT involves measuring each student's growth in the skill being taught, and graphing that growth daily. This allows the educator to see the student's progress in learning.

What are the possible benefits in taking part?

The possible benefits to your child taking part are that your child's early reading skills will be enhanced, and this may improve their overall reading performance. In addition, your child's participation will make an important contribution to research that will advance our understanding of assessment and educational intervention with Senior Infant students in Ireland.

What are the possible risks of taking part?

There are no possible risks to your child as a result of taking part in this study.

What happens at the end of the study?

At the end of the study the results from the various reading assessments will be combined with those from other participants and analysed using visual inspection of the data, and data analysis software. The findings from these analyses will be used to write a scientific study that may be submitted to academic journals and/or presented at conferences. A laypersons version of the findings will be made available to both parents and teachers no later than 3 months from the end of the study. In addition, copies of any published work will be provided to the school for parents or teachers, alternatively you can directly request published works from the researcher. As all the findings will either be presented in group form or using pseudonyms for individual examples: no student should be identifiable from the findings. At the end of the study, the intervention - Precision Teaching - will be discontinued with your child, and will not be available as an intervention.

What happens if I change my mind during the study?

If for whatever reason, at any time in the study, you or your child have a change of mind about participating there will be no negative consequence. It will not offer your child any disadvantage to stop participating in the study at any time.

What happens if I have a complaint during the study?

Should you wish to make a complaint during the study the research team will be available to you at the contact details provided below. Alternatively if you wish to speak to someone independent of the research team, contact details for a special contact person in NUI Galway is provided below also.

Whom do I contact for more information or if I have further concerns?

Should you require any further information; or if you have any concerns the research team will be available to you at the contact details provided below. Alternatively if you wish to speak to someone independent of the research team, contact details for a special contact person in NUI Galway is provided below also.

Confidentiality: All information that is collected about your child during the course of the research will be kept strictly confidential, it will only be made available to the researchers involved in the study, and will not be shared with anyone else. The information collected in this research study will be stored in a way that protects your child's identity. Anonymised numbers will be used instead of names so there is no way that your child will be identified. In addition, all paper records will be kept in a locked filing cabinet in the department of Psychology in NUI Galway. Computer records will be held securely in password protected files, access to which will be limited to the researchers involved in the study. Future presentations of findings (in the form of conference presentations, or publication in a scientific journal) will be reported as group data and will not identify your child in any way.

Remember: You are free to refuse your child takes part in this study without them suffering any disadvantage. Should you agree for your child to take part, you or your child can change your mind at any point during the study without any disadvantage. Having read this information sheet should you find that you are still unclear about any part of it, or have any questions/concerns you wish to raise, please feel free to contact:

Julie Brosnan, BA, MSc.
School of Psychology
Room 453 New Engineering Building,
National University of Ireland, Galway
University Road,
Galway
Telephone: (087)2891478
E-mail: jbrosnan4@gmail.com

If you have any concerns about this study and wish to contact someone independent and in confidence, you may contact ‘the Chairperson of the NUI Galway Research Ethics Committee, c/o Office of the Vice President for Research, NUI Galway, ethics@nuigalway.ie.

Date: 15.06.11

Participant Information Sheet Version number: 02

--

I am **NOT** interested in being contacted about the possibility of my child’s participation in this study

Name: _____

APPENDIX B

Consent Form



Informed Consent

Title of Project: Supporting the Development of Senior Infant Students' Early Reading Skills.

Researcher: Julie Brosnan, BA, MSc.

Supervisors: Dr. Olive Healy, Dr. Kendra Brooks Newsome

In order to participate in this research study, it is necessary that you give your informed consent. By signing this informed consent statement you are indicating that you understand the nature of the research study and your child's role in that research, and that you agree for her/him to participate in the research. Please consider the following points before signing:

Please initial box

1. I confirm that I have read the information sheet (version 003) for the above study and have had the opportunity to ask questions.

[]

2. I am satisfied that I understand the information provided and have had enough time to consider the information.

[]

3. I understand that my child's participation is voluntary and that I am free to withdraw him/her at any time, without giving any reason, without my legal rights being affected.

[]

4. My child does not have a diagnosis of a speech and language difficulty, or of developmental disability. My child does not receive special education services.

[]

5. I agree for my child to take part in the above study.

[]

By signing this form I am stating that I understand the above information and consent for my child to participate in this study.

Signed: _____ Date:

First Name (print): _____ Second Name (print):

I have explained the above and answered all questions asked by the participant:

Researcher Signature: _____ Date:

APPENDIX C

Procedural Integrity Checklist for Baseline Sessions**Procedural Integrity Checklist for Baseline Sessions**

Student ID _____ Date: _____

Target Skill _____

Observer _____

	Yes	No
Organisation		
Instructional area is neat and clean		
All stimulus materials ready		
Provides a clear cue that the session is starting		
Instructional Delivery		
Secures student attention before delivering first cue		
Provides a clear focus cue (e.g. listen/look)		
Tells participant to 'do the best they can and keep going until the timer beeps'		
Provides appropriate random display of stimuli		
Provides no feedback on performance nor make reference to speed		
Data Collection/Analysis		
Needed charts are present and set up before instruction begins		
Timer is set between each timing		
Counts as instruction progresses through timings		
Data are recorded as numbers on sheet		
Reinforcement		
Reinforces effort		
Reinforces appropriate learning skills		
Does not give reinforcement for correct responding		
If questioned on performance redirects 'what's important is that you do the best you can'		
Total		

APPENDIX D

Procedural Integrity Checklist for Intervention Sessions**Procedural Integrity Checklist for Intervention Sessions**

Student ID _____ Date: _____

Target Skill _____

Observer _____

	Yes	No
Organisation		
Instructional area is neat and clean		
All stimulus materials ready		
All charts ready		
Sets one timer to record the session duration		
Provides a clear cue that the session is starting		
Instructional Delivery		
Secures student attention before delivering first cue		
Provides a clear focus cue (e.g. listen/look)		
Correct learning channel/slice as indicated on daily chart		
Provides no inadvertent cueing		
Speech is rhythmic and tone of voice varies		
Provides appropriate random display of stimuli		
Provides differential reinforcement for correct responding		
Data collection/Analysis		
Timer is set between each timing		
Count as instruction progresses through timings		
Data are recorded as numbers on data collection sheet		
Data are transferred to TPM after each timing		
Median data points (corrects and errors) are recorded on the DPM chart		
Student behaviour brought into contact with the chart		
Error Correction		
Provides encouragement during the timings where needed		
Errors are corrected only between timings		
Errors are corrected when student data indicate this is needed		
Primes previously missed targets where required		
Prompts rather than corrects		
Calls mistakes learning opportunities		
Reinforcement		
Different reinforcers used in the session		
Reinforces appropriate learning skills		
Reinforcement related to student expectation and learning goals		
Provides more reinforcement on harder items/tasks, less on easier items/tasks		
TOTAL		

APPENDIX E

Social Validity Measure for Teachers**Social Validity Measure of Teacher Satisfaction with Intervention Goals, Procedures, & Outcomes.**

It is important to us that you, as a teacher are satisfied with the goals of the intervention, the procedures used to implement it, and the outcome for students. Please rate each statement to the degree that you agree with it, this can be anonymous so please be as honest as possible! Alternatively if you would like to leave your name and/or make other comments or suggestions please do.

KEY: 1 = Strongly agree, 2 = Agree, 3 = Don't know, 4 = Disagree, 5 = Strongly disagree

It is important to assess children's literacy needs as early as Senior Infants.

1	2	3	4	5
Strongly agree	Agree	Don't know	Disagree	Strongly disagree

When assessments are carried out, it is important that the results are used to guide teachers in their students education needs.

1	2	3	4	5
Strongly agree	Agree	Don't know	Disagree	Strongly disagree

I would use The Children's Progress Academic Assessment with my students as a means of assessing their educational needs.

1	2	3	4	5
Strongly agree	Agree	Don't know	Disagree	Strongly disagree

I would use Precision Teaching with my students as a way of teaching fluency in core skills.

1	2	3	4	5
Strongly agree	Agree	Don't know	Disagree	Strongly disagree

I would use Precision Teaching with my students as a way of monitoring their progress in learning.

1	2	3	4	5
Strongly agree	Agree	Don't know	Disagree	Strongly disagree

A Response to Intervention Framework makes sense in terms of helping struggling readers as they fail.

1	2	3	4	5
---	---	---	---	---

Strongly agree	Agree	Don't know	Disagree	Strongly disagree
-------------------	-------	------------	----------	----------------------

Precision Teaching helped students gain core skills in the development of reading.

1 Strongly agree	2 Agree	3 Don't know	4 Disagree	5 Strongly disagree
------------------------	------------	-----------------	---------------	---------------------------

Because of the intervention, student's overall reading levels have improved.

1 Strongly agree	2 Agree	3 Don't know	4 Disagree	5 Strongly disagree
------------------------	------------	-----------------	---------------	---------------------------

Thank you for completing this questionnaire!

If you would like to leave your name, or have any contacts or queries there is space provided below. Alternatively if you would like to remain anonymous leave the space blank.

Also provided are my contact details, please feel free to contact me regarding any query or concern you may have

Name:

Comments: _____

Contact Details for the Researcher:

Julie Brosnan, BA, MSc.

School of Psychology

Room 453 New Engineering Building,

National University of Ireland, Galway

University Road,

Galway

Telephone: (087)2891478

E-mail: jbrosnan4@gmail.com

APPENDIX F

Social Validity Measure for Parents**Social Validity Measure of Parental Satisfaction with Intervention Outcomes.**

It is important to us that you, as a parent or guardian are satisfied with your child's role in the research, and the outcome of their participation.

Please rate each statement to the degree that you agree with it, this is anonymous so please be as honest as possible! Alternatively if you would like to leave your name and/or make other comments or suggestions please do.

KEY: 1 = Strongly agree, 2 = Agree, 3 = Don't know, 4 = Disagree, 5 = Strongly disagree

It is important to assess children's literacy needs as early as Senior Infants.

1	2	3	4	5
Strongly agree	Agree	Don't know	Disagree	Strongly disagree

When assessments are carried out, it is important that the results are used to guide teachers in their students education needs.

1	2	3	4	5
Strongly agree	Agree	Don't know	Disagree	Strongly disagree

If students fall behind in class and they get some extra help with what they need to learn, they can catch up with their class-mates.

1	2	3	4	5
Strongly agree	Agree	Don't know	Disagree	Strongly disagree

I feel that there was an overall benefit to my child receiving the intervention.

1	2	3	4	5
Strongly agree	Agree	Don't know	Disagree	Strongly disagree

My child is more positive towards reading since they received the intervention.

1	2	3	4	5
Strongly agree	Agree	Don't know	Disagree	Strongly disagree

My child's reading has improved since they received the intervention.

1	2	3	4	5
Strongly agree	Agree	Don't know	Disagree	Strongly disagree

Thank you for completing this questionnaire!

If you would like to leave your name, or have any contacts or queries there is space provided below. Alternatively if you would like to remain anonymous leave the space blank.

Also provided are my contact details, please feel free to contact me regarding any query or concern you may have

Name:

Comments _____

Contact Details for the Researcher:

Julie Brosnan, BA, MSc.

School of Psychology

Room 453 New Engineering Building,

National University of Ireland, Galway

University Road,

Galway

Telephone: (087)2891478

E-mail: jbrosnan4@gmail.com

APPENDIX G

Social Validity Measure for Participants

Social Validity - Participant Interview

Participant: _____

Interventionist: _____

Interviewer: _____ Date: _____

1. Did you like being in this special group/ program? _____ Yes
_____ No

2. What did you like best about this special group/ program?

2. Did you like..... Being part of the group?
(Not Much/ A Little /A Lot)

☺ ☺ ☺ ☺ ☺ ☺

Spending time with the Ms. _____ in the group?

☺ ☺ ☺ ☺ ☺ ☺

Earning rewards/ stars? Learning new skills?

☺ ☺ ☺ ☺ ☺ ☺

3. Do you feel you learned important things? _____ Yes
_____ No

If yes, what is/ are the most important thing(s) you learned?

5. Did you learn things that will

Help you do better work in school? Help you at home?

(Not Much/ A Little /A Lot)



6. Do you use the skills that you learned in the special program/ group?

____ Yes ____ No

If yes, where do you use these skills?

7. Do you wish our group/ program could have lasted longer?

(Clarify.....meet for more time, like for another week?) _____

Yes _____ No

If yes, how much longer would you like to have met?

8. Is there anything else you would like to tell me about your special group/ program?

APPENDIX H

SAFMEDS Decoding Word Sets

SAFMEDS Set	MLGPC	Words Targeted		
Set 1	/oo/	food	zoo	
		pool	too	
	/ay/	way	may	
		day	say	
	/ch/	chin	chop	chat
Set 2	/ee/	see	sleep	
		been	seen	
	/th/	thin	this	
	/ou/	shout	out	
		loud	round	
	/sh/	shop	fish	ship
Set 3	/ar/	car	hard	
		part	start	
	/ir/	third	dirt	
		girl	twirl	
	/ew/	blew	drew	
		new	grew	
Set 4	/igh/	might	high	
		night	light	
	/aw/	paw	dawn	
		saw	crawl	
	/or/	fork	sort	snort

APPENDIX I

SAFMEDS 100 High Frequency Word Sets

<p><u>Set 1</u></p> <p>a away did* had* into new off* some their to</p>	<p><u>Set 2</u></p> <p>about back* could has* I next* old them today two</p>	<p><u>Set 3</u></p> <p>after big* come he in* not* on* then up* very</p>	<p><u>Set 4</u></p> <p>all but* do her* is now one saw* there too*</p>	<p><u>Set 5</u></p> <p>am* be down have it* my once see* they us</p>
<p><u>Set 6</u></p> <p>an* because for* here last* me other three* was you</p>	<p><u>Set 7</u></p> <p>and* by from* his live our* said take* we with</p>	<p><u>Set 8</u></p> <p>are came* get* little make* of out* she went* will*</p>	<p><u>Set 9</u></p> <p>as can* go him* made* over that this* were what</p>	<p><u>Set 10</u></p> <p>at* call got* like* look put so the time* when*</p>

100 High Frequency Words (Solity & Vousden, 2009), words marked with an asterix can be decoded.

