Challenges of teaching phonics-based reading remediation in Singapore

Leung, Wai Tung

2019

Leung, W. T. (2019). Challenges of teaching phonics-based reading remediation in Singapore. Master's thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/137123>

<https://doi.org/10.32657/10356/137123>

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Downloaded on 26 Apr 2023 22:00:30 SGT

**CHALLENGES OF TEACHING PHONICS-BASED READING REMEDIATION IN
SINGAPORE**

LEUNG WAI TUNG

SCHOOL OF SOCIAL SCIENCES

NANYANG TECHNOLOGICAL UNIVERSITY

2019

**CHALLENGES OF TEACHING PHONICS-BASED READING REMEDIATION IN
SINGAPORE**

LEUNG WAI TUNG

SCHOOL OF SOCIAL SCIENCES

**A thesis submitted to the Nanyang Technological University in partial
fulfillment of the requirement for the degree of Masters of Arts**

2019

Statement of Originality

I certify that all work submitted for this thesis is my original work. I declare that no other person's work has been used without due acknowledgement. Except where it is clearly stated that I have used some of this material elsewhere, this work has not been presented by me for assessment in any other institution or University. I certify that the data collected for this project are authentic and the investigations were conducted in accordance with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

15/08/19

.....

Date

.....

Leung Wai Tung

Supervisor Declaration Statement

I have reviewed the content of this thesis and to the best of my knowledge, it does not contain plagiarised materials. The presentation style is also consistent with what is expected of the degree awarded. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accordance with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

18/08/2019

.....
Date

.....
Asst Prof Suzy Styles

Authorship Attribution Statement

This thesis describes research that has been previously reported elsewhere.

Chapter 2 describes research that was preregistered on NTU's institutional repository:

Leung, W. T., & Styles, S. J. (2018). *Reading Intervention Quality: Preregistration Docs*. DR-NTU (Data), V1.
<https://doi.org/10.21979/N9/RAQM8H>

The contributions of the co-authors of the preregistration are as follows:

- Leung WT & Styles SJ collaborated on the design of the study.

The preliminary outcomes of this research were first described in a paper prepared under HP7888, Leung WT (2018), "Evaluating quality of evidence in reading interventions with phonics training and those without, for children aged 6 to 9 years old with reading difficulties: A systematic review".

Selected outcomes of this research were also presented at the following international conferences:

- Leung WT & Styles SJ (2019). "What is the quality of evidence for reading intervention studies? A systematic review of small-group reading support programmes" (poster), *International Symposium on Cognition and Neuroscience 2019*, NTU, Singapore, 8 March 2019.
- Leung WT & Styles SJ (2019). "What is the quality of evidence for reading intervention studies? A systematic review of small-group reading support programmes" (oral presentation), *Linguistics and Knowledge About Language in Education (LKALE) SIG 2019*, University of Oxford, UK, 5 July 2019.

The contributions of the co-authors of both conference presentations are as follows:

- LWT & SJS jointly designed and preregistered the study.
- LWT conducted the systematic review, and created the graphs.
- LWT & SJS jointly created the poster and presentation slides.

Chapter 3 describes research that was preregistered on NTU's institutional repository:

Leung, W. T., & Styles, S. J. (2019). *Preregistration Documents: What do words sound like in Singapore?* DR-NTU (Data), V1. Retrieved from <https://doi.org/10.21979/N9/UAZ7AS>

The contributions of the co-authors of the preregistration are as follows:

- LWT & SJS collaborated on the design of the study.
- LWT completed data collection and analysed the data.

18/ 08 / 19

.....
Date

.....
Leung Wai Tung

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to my supervisor, Dr Suzy Styles, for her patient guidance, insightful discussions, and invaluable expertise. I would never look at research and children's language development in the same light. Thank you for introducing me to exciting ideas and concepts that will guide me through much exploration in the years to come!

I would also like to thank all members of the NTU BLIP Lab for their immense help and encouragement during this period of time.

Finally, I wish to thank my family members and friends for their support and encouragement throughout my study. I wish to thank my ever-supportive husband, Whelan. This would not be possible without him.

Contents

Chapter 1: Introduction.....	1
How children learn to read.....	1
Methods of reading instruction	5
Bilingualism in Singapore	8
English in Singapore	9
Vowels.....	10
How English is taught in Singapore	11
Overview	15
Chapter 2: Systematic review of reading interventions for children with reading difficulties	17
Disparity between research and classroom practices.....	17
Quality of methodological practices in education research	19
Methods	22
PRISMA Protocol	23
Quality indicators.....	23
Results	26
Discussion	32
Application of quality indicators	33
Open access.....	34
Linguistic differences	36
Recommendations for researchers	37
Recommendations for educators	39
Chapter 3: Phonology of Singapore English: An investigation of readers' intuitions about vowel sounds	41
Introduction	41
The current study	45
Method	46
Participants.....	47
Test stimuli	48
Experimental design	50
Hypothesis and analysis plan	52
Results	54
Data exclusion	54
Possible vowel merger.....	54
Degree of vowel merger	58
Speed of decisions.....	62
Discussion	63

Recommendations for researchers	67
Recommendations for educators	69
Chapter 4: Challenges in designing a phonics scope and sequence appropriate for teaching English in Singapore.....	71
Implications of the Singapore English context on the use of phonics.....	72
Developing a scope and sequence for phonics instruction	75
Recommendations for researchers	82
Recommendations for educators	82
Conclusions	84
References	86
Appendix 1: Preregistration codebook for Systematic review.....	94
Appendix 2: Preregistration report for Systematic review.....	100
Appendix 3: PRISMA Protocol for Systematic review	102
Appendix 4: Differences between protocol and review	107
Appendix 5: Preregistration report for ‘What do words sound like in Singapore?’	109
Appendix 6: Demographics and language questionnaire	117
Appendix 7: PEBBLES – The Perceptual Experience Battery.....	119
Appendix 8: List of words used in ‘What do words sound like in Singapore?’	121
Appendix 9: Data analysis for “What do words sound like in Singapore?”	123
Appendix 10: Grapheme-phoneme correspondences.....	124
Appendix 11: Full-text articles excluded with reasons (Systematic review).....	125

Abstract

Many early readers struggle to master grapheme-phoneme correspondences – the relationship between letter patterns and the speech sounds they represent in a language. There has been extensive evidence for the effectiveness of explicit systematic phonics, which teaches early readers grapheme-phoneme correspondences. In order to examine the availability and quality of evidence of school-based reading interventions for struggling early readers, a systematic review was conducted comparing phonics versus non-phonics interventions. We found limited information about group interventions targeting reading difficulties from the past decade of research. While there is more evidence supporting phonics-based reading intervention than non-phonics based, the pool of available evidence is small. We also found no bilingual studies with sufficient quality.

Speakers of Singapore English are known to differ substantially in their pronunciations of vowels, compared to speakers of other varieties of English. Given the importance of grapheme-phoneme correspondences in phonics instruction, it is therefore critical to consider how differences in pronunciation of vowels in Singapore English can affect phonics instruction. In order to find out whether there are possible vowel mergers, we tested English-speaking adults with a minimal pair task in which they judged whether two words sound the same or different. We found that the adults judged word pairs differently, depending on whether the word pairs involved vowels suspected of merger.

Ideally, the scope and sequence of a phonics-based instruction should begin with a small set of one-to-one letter-sound pairs that represent the most regular, highest frequency correspondences in a written language. Given the mismatches between standard scope and sequence and the regularities of the Singapore variety of English, we then discuss how this mismatch has implications for students in Singapore, and the challenges in developing a scope and sequence appropriate for Singaporean learners.

Chapter 1: Introduction

The ability to read allows one to acquire knowledge and access better employment opportunities. Low literacy is estimated to cost the global economy 970 billion dollars a year (World Literacy Foundation, 2018). People with low literacy tend to have limited employment opportunities and are trapped in poverty. Literacy is essential for daily living and ranges from reading the newspaper to completing a job application. The costs of illiteracy are not limited to poorer economic outcomes, but also negative outcomes such as poorer physical and mental well-being, higher level of crime rates, and welfare dependency (World Literacy Foundation, 2018).

For literate adults, reading and writing is usually effortless and automatic. This makes it hard to appreciate how challenging and complex it is for children who are acquiring these skills for the first time. Literacy covers both word-level (reading, spelling) and text-level skills (reading comprehension and writing). Since children start the journey towards literacy by learning how to read, this thesis will focus on the core skill of reading.

How children learn to read

Learning to read is different from learning to speak. Various researchers (e.g., Castles, Rastle, & Nation, 2018; Stuart & Stainthorp, 2015; Treiman, 2018) have reiterated the point that while spoken language is natural and considered a normal behaviour in the human species, writing is a recent invention in human history. Spoken language acquisition begins from infancy, even before infants understand words, and their auditory processing systems begin attuning to

sounds in their language(s). While exposure and experience is crucial for spoken language, by the time typically-developing children start school, they normally have no difficulty understanding spoken language and speaking. This means that for most children, spoken language precedes reading and oral language skills remain stronger than reading skills for some time. As a beginner reader, children have to first learn how a spoken language they are familiar with is linked to a writing system.

Reading is a skill that needs to be learned, requiring accumulation of learning experience and practice. In order to read words printed on a page, children have to make connections between the visual symbols they encounter and what they already know about their spoken language. This is regarded as a key challenge of reading (Rastle, 2019a). For example, as shown in Figure 1.1, a child needs to connect the printed word 'cat' with its meaning with its pronunciation [kæt]. Assuming that the child already knows about the connection between sound and meaning beforehand, the child now needs to make the connections from what he/she already knows onto spelling.

Figure 1.1. Learning spelling-meaning and spelling-sound mapping. The image used in this figure is licensed under Creative Commons (free to use without attribution). This image is made, after Rastle (2019b).

Willingham (2017) highlighted that for a child, learning to read involves overcoming the following challenges: visually discriminating between letters (visual symbols), distinguishing between speech sounds, and learning the mappings between letters and speech sounds. For the writing system of English, phonemes are critical in the process of learning to read. In this paper, the term phoneme is understood as a category of speech sounds that are treated as functionally equivalent in a language, and that signal differences in meaning. Acoustic analysis of spoken words show that most words occurring in everyday speech are pronounced continuously without breaks within the word. However, linguists would regard 'cat' as being comprised of 3 phonemes /k/ /æ/ /t/ which can be regarded as discrete sounds. The writing system of English represents each of these speech sounds using written letters, hence readers of English learn to map the three phonemes to the three letters, also known as graphemes. One needs to learn through experience with print, that phonemes of English are represented by graphemes. This is known as the alphabetic principle. The writing system of English is not transparent such that one grapheme maps to one phoneme all the time. There are 44 phonemes in standard varieties of English (e.g. the UK), but represented by 26 letters. This is considered complex, in comparison with transparent languages, such as Italian, where there are direct one-to-one correspondences. As such, learning to read English is hard because the writing system cannot be inferred directly from print all the time.

Researchers highlight that effective reading instruction should be determined by the characteristics of the writing system, instead of being guided by prevalent educational dogma (for e.g., Castles et al.,2018). In English, the link

between spelling and meaning is arbitrary, such that similar-looking words tend to have dissimilar meanings. For example, “bat” refers to a mammal that flies while “hat” refers to covering for the head. If a child is taught the phoneme-grapheme relationships for “b”, “a”, “t” and “n”, then knowledge about the pronunciation and spelling of “bat” can be generalised to unfamiliar words such as “tan” and “ban”. Hence, learning the mappings between the phonemes and graphemes of English enables children to read new unfamiliar words independently, and to link the printed words to their meaning. Methods of instruction that focus on the link between phonemes and letters are more effective than methods that focus on learning mappings between whole printed words and their meanings (Castles et al., 2018).

Discovering the alphabetic principle can be challenging for children learning to read, especially if they are left to figure out letter/sound relationships on their own. Studies by Byrne and colleagues found that after being taught to read pairs of written words such as *fat* and *bat*, children without letter/sound knowledge were unable to transfer learning to a new transfer task. In the transfer task, they were presented with the written word *fun*, and then asked whether the word was “fun” or “bun” (Byrne, 2005). As the children were not taught explicitly the alphabetic principle or how to decode and map individual sounds, none of the children were not able to transfer the learning from the previous fat/bat example. Research has concluded that most children must be taught the alphabetic principle, and should be provided with explicit teaching (Castles et al., 2018).

Methods of reading instruction

Reading instruction that explicitly teaches children the relationship between graphemes and phonemes is known as phonics. The basis for phonics instruction is that teaching children a manageable number of grapheme-phoneme correspondences can provide them with the ability to read words on their own.

Mastering grapheme-phoneme correspondences through explicit and systematic phonics for beginning readers, has been recommended by national inquiries in the Australian, UK and US education systems for more than a decade (National Reading Panel, 2000; Rose, 2006; Rowe, 2005). Phonics benefit both typically developing readers and those with learning disabilities. For example, approximately one-third of school children in the US perform below age expectations (NAEP, 2019) and these poor readers are likely to benefit from systematic phonics. The term “science of reading” has generated much online discussion in the recent years in the US, UK and Australia. A qualitative meta-analysis comparing systematic phonics to nonsystematic or no-phonics, as part of the investigation by the U.S. National Reading Panel, found that phonics instruction is more effective compared to nonsystematic or no-phonics instruction and the difference yielded a moderate effect size of 0.41 compared to nonsystematic or no-phonics instruction (Ehri et al., 2001). This meta-analysis has been frequently cited as strong evidence for using phonics instruction to teach both children learning to read, as well as in reading remediation for children with reading difficulties. The use of systematic phonics has been found to result in better outcomes for children in mainstream classrooms (de Graaff,

Bosman, Hasselman, & Verhoeven, 2009; Rose, 2006; Torgerson, Brooks, Gascoine, & Higgins, 2018). However, Torgersen and colleagues (2018) caution that while the majority of systematic reviews conducted over the years support the use of phonics, sources of bias in the published literature (design and publication bias) might have overstressed its benefits. Despite this note of caution, it is undeniable that systematic phonics instruction currently has the best evidence base for teaching young readers how to read in English.

According to the Simple View of Reading (Gough & Tunmer, 1986), decoding and linguistic comprehension are two separate sets of skills that contribute to reading comprehension. Decoding can refer to both “sounding out” a word or skilled word recognition, while linguistic comprehension refers to the ability to comprehend spoken language. This theory has been supported by various studies not only in the English Language, but in other languages as well. Decoding, along with linguistic comprehension, forms a foundation from which reading comprehension can be built upon. Although reading comprehension is not the focus of discussion in this thesis, it should be highlighted that reading comprehension is complex, and that both processes are critical for skilled reading comprehension. The contribution of these two components to reading comprehension changes over time, with decoding playing a heavier role in the early stages of learning to read (Nation, 2019).

Gough and Tunmer (1986) used the Simple View of Reading to discuss reading difficulties in children. Figure 1.2 shows decoding on the horizontal axis and linguistic comprehension on the vertical axis, with four quadrants labelled A, B, C and D, referring to different characteristics of individuals who fall into the

Figure 1.2. Simple View of Reading according to Gough and Tunmer (1986).

quadrants. Strong readers with good spoken comprehension skills will be in Quadrant B, and children in the other quadrants would show different profiles of deficits. For example, children in Quadrant A are characterised as having good comprehension but poor decoding, while those in Quadrant C are characterised as having poor comprehension but poor decoding. This model helps to explain how systematic explicit phonics instruction would be important for improving word reading outcomes for children with reading difficulties by moving those children on the left of Figure 1.2 (poor decoding) to the right (good decoding), regardless of their level of linguistic comprehension.

We have established that in order for children to learn an alphabetic language such as English, they need to grasp the concept that graphemes represent phonemes and that it is best taught through explicit systematic phonics, which typically comes with a pre-determined and sequenced set of teaching materials (Stuart & Stainthorp, 2015). In the following paragraphs, we

introduce bilingualism in Singapore, Singapore English and its linguistic characteristics, and review how English is taught in Singapore.

Bilingualism in Singapore

In Singapore, English is identified as the official working language, while Mandarin, Malay and Tamil represent the official Chinese, Malay and Indian ethnicities are designated as mother-tongue languages. The bilingualism policy requires students in mainstream schools to study English as the main language of instruction, along with one mother tongue, as assigned according to the official classification of one's ethnicity (Sim, 2016).

Pakir (2001) highlighted the increasing perceived importance associated with English in Singapore since the bilingual education policy was introduced. She attributed this perception to long hours of instructional time in classrooms using English and increasing use of English as the dominant language at home. Since Pakir's observation, there has been a continued trend of more households reporting English as the main language at home. As seen Figure 1.3, more English is reported as the main language spoken at home in 2010 compared to 2000 across all three ethnic groups.

As noted by Styles (2019a), the Singapore census data does not provide information on how different languages are used in the home environment. Such information will help us to better understand how children's literacy skills develop from the diverse bilingual language backgrounds in Singapore. With bilingualism and biliteracy as the background to understanding the education landscape in Singapore, we will proceed to discuss English in Singapore.

Change in main language spoken at home in Singapore

Figure 1.3. More English reported as main language spoken at home. Reproduced CC-BY from Styles (2019b), *Singapore's languages at a glance: A visual overview*.

What is the main language spoken at home?

Among 100 people in Singapore...

Ethnic Groups:

Figure 1.4. More English reported as main language spoken at home. Reproduced CC-BY from Styles (2019b), *Singapore's languages at a glance: A visual overview*.

English in Singapore

There is literature documenting how Singapore English differs from other varieties of English. For example, Singapore English (SgE) differs in terms of choice of vocabulary, grammar, pronunciation, word stress, rhythm and intonation (Deterding, 2007; Low & Brown, 2005). While researchers have

typically compared pronunciations in Singapore English with Standard Southern British English (SSBE), they noted that the comparison is made for the purpose of discussion, and that SSBE is recognised internationally and familiar to Singaporeans (Deterding & Hvitfeldt, 1994; Low & Brown, 2005). The focus of this discussion is on pronunciation differences, specifically for vowels, which have design and instructional implications for teaching of phonics in Singapore.

Vowels

According to existing analysis of Singapore English vowels (Deterding, 2007), several pairs of vowels are suspected of merger – for example, Singaporeans do not typically distinguish between the /æ/ vowel in ‘bag’ and the /e/ the vowel in ‘beg’. Figure 1.5 compares the vowel systems between British and Singapore English, as shown in blue and red respectively. The vertical axis refers to Formant 1 (F1), which in typical acoustic analysis results from tongue height in the oral cavity while the horizontal axis refers to Formant 2 (F2), which results from how far back the tongue is positioned in the oral cavity. Figure 1.5 shows that SgE monophthong vowels differ in their acoustic structure, as can be seen in the average formant values, compared to British English vowels. The SgE vowels are visibly closer to one another for certain pairs of vowels, which suggests that these vowels are pronounced more similarly compared to British English. These suspected mergers between pairs of vowels have resulted in Singaporeans’ difficulties in distinguishing between: i) /i:/ *peach* and /ɪ/ *pitch* (*heed* and *hid* in Figure 1.5), ii) /ɑ:/ *cart* and /ʌ/ *cut* (*heart* and *hut* in Figure 1.5),

iii) /ɔ:/ *caught* and /ɒ/ *cot* (*hoard* and *hot* in Figure 1.5), and iv) /e/ *bed* and /æ/ *bad* (*head* and *had* in Figure 1.5).

Figure 1.5. Vowel differences between SgE and SSBE. Reproduced CC-BY from Styles (2018), *Where have all the vowels gone? Vowel systems in British and Singaporean English*.

Given this initial analysis of differences between vowel differences in acoustic structure in Singapore compared to British English, implications on reading instruction in Singapore need to be explored further. Phonics depends on teaching relationships between phonemes and graphemes. Hence, when vowel differences exist, they might have implications for phonics instruction.

How English is taught in Singapore

The Strategies for English Language Learning and Reading (STELLAR) programme is currently implemented in the National Curriculum for primary schools in Singapore. Materials used in the STELLAR curriculum were based on

“high interest, enlarged children’s books for lower primary” (Pang, Lim, Choe, Peters, & Chua, 2015) supplemented by guidelines written by the Ministry of Education’s curriculum planners. Currently, the guidelines highlight students’ enjoyment in driving learning through the use of children’s books and literature. The main teaching methods focus on teacher-guided reading and questioning to engage students. Pang and colleagues (2015) reported that children in the pilot STELLAR schools performed significantly better in both speaking and reading from grade three to five, compared to those in control schools who were using English materials developed by commercial publishers. At present, the best evidence for teaching English Language and literacy in Singapore comes from this evaluation study, which was conducted over a period of six years from 2007 to 2012. However, systematic phonics was not taught in STELLAR programme. Given the evidence on the effectiveness of systematic phonics instruction, this appears to be a gap in the existing syllabus of English in Singapore schools.

As children enter Primary One, those with weak literacy skills are identified for the two-year Learning Support Programme (LSP). The LSP children will then receive daily half-an-hour support conducted in small groups. Children who have made sufficient progress in Primary One will no longer be in the programme for Primary Two. Some who continue to experience literacy difficulties are then screened for dyslexia (Ministry of Education, 2018). Children diagnosed with dyslexia are supported in the School-based Dyslexia Remediation (SDR) Programme for a further two years (Ministry of Education, 2014). It is conducted four times a week after school in small groups. Both programmes are

regarded as intensive support to help children in their literacy development and are resourced by teachers who have received relevant professional training.

The 2010 English language syllabus supported “a balance between decoding and meaning-based instruction, word recognition and passage comprehension, and phonics and whole language” (Ministry of Education, 2010, p.30). However, the National Curriculum has not recommended a set of specified correspondence between letters and sounds, nor guidance for teachers on how to teach students to read words. Since systematic phonics refers to a programme with a pre-determined and sequenced set of teaching materials (Stuart & Stainthorp, 2015), this implies that phonics is not taught explicitly or systematically in the STELLAR programme at the lower primary. As highlighted by Snow (2019), the issue is how all the critical components to reading instruction, such as phonemic awareness, phonics, vocabulary, comprehension and fluency, are arranged relative to each component within the scope and sequence. It is not sufficient that these components of reading instruction are merely mentioned in the curriculum. Comparing the approaches used by STELLAR and the additional support programmes, it appears that there are discrepancies between them, with STELLAR emphasising on meaning and additional programmes emphasising on phonics. Given the importance of word-level literacy development, especially in word reading, there is value in exploring how systematic phonics can be incorporated into the Singapore syllabus consistently for all learners.

This problem of the lack of systematic phonics instruction is not unique to Singapore. With the exception of the UK, where they implemented mandatory

systematic phonics in all schools, classroom practices in countries such as the US and Australia vary across states, and have not consistently kept up to date with reading science. Buckingham and colleagues (2013) attributed political ideologies, reliance on education experts' misplaced beliefs in reading, and poor practices in the evaluation of programmes as contributing factors to resistance in accepting evidence-based reading instruction. The disparity between reading science and classroom practices gained more attention and discussion in the US and Australia by parents and educators in recent years, as reflected in increased reporting and discussions on the Internet (e.g. Hanford, 2018; Snow, 2019; Snow & Clarke, 2015). However, educators and parents in Singapore do not seem aware of these issues, given the lack of such discussions locally. Hence, open discussion between researchers, educators and even parents might be necessary to address the inconsistencies in Singapore on teaching typical-developing children how to read the English Language, as compared to children who are struggling. Early intervention is found to be most effective (Torgesen et al., 2018), and children who are at mild risk of reading difficulties could have better reading outcomes when given appropriate instruction (Moats, 1999). Some of these children might even meet grade-level reading abilities if explicit systematic phonics instruction is provided early (Moats, 2007). This suggests that some children who are at mild risk but not identified to be in the specific support programmes, might continue to struggle in mainstream classroom without help. Even among those identified as having literacy difficulties might not require specialist intervention if explicit phonics instruction were provided in the mainstream classrooms.

While the identification and support for students with reading difficulties in Singapore follows the principles of Response to Intervention (Fuchs & Fuchs, 2006), it should be noted that children in Singapore who fail to meet the criterion for LSP screening would miss out on the year-long programme. Given that there is only one window of opportunity to identify the students who need more support, this implies that students who develop reading problems later, such as difficulties in developing fluency, would not be supported. Only those who are diagnosed to have dyslexia would then be supported through SDR at Primary Three and Four. This suggests that some students will potentially miss out on essential skills in learning to read, given the current design and structure of the English curriculum and support programmes.

Overview

In summary, explicit systematic phonics may be an important element for improving reading instruction in Singapore. Phonics-based instruction is mentioned in the context of a whole-literacy approach for general classroom instruction in the Singapore National Curriculum, but there is little guidance on the way to implement phonics-based instruction for these children, in the context of Singapore English (distinctly different from British English as presented in this chapter). Furthermore, phonics-based instruction is provided at the primary levels to children with reading difficulties in Singapore. But there are substantial existing gaps in the information currently available to educators.

What do we know about intervention strategies to support children with reading difficulties? Are current additional programmes in Singapore supported

by evidence in reading science? In Chapter 2, we investigate best practices to provide group-based interventions in a school setting through a systematic review, and consider the availability and quality of evidence.

Since effective phonics instruction depends on teaching grapheme phoneme correspondences, it is important that we understand the phonemic structure of English in Singapore. In Chapter 3, we explore the possibility of vowel mergers in Singapore English through the intuitions of English-speaking adults. Following this detailed investigation of the differences between the phonemic system of Singapore English and other prominent varieties of English (i.e., the UK, US) in Chapter 4, we then discuss the implications of understanding the vowel system for phonics instruction, and explore the challenges in developing a phonics-based intervention that is suited to the linguistic context in Singapore.

Chapter 2: Systematic review of reading interventions for children with reading difficulties

Disparity between research and classroom practices

Reviews of reading and literacy development over the years have confirmed the importance of systematic and explicit phonics instruction as an essential part of effective reading instruction (See Rose, 2006; Rowe, 2005; National Reading Panel, 2000 for reviews). However there has been a disparity between what formal research has found and current instructional classroom practices in a number of jurisdictions, including the US and Australia (for e.g., Rayner et al., 2001; Seidenberg, 2017; Snow, 2019). Castles and colleagues (2018) even titled their recent paper “Ending the Reading Wars”, highlighting the perspective that the weight of evidence is now so large in the science of learning that the reading wars should no longer exist. However, counter to this perspective in reading science, some educators and even educational systems continue to advocate for meaning-focused instruction or *Whole Language* approach, and more recently, *balanced literacy*, for which little evidence of effectiveness exists (Moats, 2014; Snow, 2016). Such beliefs persisted in spite of phonics instruction and the science of reading being championed by numerous researchers (Dehaene, 2011; Seidenberg, 2017).

One key difference between systematic phonics and meaning-focused instruction is the underlying assumption of how children learn to read. Systematic phonics assumes that most beginner readers have difficulty learning the English spelling system. Therefore, educators have to explicitly teach the

alphabetic code to beginning readers. By contrast, meaning-focused, or whole language instruction assumes that beginning readers will implicitly acquire word recognition, and text comprehension if they are provided with high-quality and engaging reading materials, just as speech comprehension and production are acquired by infants exposed to high-quality and engaging speech. However, it appears that this kind of implicit acquisition of the writing system from text exposure is only possible for a small minority of children in a typical classroom, leaving the majority unable to read (Moats, 2014).

There has been a growing body of research that supports applying the same principles (i.e. explicit systematic phonics) for reading interventions to support children with reading difficulties (Hulme & Snowling, 2016; Rose, 2006). A major risk factor associated with reading difficulties in children is a deficit in phonological processing (Snowling, 2014). Hence, the main ingredients of an intervention for children with reading difficulties would typically comprise of a combination of phoneme awareness and phonics instruction. However, the evidence should not merely be based on the theory of how it can help children with reading difficulties (Snowling & Hulme, 2011). Instead, before an intervention can be recommended and considered for implementation in schools, there needs to be evidence that the intervention works within the school context. This is in line with intervention science research that successful implementation cannot be assumed despite having theoretically sound interventions (Moir, 2018). Evidence must be based on high standards, with practical implementation guidelines in order for it to be viable in school settings.

In order to establish the best practices for reading remediation of the kind

that could be conducted for struggling readers in a school setting, this chapter reviews the published literature on long-lasting interventions conducted in groups.

Quality of methodological practices in education research

There has been ongoing discussion about the dangers of weak methodological practices in education research for more than two decades. Torgerson & Torgerson (2003) drew parallels between healthcare and education research, highlighting the importance of methodological improvements. Poor methodological practices in the past have resulted in misleading medical conclusions and brought harm to people. Health care treatments are expected to be evaluated with rigour, typically through randomised controlled trials (RCTs) in which people are randomly allocated to the new treatment compared to those in standard practice. For example, when a new medical intervention becomes available, patients would expect that clinical trials on the use of the intervention have been conducted and the intervention is found to be equally if not more effective compared to an existing intervention, before it is made available. Hence, if there is weak methodology in the medical field, consequences would be grave. For example, the use of a particular antibiotic medicine given to premature infants for infection prevention in late 1940s only stopped after RCTs found a damaging effect (Torgerson & Torgerson, 2003). Lives were lost unnecessarily due to failed treatments. Should we expect any less from education interventions? In education, failed interventions might mean that children's early

reading difficulties are not addressed, and become persistent reading difficulties later on in life, resulting in a host of related issues such as academic underachievement, increased risk in mental health problems and poorer employment in future. If children miss out on the windows of opportunities to read due to ineffective interventions, it would be as unethical as failed medical interventions. The opportunity cost for ineffective educational interventions is huge, not merely from the perspective of individuals, but more broadly, from a macro perspective of an education ecosystem, in terms of the amount of resources and funding it takes for schools to conduct these interventions. Methodological reforms in medical practices have existed for years. However, such reforms have not been observed in education research and weak methodological practices still persist up till today.

Gersten and colleagues (2005) and Bishop (2008) proposed that using quality indicators could improve the quality of research and minimise bias. Despite ongoing discussions on these issues since then, weak methodological practices in education research continue to persist till today. Cook (2014) observed that studies with “small n ; small effects; variable research designs, definitions, outcomes and analyses are not uncommon” (p. 239). Misleading conclusions from studies with poor methodological practices might lead to ineffective interventions chosen over more effective ones at the school or state level if educators and policy makers are not aware of or do not have sufficient skills to discern the quality of research. Since interventions cost a substantial amount of resources to be implemented, there is opportunity cost to children and teachers who are implementing the interventions, in terms of time and

energy.

Whether an intervention programme is deemed to be effective needs to be informed by close examination of evidence provided (Duff & Clarke, 2011). A systematic and objective approach to review the quality of the evidence is required. Systematic review is, therefore, a method of surveying the published evidence, through an exhaustive literature review that follows a step-by-step method, with documented procedures before the literature review begins. Cochrane, a well-established organisation that supports evidence-based practices in medicine, defines systematic reviews as “attempts to identify, appraise and synthesize all the empirical evidence that meets pre-specified eligibility criteria to answer a specific research question” (Higgins & Green, 2011, 1.2.2). This approach minimises researcher bias compared to methods such as a traditional literature review in which both evidence and perspective can be heavily influenced by the researcher’s preferences or by the undue prominence of certain perspectives in an unrepresentative subset of the published literature.

In order to establish the quality of evidence for group-based reading interventions targeting struggling readers to be conducted within school setting, we chose to conduct a systematic review. In this systematic review, we therefore aimed to examine the following:

1. What is the quality of the evidence on group-based reading interventions for children with reading difficulties?

2. In line with current theories about the importance of phonics-based intervention, does the evidence for phonics-based interventions outweigh the evidence for alternative literacy interventions?
3. How much evidence is accessible to educators (e.g., open access), given the differing opinions between reading science researchers and educators about how to teach children reading?
4. What evidence is available about reading remediation in bilingual communities?

The results of this review will be useful in informing us of current research gaps and areas that more research would be necessary.

Methods

We conducted an exhaustive search of the existing literature using a pre-determined search strategy (Moher, Liberati, Tetzlaff, & Altman, 2009). We set our inclusion criteria for the systematic review as such:

- i. peer-reviewed journal articles published between 2008 to 2017 – we limited the time frame so that the interventions discussed may be current and has potential for continued use. This time window also coincided with the call for better quality of research in intervention studies, for example, see Bishop (2008).
- ii. group-based interventions lasting at least 10 weeks – the duration of a typical school term is 10 weeks; we limited interventions to group-based so that the interventions can realistically be conducted by teachers in schools in Singapore.

- iii. age of students with reading difficulties between 5 and 9 – this coincides with the age of children who started learning to read up to lower primary level.
- iv. conducted in school settings in English-speaking countries – we limited interventions to those conducted in schools so that the findings can be applied to educators working in schools.

First, we searched five databases (PubMed, PsycINFO, ERIC, Scopus, and SSCI) with the key search terms “reading intervention” and “reading difficulties”. After screening the titles and abstracts, we identified articles that met the inclusion criteria. After we determined the articles that met the inclusion criteria, we assessed the quality of all the articles according to a predetermined checklist of 18 quality indicators using a data extraction table.

PRISMA Protocol

The PRISMA protocol and related preregistration documents were lodged with timestamp on an institutional data repository prior to beginning the search (Leung & Styles, 2018; see Appendix 1 to 3).

Quality indicators

The quality of each article included in the review was evaluated with 18 quality indicators (see Table 2.1). The majority of the quality indicators were adapted from Bishop (2013), addressing methodological problems in intervention studies related to study design, data collection and analysis. Of the

18 indicators, the first five quality indicators address whether a study's design would yield meaningful results that can be further analysed.

Table 2.1. Quality indicators evaluated in the systematic review

Type	No.	Quality Indicator	Source
Essential	1	Were study participants appropriately selected?	1
	2	Were study participants selected by objective criteria?	1
	3	Did the samples give adequate power?	1
	4	Was the intervention adequately described?	1
	5	Was there an appropriate control group?	1
Additional	6	Were there random/matched clinical controls?	1
	7	Did the article provide information on dropouts?	1
	8	Were the primary outcome measures specified?	1
	9	Were reliable and standardized instruments used to measure participants' abilities?	1
	10	Was the measurement blind to group?	1
	11	Were all key data reported?	1
	12	Was the intervention effect appropriately analysed?	1
	13	Was there correction for multiple comparisons?	1
	14	Was the risk of selective reporting bias minimised?	2
	15	Was the interpretation of results fair?	3
	16	Open access to journal article?	-
	17	Open access to data?	-
	18	Open access to materials?	-

Sources: 1 Bishop (2013); 2 Cochrane Collaboration (Higgins & Green, 2011); 3 Bishop (2008)

To fulfil the first quality indicator, participants had to be appropriately selected, as it was important to know whether there were consistencies between the interventions used in the study and participants' characteristics. To fulfil the second quality indicator, participants had to be selected by objective criteria, through the use of clear guidelines on selection criteria. In the context of this systematic review, we looked at whether study participants were identified

using national or state criteria to be considered at-risk of reading difficulties, or appropriate screening tools.

The third essential quality indicator requires the studies to have sufficient power of 0.85. According to G*power calculation, a sample size of 59 per group will be sufficient to detect a medium-sized effect (Refer to Appendix 4 for a detailed description of the G*power calculation). A typical study with a treatment and control group would require 118 participants in total. When power analysis was reported, we checked effect sizes and sample sizes reported to determine whether there was sufficient power.

The fourth essential quality indicator requires that a study provide clear descriptions of intervention such as intervention procedures and materials, allowing for replications. Lastly, to fulfil the fifth essential quality indicator, a study needs to have an appropriate control group included with comparison of pre and post intervention between groups.

We decided to administer the first five quality indicators as a ladder, with studies having to achieve all five essential quality indicators before being assessed further on 13 additional indicators. For example, this means that a study that fails on the third quality indicator will not be assessed on any subsequent criteria beyond that. Only studies that achieve all five essential quality indicators are considered to have 'sufficient' evidence for further evaluation.

The additional quality (6 – 15) indicators addressed areas covering study design, data analysis and data reporting. Indicators 6 to 15 looked into areas such as presence of random allocation of participants to treatment and control

groups; information on dropouts; distinction between primary and secondary outcome measures; assessment methods that are reliable and standardised; presence of blind assessment; reporting of key data; appropriate analysis of intervention effect; correction for multiple comparisons; selective outcome reporting bias minimised; and fair appropriate interpretation of results. For further details about how each indicator was assessed, see Leung & Styles (2018); also found in Appendix 1.

In addition to these indicators of the quality of the research, we also included three indicators (16-18) designed to evaluate how easily an educator could gain access to the research in terms of the article, data and materials. "Open access literature is digital, online, free of charge, and free of most copyright and licensing restrictions" (Suber, 2012, p.4). Benefits of open access literature have been increasingly discussed, especially in academic research fields. Roehrig, Soper, Cox, & Colvin (2018) discussed the positive impact open access literature can have in education research, allowing educators, policy makers and researchers to have access to articles that would typically be locked behind paywall. As such, indicators 16 to 18 regarding open access were identified to be important for the systematic review as we were interested to find out whether articles, data and materials are made available to members of the public without having additional barrier to accessibility.

Results

A total of 1730 studies were obtained by the database searches. We identified an additional 72 studies from Wanzek et al. (2016), which investigated relevant reading interventions within the scope of our systematic review. After

removing 385 duplicate studies, 1417 studies remained. 1258 studies were dropped after examining the titles and abstracts.

Figure 2.1. Study flow diagram. This figure illustrates the flow of information in the review (Moher et al., 2009).

We investigated group-based interventions for children with reading difficulties conducted within school setting, lasting for at least 10 weeks. Out of 160 full-text articles examined, a further 128 were excluded, leaving 32 studies that were included (See Appendix 11 for the 128 articles excluded with reasons). Figure 2.1 shows the flow of information in the review.

After systematically searching the literature, only 32 studies met the criteria for the review. Given that a time period of 10 years was applied in the search, this was an unexpected finding that such a small number of interventions

were found. Table 2.2 summarises the number of studies reporting phonics-based and non-phonics based interventions identified in the systematic review.

Table 2.2. Studies included in qualitative synthesis by type of intervention and quality of evidence.

Type of intervention	Insufficient evidence ⁽¹⁾	Sufficient evidence ⁽²⁾	Total
Phonics	15	11	26
Without phonics	3	3	6
Total	18	14	32

(1) Failed to achieve all five essential quality indicators

(2) Achieved all five essential quality indicators

Out of these 32 studies, there was more evidence for interventions with phonics compared to those without phonics (Table 2.2). Specifically, 26 studies involved intervention with phonics while six involved interventions without phonics. However, a number of the studies failed to achieve the 5 essential quality indicators, which would provide sufficient evidence for a reader to know whether or not the reading intervention made a difference to the struggling readers in their study. Out of 32 studies that qualified for the analysis, only 14 studies were judged to have sufficient evidence to evaluate whether an intervention had an effect while the rest did not. Of those 14 studies, 11 studies including phonics in their intervention and 3 studies did not. As the observed frequencies in two cells were too small for a chi-square test of independence, Fisher's exact test was performed, comparing interventions with and without phonics training. The relation between these variables was not significant in a two-sided test ($p > .99$).

Figure 2.2 gives a detailed breakdown of the number of articles that fulfilled each quality indicator. The first five quality indicators were assessed incrementally, and for those studies that achieved the first five essential quality indicators, all remaining quality indicators were assessed. All 32 studies fulfilled the requirements of the first two essential indicators on appropriate and objective criteria for study participants. 15 studies failed on the third essential indicator to have a sample size with adequate power to detect an intervention effect.

Figure 2.2. Dot plot showing the number of articles that scored for each quality indicator. Each dot represents a single study.

All 32 studies achieved the fourth essential quality indicator, i.e. whether the intervention was adequately described. Three studies did not meet the fifth essential quality indicator, i.e. the use of appropriate control group. Hence, we were left with 14 studies with 'sufficient evidence', which were subsequently assessed on the additional quality indicators.

Most of the studies with sufficient evidence performed well for the following additional quality indicators: (6) random/matched clinical control; (9) standardised assessment methods; (14) selective outcome reporting bias minimized; and (15) appropriate interpretation of results. Performance of studies was moderate for the following additional quality indicators: (7) information on dropouts; (11) all key data reported; and (12) intervention effect appropriately analysed. Studies performed poorly on the following additional quality indicators: (8) primary outcome measures specified; (10) blind assessments; and (13) correction for multiple comparisons. Notably, none of the articles met any of the availability criteria (16 to 18), which explored whether articles, data and materials are available to members of the public without having to make special requests or paid subscription. This means that the articles are not openly accessible to educators and policymakers working with students with reading difficulties.

Table 2.3 shows all 14 studies with sufficient evidence to evaluate whether an intervention had an effect. The studies investigated a large range of outcome measures, suggesting huge variability even within the investigation of

Table 2.3. Studies with sufficient evidence, with the outcome measures reported in the studies.

Studies		Outcome measures														Age in years/ Grade
		Quality score (1)	Expressive Vocabulary	Letter knowledge	Letter Sound	Nonword Reading	Phonological	Rapid Automatized	Reading	Reading fluency	Receptive	Sight Word	Single Word	Spelling	Word reading	
Phonics		14													*	K to 2nd grade
	Kyle, Kujala, Richardson, Lyytinen, & Goswami (2013)	12				*	*				*		*			Mean= 6:7
	Denton, Fletcher, Taylor, Barth, & Vaughn (2014)	12				*			*	*		*				1st grade
	Torgesen, Wagner, Rashotte, Herron, & Lindamood (2010)	12							*							Mean= 6.5
	Morris et al. (2012)	11							*				*	*	*	6- to 9-years
	Duff, Reen, Plunkett, & Nation (2015)	11	*			*	*	*			*		*	*		5- to 7-years
	Lee & Scanlon (2015)	11		*	*	*	*						*	*		Mean= 5:2
	Kamps et al. (2008)	11				*			*				*		*	K
	Hagan-Burke et al. (2011)	10		*		*	*	*					*			Mean= ~5.5years
	Hagan-Burke et al. (2013)	9		*	*	*	*	*								Mean= ~5.5years
	Simmons et al. (2014)	9		*	*			*	*	*	*	*				Mean= ~5.5years
Non- Phonics	Vadasy & Sanders (2008)	13						*	*	*			*			2nd and 3rd grade
	Vadasy & Sanders (2008)	13			*			*	*	*	*		*			2nd and 3rd grade
	(Clarke, Snowling, Truelove, & Hulme, 2010)	13	*						*							8- to 9-years

(1) Quality score represents the studies' performance on all quality indicators.

reading interventions for children with reading difficulties. The studies scored a minimum of four and up to a maximum of nine for additional quality scores, over and above the maximum score of 5 for the essential indicators. Intervention studies involving phonics scored a range of 4 to 9 secondary quality indicators while intervention studies not involving phonics scored 8 secondary quality indicators. In line with the preregistered goals of the research, effect sizes were not reported alongside quality scores. In part, this decision was because studies handled effect sizes differently, and were not easily comparable with varying outcome measures. Students' age in studies also differed.

We originally planned a secondary objective to explore the quality of evidence for reading interventions in bilingual communities. However, only one out of the 32 studies included in the systematic review was from a bilingual community (Hwee & Houghton, 2011). This paper from Singapore failed to meet the first five essential indicators and was not evaluated on additional indicators.

Discussion

The quality of evidence for group-based interventions for children with reading difficulties was much lower than expected. This systematic review found that there are still serious methodological weaknesses in education research even from the past recent decade of reading research. We also found that there was limited information about group interventions targeting children identified as having reading difficulties of the kind that would be suitable to administer in small groups in a school setting.

While there is more evidence supporting phonics-based reading intervention than non-phonics based, the pool of available evidence is small. It is

therefore important to note that more high-quality research targeting children with reading difficulties in small groups will be necessary to establish what constitutes 'best practice' in this arena. Non-phonics reading intervention does not provide support to whole-language or balanced literacy related instruction. Instead, the three non-phonics reading intervention studies are targeting reading comprehension and reading fluency, which can be considered essential skills to remediate once word-reading skills are addressed. They are also targeting older students compared to beginner readers in phonics-based interventions. Thus, a comparison between phonics and non-phonics interventions becomes meaningless as it fails to capture the phonics versus meaning-focused dichotomy present in classroom practices.

Application of quality indicators

Peer-reviewed journal articles should ideally be deemed by educators and policymakers as critical primary sources of information, providing guidance on best practices and interventions based on evidence. As such, it is imperative for educators and policymakers to investigate the quality of evidence in peer-reviewed journal articles. It is also crucial to recognise that research published in peer-reviewed journals differs in the quality of evidence provided (Bishop, 2013; Duff & Clarke, 2011). Quality of evidence, especially those items grouped under essential indicators in our systematic review, need to be in place before results of studies can be understood and interpreted. As highlighted earlier, checklists of quality indicators have been developed, but are not broadly applied in education research by researchers. Applying checklists would help to ensure that research

is designed, conducted and analysed appropriately. In this systematic review, quality indicators were adapted from Bishop (2013), and applied to studies to evaluate the quality of evidence. Drawing upon findings of the systematic review we conducted, quality of evidence that are frequently overlooked in reading research literature are summarised in Table 2.4.

Table 2.4. Quality of evidence frequently overlooked in reading research literature.

Quality indicators frequently overlooked	Type
Having sufficient sample size to detect intervention effect	Study Design
Appropriate control group	
Assessors blinded to group assignment	
Conducted power analysis	Data analysis
Adjustments of multiple comparisons	
Reporting of effect sizes	Data reporting

Open access

The lack of open access to the articles, data and intervention materials presents substantial layers of barrier, as educators do not have access to relevant information. In particular, pay-walled and low-quality studies make it difficult for educators to draw evidence-based conclusions from the existing literature. For example, if educators want to find out about how good phonics-based instruction is in a group setting, they would not be able to read a primary source report unless they have access to paywalled articles. Even for those who have access to the article, lack of open access data would present an additional barrier if one cannot check the data to see if they agree with the published

interpretation. This might prove crucial in convincing sceptical readers who might still not be convinced about the effectiveness of phonics-based intervention. Finally, for educators who have been persuaded by what they have read, and wish to implement an effective intervention, they have no access to the materials used for an intervention. Given the current wide prevalence of internet resources, and limited availability of journal articles, educators typically have no choice but to turn to Internet resources such as websites. Test, Kemp-Inman, Diegelmann, Hitt, & Bethune (2015) reported that amongst 47 websites claiming to provide information on evidence-based practices in special education, about 57% were classified to be trustworthy while the 43% were classified as not trustworthy based on quality and level of evidence. This is alarming as educators need to be critical readers as *knowledge consumers*, and cannot afford to believe what is presented as information without assessing the source and quality of evidence.

Lysenko, Abrami, Bernard, Dagenais, & Janosz (2014) found that educators' attitude towards research was the largest predictor of the use of research-based information in Canadian schools, followed by research expertise (which refers to their ability to read and assess quality of research) as the second largest predictor. This suggests that developing educators' positive attitude and educating them about research would help to bridge the gap between educational research and practice. Though more research is necessary in this area, it is our belief that open access resources can potentially help to enhance educators' attitude towards research through better access to articles, data and materials. We are hopeful that along with better quality of evidence in education

research and open access literature, educators can eventually build up sufficient skills and knowledge in identifying biases present in the literature, and evaluating effectiveness of interventions. In the meantime, one good practice researchers can explore is to provide one-page summaries of research that are easily accessible. One excellent example would be the Open Accessible Summaries In Language Studies (OASIS) initiative (OASIS, n.d.).

Linguistic differences

While there is empirical evidence on reading interventions targeted at children with reading difficulties, the research is primarily based in monolingual English-speaking countries such as the United States and United Kingdom. As reported earlier, no studies from bilingual communities reported research with sufficient quality that it was possible to know whether or not the intervention had an effect. When faced with a lack of quality evidence from bilingual communities, researchers and educators might have to assume that to a large extent, what works effectively in monolingual English-speaking countries would apply similarly to children in bilingual communities. Linguistic differences arising from being brought up in a bilingual community should not be overlooked. In Chapter 1, we have presented some information on how linguistic context in Singapore might be different from monolingual English-speaking countries. For example, there is no clear information on the repertoire of speech sounds of Singapore English. Better understanding is necessary regarding how learning more than one language, in different linguistic contexts, would affect how reading interventions can be better suited for bilingual communities. This will be explored further in the subsequent chapters.

In summary, with reference to the four aims discussed at the beginning of this chapter, we found that:

1. The quality of evidence for group-based reading interventions for children with reading difficulties is limited. More high-quality research for group intervention will help to inform educators in the area.
2. There is more evidence for phonics-based intervention compared to non-phonics based ones. Non-phonics interventions targeted reading fluency and reading comprehension, which are also evidence-based skills to remediate, typically after word-level skills such as phonics have been addressed. Contrary to what we expected to find, there were no intervention studies based on Whole Language philosophy.
3. There is a dearth of open access articles, data and materials in the published literature for the time period that we investigated. However, with rising adoption of open science practices from researchers in the recent years, more open access is to be expected.
4. There is no evidence about reading interventions from bilingual communities. This is an area for more research in future as language exposure in two or more languages might affect how learning is best delivered.

Recommendations for researchers

First, educational researchers conducting intervention studies would need to recognise the importance of rigorous scientific research in ensuring the

validity of intervention findings. Higher quality research is necessary, with particular attention to appropriate and objective selection of study participants, adequate sample size for power, clear descriptions of intervention and the use of an appropriate control group. Second, educational researchers should improve the overall quality of research by adhering to the highest standard of research reporting. Change is particularly important in the areas of specifying primary outcome measures, making measurement blind to group, and correcting for multiple comparisons. Both of these conditions (i.e. higher quality research and highest standard of research reporting) can be achieved through preregistration, in which methods, sample sizes and analysis plan would be documented in advance before interventions are implemented. Thirdly, educational researchers should consider the use of open access to increase educators' and policymakers' access to peer-reviewed journal articles, data and materials, thus minimising barriers to knowledge for the main stakeholders who work closely with students. However, it should be noted that although open sharing of materials is highly recommended in the new era of Open Access research, it may not have been possible at the time these previous studies were conducted, or due to copyright restrictions. Fourth, plain-language summaries of peer-reviewed journal articles can help to bridge the researcher-practitioner gap by having information provided in a succinct, accessible manner, thus helping educators to make the link between research and application of educational practices. Finally, given the lack of bilingual intervention studies on group-based phonics instruction, this is an area that researchers in bilingual communities can

consider investigating. Other possible exploration could be how phonics-based interventions to learn English can be influenced by differing linguistic contexts.

Recommendations for educators

First, educators need better professional development on critical reading of research so that they can learn to rigorously understand and evaluate studies. Given that such skills take time to learn and apply, educators might find it more beneficial for critical reading of research to be applied to relevant areas of work, depending on their specialisations. Second, educators should start requesting researchers whose materials are not open access to produce accessible summaries so that critical evaluation can be possible. This could be seen as an interim practice that can benefit both researchers and educators, even as education research is starting to see the benefits of open access.

Third, educators can team up with researchers to test the effectiveness of interventions in schools and hence increasing transferability of research to practice. However, it is crucial to insist on best practices (including control groups, high power and preregistration) to ensure the highest quality evidence. The use of a control group in education research has not always practised. Admittedly, children should not be left behind and be denied an effective intervention. However, if a study did not include a control group, one would never be able to determine whether an intervention works or not. While it can be challenging in education research to justify for control groups when one treatment is thought to be more effective, thus making the decision to have a control group with a less effective intervention questionable, there are

alternative designs using control groups. A treat/wait design can be used instead, where one group receives intervention while the other group waits and receives intervention later. Another possibility is the use of 'current best practice' control group where new intervention can be compared against a control group receiving the best current intervention.

Such collaborations, in my opinion, have potential to break down barriers between researchers and educators. Lastly, educators need to exercise caution when adapting intervention materials from different contexts (e.g., monolingual materials for bilingual learners), given the lack of evidence.

Chapter 3: Phonology of Singapore English: An investigation of readers' intuitions about vowel sounds

Introduction

As highlighted in Chapter 1, many English-speaking Singaporeans do not distinguish between: i) /i:/ *peach* and /ɪ/ *pitch*, ii) /ɑ:/ *cart* and /ʌ/ *cut*, iii) /ɔ:/ *caught* and /ɒ/ *cot*, and iv) /e/ *bed* and /æ/ *bad* (See Figure 1.5 for a visual comparison between Singapore English and SSBE). Existing descriptions of Singapore English suggest that 'pitch' and 'peach' would be pronounced the same, whereas the vowels in these words would sound noticeably different when spoken by speakers of other varieties of standard English (OASIS, n.d.).

Literature on Singapore English has consistently shown that speakers of Singapore English use a reduced repertoire of vowels compared to speakers of other varieties of English. Many speakers of Singapore English do not distinguish between vowels /æ/ and /e/ when asked to read out individual words (Suzanna & Brown, 2000). Acoustic studies from (Suzanna & Brown, 2000) found that distinction between various monophthong vowel pairs were reduced in conversational speech of educated speakers of Singapore English, as compared to speakers of other varieties of English. When vowel mergers occur, speakers typically do not show a consistent difference in their pronunciations and do not tell the sounds apart (Gordon, 2004). So far, researchers have discussed the possibility that certain vowel pairs have 'merged' into single phonemic categories in Singapore English (e.g., tense/lax vowel pairs like in "peach" and "pitch"), due to substantial overlap in where they are articulated in vowel space.

However, there was no definite agreement on which vowel sounds are merged in Singapore English and how they are articulated in terms of acoustic values.

Current literature on the vowel sounds of Singapore English is limited. (Gordon, 2004) suggested a '*minimal vowel inventory*' for reference: [i] (*fleece, kit*), [e] (*face*), [ɛ] (*dress, trap, square*), [ʌ] (*palm, strut*), [ɔ] (*lot, thought*), [oʊ] (*goat*), [u] (*foot, goose*), [ə] (*nurse*), [aɪ] (*price*), [ɔɪ] (*choice*), [aʊ] (*mouth*), [iə] (*near*), [uə] (*poor*). However, the authors highlighted variation across speakers and suggested that one should not make assumptions about all speakers of Singapore English pronouncing these vowels in the same manner. Low and Brown (2005)'s initial vowel inventory was drawn up mainly from impressions and observations made by them.

Deterding (2007) suggested another system of eight monophthong vowels for Singapore English, based on data from one participant. The author recognised limitations of using one speaker, but decided that it allowed for an in-depth investigation of features of Singapore English. The vowel system reported in Deterding (2007) share high degree of similarities with Low & Brown (2005): [i] (*beat, bit*), [e] (*say*), [ɛ] (*send, sand*), [ə] (*heard, her*), [ʌ] (*cart, cut*), [u] (*pool, pull*), [o] (*know*) and [ɔ] (*caught, cot*). The exception would be the classification of [oʊ] in Low & Brown (2005) compared to [o] in Deterding (2007). Deterding (2007) explained that both [e] in 'say' and [o] in 'know' do not change in quality in Singapore English pronunciation, compared to other standard varieties of English. It is uncertain why Low & Brown (2005) considered [e] as a monophthong vowel but not [oʊ]. It should be noted that Deterding (2007) cautioned that even with the classification of vowels, predicting whether a word

contains a particular vowel can be problematic. Table 3.1 shows an analysis of the classification of vowels between the two papers discussed above.

Table 3.1. An analysis of classification of vowels in Low & Brown (2005) and Deterding (2007).

Analysis	Low & Brown (2005)	Deterding (2007)
Similar – merger of i:/ɪ	[i] (<i>fleece, kit</i>)	[i] (<i>beat, bit</i>)
Similar	[e] (<i>face</i>)	[e] (<i>say</i>)
Similar – merger of æ/e	[ɛ] (<i>dress, trap, square</i>)	[ɛ] (<i>send, sand</i>)
Similar – merger of ɑ:/ʌ	[ʌ] (<i>palm, strut</i>)	[ʌ] (<i>cart, cut</i>)
Similar – merger of ɔ:/o	[ɔ] (<i>lot, thought</i>)	[ɔ] (<i>caught, cot</i>)
Deterding (2007) considered it to be a monophthong in SgE.	[oʊ] (<i>goat</i>)	[o] (<i>know</i>)
Similar - merger of u:/ʊ	[u] (<i>foot, goose</i>)	[u] (<i>pool, pull</i>)
Similar	[ə] (<i>nurse</i>)	[ə] (<i>heard, her</i>)
Not monophthong vowel	[ai] (<i>price</i>)	-
Not monophthong vowel	[ɔi] (<i>choice</i>)	-
Not monophthong vowel	[au] (<i>mouth</i>)	-
Not monophthong vowel	[iə] (<i>near</i>)	-
Not monophthong vowel	[uə] (<i>poor</i>)	-

The ones highlighted in orange are of interest in the current study. The items highlighted in purple are classified similarly in both studies. Cells corresponding to ‘Not monophthong vowel’ group are left uncoloured, these are diphthongs.

Although Low (2012) updated the vowel inventory after considering experimental findings on Singapore English, the updated inventory did not present an unifying picture about how vowels are pronounced in Singapore. For example, keyword ‘trap’ which is pronounced [æ] in British English, was presented as [æ] and [ɛ] in two separate experimental studies. Similarly, ‘bath’ which is pronounced [ɑ:] in British English, was presented as [ɑ:] in one study (Lim, 2004) and not tested in another study (Low & Brown, 2005).

Investigations of pronunciation of vowels were typically made using observations by linguistic researchers. Only a few investigations involve

instrumental designs, usually through measurements of recorded speech of speakers, either during conversations or of individual words read from a list. For example, Suzanna & Brown (2000) examined pronunciation of eight female participants while Deterding (2003), on which Figure 1.5 was based on, examined pronunciation of five male and five female participants. Both studies depended on a small number of participants, making it hard to generalize the findings from the studies. Deterding (2005) investigated vowel patterns of Singapore English in 38 participants. However, the focus of the paper was on distinctive patterns in specific words and not all monophthong vowels were investigated. Deterding (2007) investigated monophthong vowels, focusing on conversational data with only one participant. As summarised in Table 3.2, the four studies used varying sources of data from which vowel information was extracted.

Table 3.2. Instrumental studies on Singapore English.

Instrumental studies	Vowel pairs investigated	<i>n</i>	Source
Suzanna & Brown (2000)	æ/e	8	Individual words (word list used not available); conversational data
Deterding (2003)	i:/ɪ ; æ/e ; ɑ:/ʌ ; ɔ:/o ; u:/ʊ ; ɜ:/	10	Conversational data
Deterding (2005)	eɪ/e ; ɑ:/ʌ ; ʊə/o:/	38	12 sentences which contain target words
Deterding (2007)	i:/ɪ ; æ/e ; ɑ:/ʌ ; ɔ:/o ; u:/ʊ ; ɜ:/	1	Conversational data

In short, the previous literature documenting vowel patterns was mainly observation studies conducted with only few studies involving acoustic analysis. In addition, there were small numbers of participants, varying from one to 38 participants. This makes it challenging to draw conclusions about possible vowel mergers in Singapore English from previous studies. Furthermore, many studies

have relied upon either a limited number of words or depended on conversational data, making it difficult to determine the range of acoustic measurement. In order to draw any possible conclusions about vowel mergers in Singapore English, we would require information about acoustic measurement and the extent of overlap between vowels in formant values. All studies highlighted in Table 3.2 collected acoustic measurements but differed in the vowel pairs under investigation, methodology, and number of participants. The question of whether there are vowel mergers in Singapore English is a pertinent one for research, as the basis of phonics instruction depends on teaching grapheme-phoneme correspondence. Hence, any vowel differences in Singapore English compared to phonics designed for other varieties of English have implications for instruction.

The current study

Given that there is limited information available about the sounds in Singapore English, a more systematic investigation of the variety of English spoken in Singapore will be required. This will enable a better match between English spoken in Singapore and materials that can be used in reading remediation targeted for Singaporean children. We present an experiment that explores intuitions of English-speaking adult Singaporeans on different vowel pairs that are contrastive in other standard varieties of English (e.g. standard Southern British English). This investigation will enable us to map out a phonemic model of Singaporean English vowel-space that is aligned to speaker intuitions.

We asked skilled adult Singaporean readers to make decisions about whether minimal pairs, which are pairs of words with different meanings that differ in only one sound in the same position of each word (e.g. “pin” and “pan”), are pronounced the same or different in an onscreen reading task.

We predicted that:

- 1) Speakers of Singapore English would judge written minimal pairs containing vowels suspected of merger as more similar compared to minimal pairs containing vowels not suspected of merger.
- 2) Different vowel pairs might show different degrees of merger.
- 3) Speakers of Singapore English would take longer to respond to written minimal pairs containing vowels suspected of merger than they would for minimal pairs containing vowels not suspected of merger.

The predictions were preregistered and documented (Leung & Styles, 2019; Appendix 5).

Method

Prior to data collection, questions from Open Science Framework were used to create a document detailing the hypotheses, proposed data-handling, and analysis plan. This preregistration document is now archived on NTU's institutional Dataverse (see Leung & Styles, 2019; also found in Appendix 5). The number of participants for the study was pre-determined using G*Power to detect a medium-sized effect of 0.5.

Participants

As preregistered, 60 participants were recruited. All participants in this experiment were English-speaking Singaporeans with typical hearing. Six participants were undergraduates enrolled in an introductory psychology module and were given course credits for their participation, and the remaining 54 participants were obtained through a combination of university campus recruitment and convenience sampling from the general public. Recruitment posters were put up on campus and sent digitally to interested parties. These participants were paid \$5 upon completion of the study. The participants were aged between 19 and 46 years old (M age = 26 years 1 month, SD = 5.80 months). 38 were female and 22 were male. All participants described themselves as bi- or multi-lingual. Out of those who described themselves as Bilinguals (71.7%), the breakdown of language identity was as follows: English-Mandarin (N = 36), English-Malay (N = 4) and English-Tamil (N = 3). For those who described themselves as Multilinguals (28.3%), the breakdown of language identity for the dominant languages in Singapore (in addition to other languages) was as follows: English-Mandarin (N = 13), English-Malay (N = 2) and English-Tamil (N = 1). In addition, one participant reported being multilingual in English-Mandarin-Malay.

In terms of proficiency across their languages, participants provided self-rated proficiency for the languages that they speak and understand on the following scale, ranging from 1 to 7: 1 indicates understanding/speaking just a few words, and 7 indicates native-level proficiency in understanding/ speaking. The mean proficiency for understanding English was 6.63, with scores ranging

from 4 to 7; mean proficiency for speaking English was 6.58, with scores ranging from 4 to 7. The mean proficiency for understanding Mandarin was 5.22, with scores ranging from 1 to 7; mean proficiency for speaking Mandarin was 4.77, with scores ranging from 1 to 7. The mean proficiency for understanding Malay was 3, with scores ranging from 1 to 7; mean proficiency for speaking Malay was 3.52, with scores ranging from 1 to 7. The mean proficiency for understanding Tamil was 6.75, with scores ranging from 6 to 7; mean proficiency for speaking Tamil was 5.40, with scores ranging from 1 to 7.

In terms of sensory processing and developmental factors, 55 participants reported normal or corrected to normal vision, 5 appeared to have misinterpreted the question on vision though they were able to answer questions presented on the computer screen. All participants reported normal hearing, and that they had no diagnosis of dyslexia or learning difficulties. All participants were assigned to only one condition, as the experiment has a repeated measure design. The NTU Institutional Review Board has reviewed the study for ethics approval.

Test stimuli

To test whether vowel categories are merged for speakers of Singapore English, participants were asked to make decisions about whether minimal pairs are pronounced the same or different. Table 3.3 shows the list of vowel pairs suspected of merger, as well as vowels pairs not suspected of merger. Words with vowels pairs not suspected of merger are used as control. Homophone-pairs, which are not minimal pairs, such as 'cell' and 'sell', were included as filler

trials as they sound the same but are spelt differently. The addition of control words and filler trials ensured that participants gave a mixture of ‘same’ and ‘different’ responses in the experiment.

For each vowel-pair, we aimed to identify 10 contrastive word pairs comprised of high frequency words familiar to Singaporeans. Frequency counts were based on the Zipf scale using the SUBTLEX-UK word frequencies (Heuven, Mandera, Keuleers, & Brysbaert, 2014). Due to an insufficient number of available words, the vowel-pairs of contrast ‘ʊ/u’ and ‘ɑ:/ʌ’ had only had 7 word pairs each. There were a total of 94 pairs of words as experimental items, with 44 pairs of vowel pairs suspected of merger, and 50 pairs of control vowel pairs (vowel pairs not suspected of merger) (See Appendix 8 for the list of word pairs used in the experiment).

Table 3.3. Vowel pairs with suspected merger and no-merger in Singapore English

Contrast Type	Contrast	Example Word Pairs
Suspected merger	i:/ɪ	beet / bit
	æ/e	bat / bet
	ɑ:/ʌ	Bart / but
	ɔ:/o	bought / bot
	u:/ʊ	suit / soot
No-merger	ʊ/o	look / lock
	æ/ʌ	hat / hut
	ɪ / e	bit / bet
	ɔ:/ɑ:	bought / Bart
	u:/i:	boot / beet

‘Contrast’ refers to the specific pair of vowel differences.

These experimental items were interspersed with 56 pairs of homophones, which function as filler trials. The experimental and filler items were intermixed, and were presented in a novel random order for each participant. (See Leung & Styles, 2019; Appendix 5, for a discussion of the choice of word pairs). The decision task was scripted and presented using jsPsych (de Leeuw, 2015), which is a web-based platform used to present the word stimuli and collect data on participants' responses for each trial ("same"/"different") and reaction time (RT) taken to respond.

Experimental design

The participants were tested in individual sessions that lasted about 20 minutes. Participants were seated in front of a monitor displaying the stimuli (See Figure 3.1). At the end of the pronunciation judgment task, participants also completed a word-production task for analysis in a separate study, and completed demographic questions. In each trial of the decision task, participants were asked to judge whether two written words presented on the monitor are pronounced the same or different to them for each trial. They were told that there was no need to spend too long to decide. During the task, participants were asked to indicate whether the words are pronounced the "same" or "different" by pressing two keys (i.e. left key "a" and right key "l") on the keyboard using their index fingers. Cue words "same" (in a green box) and "different" (in a red box) were shown on either sides of the screen for every trial, to remind participants which key should be pressed for their responses. Figure 3.1 provides an illustration of a trial. For each trial, the response of "same" and "different" was captured along with the RT taken to provide the response. The binary decisions

about whether minimal pairs of a particular contrast sound the “same” or “different” were used to compute a composite score per participant, with the score of 100% different when all minimal pairs within the same vowel contrast were judged to have different pronunciations and 100% same when all minimal pairs were judged to have the same pronunciations. For example, under a given contrast (e.g. i:/ɪ), a participant might have judged six out of ten word pairs to be the “same” (and the remaining four to be “different” as the choice was binary). This would mean that the participant would get a score of 60% same for the contrast i:/ɪ.

To avoid the potential confound of handedness of participants having an effect on reaction time (for e.g., right-handed participants possibly responding faster when pressing the key using their right hand), half the participants began with ‘same’ response on the left and ‘different’ response on the right. The other half of the participants began with ‘same’ response on the right and ‘different’ response on the left. The participants were randomly assigned to begin the task in either way.

Figure 3.1. A representation of a trial seen by participants on the computer screen. Half of the participants were first presented with the ‘different’ response on left side of the screen and the ‘same’ responses on right of the screen, as shown here. The other half were counterbalanced and first presented with ‘same’ response on left and ‘different’ response on right. At the halfway point, the response sides were swapped for all participants.

Participants were presented with instructions on the screen and then went through 10 practice trials to get familiarised with task instructions before the actual task started, with a total of 150 trials. After completing half the trials, participants were informed that response sides were swapped, before they completed the rest of the trials.

For the word-production task, participants were asked to read out written words that were presented one at a time on the screen. Participants' responses were recorded with a head-worn condenser microphone (AKG Pro Audio C520 L) on the Zoom H4n Pro recorder in a sound attenuated audio recording room. The experimenter controlled the speed of presenting each word such that the next word would be presented only after the participants had finished reading the word on screen. The word-production task was designed to explore a secondary research question: whether participants' actual articulation would match their intuitions in the earlier decision task. This secondary research question was also pre-registered for analysis in a future study.

At the end of the experiment, participants were instructed to complete a questionnaire on their language background and other relevant demographic information (See Appendix 6). The participants were also asked to complete the PEBBLES survey for Hearing and Vision (de Leeuw, 2015) in which they were asked about their hearing and vision (See Appendix 7). Both questionnaires were presented on Qualtrics, an online survey platform (Styles, 2018).

Hypothesis and analysis plan

We preregistered two hypotheses for the current investigation:

H1a. We predict people will give more ‘same’ responses for minimal pairs containing vowels suspected of merger (test vowel pairs) than they will for minimal pairs containing vowels not suspected of merger (control vowel pairs).

H1b. We also predict that different vowel pairs may show different degrees of merger. This prediction is exploratory, and we will report the percentage of ‘different’ responses for each vowel pair, as a way of describing the degree of merger.

H2. We predict that people will be slower to respond for minimal pairs containing vowels suspected of merger than they will be for minimal pairs containing vowels not suspected of merger.

We preregistered the plan to perform a generalized linear mixed-effects model whether Merger (merged, unmerged) significantly influenced Participants’ response in the binary decision about whether a word pair was pronounced the same or different. Since the variable was binary, the model was fitted to a logistic function, with Merger (merged, unmerged) as a fixed effect. Intercepts for items and intercepts for subjects, as well as by-subject random slopes for the effect of Merger were entered into the model as random effects¹. For further details of the analysis plan, see Appendix 5. There were no

¹ The original preregistration contained an error. In our preregistered analysis plan, we described the mixed models that we would use. Following preregistration and prior to analysis, we realised that since word pairs belong to one of the two merger categories (vowels that are merged or not merged), only random intercept (not random slope) is possible for mixed models analysis as the intersection of merger and participant leaves no residual variance for an individual ‘slope’. We therefore reported the reduced mixed model that best matches our research design.

differences in results when comparing the preregistered analysis plan and actual analysis conducted (See Appendix 9 for data analysis conducted).

Results

Data exclusion

Following pre-registered data exclusion procedures (see Appendix 5), one participant was excluded, as the participant's performance on homophone-pairs was substantially different from responses of the majority.

If participants were able to follow task instructions, we would expect all homophone-pairs should have been rated as the same for the majority of trials, as homophones sound the same by definition. Participants would have a score of 1 if they rated all homophone-pairs as sounding the same. On average, participants rated homophone-pairs as being the same more often than different with a score of .79 ($SD = .15$). One participant was an extreme outlier, rating homophone-pairs as sounding different more often than sounding the same across the 56 homophone-pair trials, with mean of .23 (See Preregistered procedure for determining outliers in Appendix 5). Hence, this participant was removed from further analysis.

Possible vowel merger

To investigate Hypothesis 1a, which predicted that people will give more 'same' responses for test vowel pairs (suspected of merger) compared to control vowel pairs (not suspected of merger), we used lme4 package (Qualtrics, 2005) in R (R Core Team, 2018) to perform a generalized linear mixed-effects model

whether Merger (merged, unmerged) significantly influenced Participants' response in the binary decision about whether a word pair was pronounced the same or different. Since the variable was binary, the model was fitted to a logistic function, with Merger (merged, unmerged) as a fixed effect. Intercepts for items and intercepts for subjects, as well as by-subject random slopes for the effect of Merger were entered into the model as random effects.

Following the procedure of comparing the final model against the matching model without the critical experimental variable (Merger), as recommended by (Bates, Mächler, Bolker, & Walker, 2015), we removed non-significant factors to identify the best-fitting model with Merger as a fixed effect. To establish whether the respective fixed effect is a significant predictor of participants' % difference score, Likelihood Ratio Test was used to obtain p-values (See Appendix 9 for data analysis conducted).

As predicted, we found significant difference between the model with and without merger, $p < .0001$. Participants consistently chose 'same' more often for vowel-pairs suspected of merger, than for vowel-pairs not suspected of merger (Merger: $M = 52\%$, $SD = 50\%$; Non-Merger: $M = 1\%$, $SD = 9\%$). As preregistered, the percentage of same-different responses for each vowel pair is reported in Figure 3.2, as a way of documenting the degree of merger for different vowel pairs. Figure 3.2 shows that control vowel pairs (those not suspected of merger) were judged as "different" 100% of the time by the majority of participants. In contrast, vowel pairs suspected of merger received more variable in pronunciation judgments, with participants' judgment ranging from 100% different to 100% same. For visual comparison, homophones are also included in

the visualization, where it is clear that they were judged to have the same pronunciation most of the time, by the majority of participants albeit with a wide spread of responses.

Figure 3.2. Dot plot of the Same/Different judgments made by each participant for word pairs in each contrast. The composite score for a particular contrast per participant was computed based on the binary decisions about whether they think minimal pairs of a particular contrast sound the “same” or “different”, meaning that 50% represents chance responding for a particular contrast. Each dot represents the composite score for a contrast for a single participant, jittered to allow overlapping dots to be seen more clearly. The degree of jitter along the x-axis and y-axis is 0.2 and 0.01 respectively. The contrasts are grouped according to the respective categories.

We conducted one-sample t-tests to compare each individual’s responses for each contrast type against the chance level of 50%. As the analysis was not preregistered, it was exploratory by nature. The responses for contrasts not suspected of merger were significantly below chance, i.e. almost all of these responses were ‘different’ ($M = .0088$, $SD = .020$, $t(58) = -187.8$, $p < .000001$). The responses for contrasts suspected of merger were not significantly different from chance, i.e. people gave a mixture of ‘same’ and ‘different’ responses

($M = .52$, $SD = .027$, $t(58) = .48$, $p = .64$). The responses for homophones were significantly different from chance, i.e. the majority of people gave 'same' responses ($M = .79$, $SD = .15$, $t(58) = 14.90$, $p = .00$).

Figure 3.3. Line plot of the mean Same/Different judgments made by participants according to language exposure.

Given that participants reported a variety of language backgrounds, we also evaluated whether language backgrounds influenced response patterns. Figure 3.3 suggests that the participants differ in same/different judgments, depending on their language exposure. For participants with Malay language exposure, there seems a dip in how similar the $\alpha:/\Lambda$ contrast is relative to their performance on other contrasts. For participants with Tamil language exposure, there appears to be a consistent tendency to judge words to be more different

compared to those with Malay or Mandarin exposure. However, given the very small number of participants with Malay or Tamil language exposure, this study should be treated with caution until further replication with representative sample sizes in the future. It should also be noted that this analysis was not preregistered in the current study and is exploratory in nature.

Degree of vowel merger

To investigate Hypothesis 1b, in which we predicted that different vowel pairs might show different degrees of merger, we used lme4 package (Bates et al., 2015) in R (R Core Team, 2018) to perform a generalized linear mixed-effects model of the relationship between different vowel contrasts and participants' response. Since the variable was binary, the model was fitted to a logistic function with Contrast as a fixed-effect factor with 5 levels (i:/ɪ, æ/e, ɑ:/ʌ, ɔ:/o, u:/ʊ). Intercepts for items and intercepts for subjects, as well as by-subject random slopes for the effect of Contrast were entered into the model as random effects.

Following the procedure of comparing the final model against a matching model without the critical experimental variable (Contrast), we removed non-significant factors to identify the best-fitting model with Contrast as a fixed effect. To establish whether the fixed effect is a significant predictor of participants' % difference score, the Likelihood Ratio Test was used to obtain *p*-values (See Appendix 9 for data analysis conducted).

The analysis did not reveal significant difference between the model with and without contrast, $\chi^2 = .63$, $df = 1$, $p = .43$. For different vowel contrasts, participants did not differ significantly in their responses ('i:/ɪ': $M = 40\%$,

Figure 3.4. Breakdown of the mean same/different judgment for contrasts suspected of merger.

Figure 3.5. Breakdown of the mean same/different judgment for contrasts not suspected of merger.

Note: (1) 'as us' was misclassified as minimal pair.

Figure 3.6. Breakdown of the mean same/different judgment for homophones. Notes: (1) ‘poll’ ‘pole’ and ‘grown’ ‘groan’ are homophones in Standard Southern British English but might not be homophones in different varieties of English. (2) ‘hairy’ ‘harry’ are homophones only in some varieties of Standard English that feature a vowel merger. (3) ‘pause’ ‘paws’ and ‘raise’ ‘rays’ are homophonous in Singaporean English where word final sibilants are voiceless.

$SD = 49\%$; ‘æ/e’: $M = 65\%$, $SD = 48\%$; ‘ɑ:/ʌ’: $M = 51\%$, $SD = 50\%$; ‘ɔ:/o’: $M = 50\%$, $SD = 50\%$; u:/ʊ’: $M = 48\%$, $SD = 50\%$). This means that the vowel pairs suspected of merger did not differ significantly in how merged they are.

Figure 3.4 and 3.5 show the mean judgment scores of all word pairs across all participants for contrasts suspected of merger and not suspected of merger respectively. Comparing both figures visually suggest that participants judged word pairs of these types of contrasts very differently– word pairs for contrasts not suspected of mergers were judged as “different” almost all of the time while word pairs for contrasts suspected of mergers were judged with a mixture of responses. Interestingly, rates for individual word-pairs did differ across the stimulus set, suggesting that there may be item-level differences in

how individual words are thought to be pronounced by speakers of Singapore English. This pattern is even more pronounced in judgment scores for the homophones, shown in Figure 3.6.

Figure 3.7 shows the line plots for all participants for contrasts suspected of mergers. This suggests that participants' responses for contrasts suspected of mergers does not remain similar for different contrasts, and there is no consistent pattern across the given set of items.

Figure 3.7. Line plot of the Same/Different judgments made by each participant for all word pairs in each contrast, arranged in order of decreasing number of “same” judgment.

Speed of decisions

We predicted that people would respond slower for minimal pairs containing vowels suspected of merger, compared to those not suspected of merger. In order to approximate normal distribution, RTs were inverse-transformed before analysis was conducted, following pre-registered procedures to choose between log or inverse transformation, based on visual inspection of histograms and values of skew and kurtosis. Extreme outliers for reaction time were then trimmed both at the group level as well as at the individual level.

To investigate Hypothesis Two, we used lme4 package (Bates et al., 2015) in R (Bates et al., 2015) to perform a generalized linear mixed-effects model of the relationship between Merger (merged, unmerged) and reaction time. Merger (merged, unmerged), and word naming times of word pairs were entered into the model as fixed effects. Intercepts for items and intercepts for subjects, as well as by-subject random slopes for the effect of Merger were entered into the model as random effects.

Participants did not take significantly more time to respond to vowel-pairs suspected of merger, compared to vowel-pairs not suspected of merger (Merger: $M = 2210.6\text{ms}$, $SD = 6403.3\text{ms}$; Non-Merger: $M = 1597.7\text{ms}$, $SD = 7161.5\text{ms}$). The analysis did not reveal significant difference between the model with and without Merger, $p = 1$. This means that speakers of Singapore English did not take longer to respond to minimal pairs containing vowels suspected of merger compared to minimal pairs containing vowels not suspected of merger. Figure 3.8 shows the back-transformed reaction time according to the contrast. Our findings suggest that skilled speakers of Singapore English did not

find deciding about the pronunciation of word-pairs suspected of merger more difficult than they found deciding about vowels not suspected of merger. This further suggests that our participants did not find these word pairs ambiguous or confusing in any way.

Figure 3.8. Back-transformed reaction time according to Contrast. Each dot represents the response pattern of a contrast for a single participant, jittered to allow overlapping dots to be seen more clearly. The degree of jitter along the x-axis and y-axis is 0.1 and 0.01 respectively.

Discussion

To our knowledge, our study is the first experiment to explore the intuitions of speakers of Singapore English using minimal pairs, focusing on monophthong vowels. We found that participants rated minimal pairs containing vowel pairs suspected of merger differently from those not suspected of merger. Specifically, they gave more “same” responses for minimal pairs containing vowel pairs suspected of merger compared to vowel pairs not suspected of

merger. As shown in Figure 3.2, there is a lot more variability across speakers in intuitions for minimal pairs with suspected vowel merger, compared to those not suspected of vowel merger. Participants rated minimal pairs with suspected vowel merger, with responses ranging from 100% different to 100% same. For example, 15 participants judged all the words containing “i:/ɪ” (suspected of merger) as sounding “different” while five participants judged them all as sounding the “same”. The remaining participants varied in their judgments along the entire spectrum. In comparison, 54 out of 59 participants judged all the words containing “ɪ /e” (not suspected of merger) as sounding “different” (See Figure 3.2).

(R Core Team, 2018) highlighted that in the case of vowel mergers, speakers typically do not show consistent differences in their pronunciation and do not tell the sounds apart. Hence, the concepts of “production” and “perception” cannot be teased apart. As the current study focused on speaker intuitions about their own pronunciations of these words, future analysis of participants’ actual production of the words from the related audio recording task (recorded in the same session, for analysis in future study) will provide valuable evidence about whether speakers’ intuitions actually match their pronunciations. Gordon (2004) noted that for a minimal pair test paradigm, the task could result in speakers producing pronunciations that are more formal than those of casual speech, and suggested comparing results from minimal pair tests along with speech from more casual contexts. If participants’ production of the words shows evidence of a merger even in formal contexts, then it can be assumed that pronunciations would be even more merged in informal contexts.

For this study, as participants were asked to judge isolated words, it may be an inaccurate representation of actual speech perception or production as we do not typically read words in isolation in everyday language use.

While further investigation incorporating the use of a corpus will add to our understanding of the vowel system in Singaporean English, this study clearly demonstrates that skilled readers in Singapore do not treat tense/lax word pairs the same way as they do minimal pairs containing other contrasts, and suggests some degree of vowel merger is underway in Singapore English.

It should be noted that we have only investigated monophthong vowels in this study. Diphthong vowels, which begins with one vowel and moves towards another (e.g. “face”), are also found to be pronounced differently in Singapore English compared to other varieties of English (Low & Brown, 2005). For example, “fairy” is pronounced by Singaporeans to sound similar to “ferry”. Further investigations into both monophthong and diphthong vowels would be necessary before a phonemic inventory of Singapore English vowel sounds can be meaningful proposed.

Whether there is presence of vowel mergers in Singapore English would affect the structure of the linguistic knowledge, which in turn affects the content of phonics instruction. The design of phonic instruction involves aligning the sound system to the English alphabetic script, so that grapheme-phoneme correspondence can be introduced in an optimal manner. However, most reading instruction programmes used in Singapore are purchased directly from overseas and hence heavily dependent on US or UK pronunciations. For example, vowels taught at the start of the phonics programme include the monophthong vowels

in Study 3, which we found to have a lot of variability in intuitions of speakers of Singapore English.

In summary, we found that people responded differently to vowel pairs suspected of merger compared to vowel pairs not suspected of merger.

Furthermore, although we predicted that different vowel pairs might have different degrees of merger, we found large variability across all vowel pairs suspected of merger.

Figure 3.9. Shifting language usage towards more English. *Reproduced CC-BY from (Low & Brown, 2005), Singapore's languages at a glance: A visual overview.*

Shifting language use in Singapore over time has been observed by researchers (Low, 2014; Styles, 2019b), with more English being spoken at home across the three main ethnic groups- Chinese, Malay, Indian. Figure 3.9 shows that more households report using English as the main language used at home when comparing trends between 2000 and 2010. With English being their dominant language, the young Singaporeans might differ from the older generations in terms of language use. It would be worth exploring whether

intuitions about vowel sounds and actual pronunciations differ according to different age groups, reflecting influence from the varying amount of different languages spoken by the individuals of various age groups. For example, (Low, 2014; Styles, 2019b) noted that in a student-conducted study, older speakers performed differently compared to younger speakers in speech production, with younger Singaporeans using the post-vocalic /r/ more frequently. Another possible exploration in future would be whether vowel intuitions and pronunciations differ according to different language backgrounds. Deterding & Poedjosoedarmo (2000) reported that Singaporeans could identify ethnically distinct features when speakers were engaging in less formal speech.

Recommendations for researchers

First, further research on vowel intuitions and pronunciations in Singapore, as discussed earlier, is critical for shaping concrete recommendations to educators on how best to teach phonics-based reading in Singapore. The acoustic data collected from the word-production task, which have not been analysed in this thesis, will allow us to have a better understanding of how vowel intuitions relate to individuals' actual pronunciations. For example, we can investigate whether participants' judgments were consistent with actual pronunciations of words. In addition, acoustic analysis of pronunciations would inform us about overlaps in vowel space.

Second, we need translational research that would inform educators about linguistic features that English-speaking children in Singapore exhibit. Given that we will have more information about the Singapore linguistic system,

we need to inform educators so that they can teach early reading in Singapore effectively using phonics. Low (2014) conducted a meta-analysis on research conducted on English in Singapore, and found that Linguistic Features, covering areas such as phonology, grammar and contact features, is one of the most researched topics locally and have garnered attention from researchers in the past decades. However, she also noted that there have been few studies on how to teach linguistic features. Jones & Deterding (2007), a book on phonics and reading of English for teachers in Southeast Asia, can be considered one of the limited resources available that infused some linguistic features within their discussion. However, Jones & Deterding (2007) had approached phonics from an arguably balanced literacy approach, which is in stark contrast with evidence provided so far in the thesis.

Third, development of data-driven reading materials designed to target Singapore children would be essential to a beginning reading programme, in addition to the considerations in designing scope and sequence for grapheme-phoneme correspondences. These reading materials would need to factor in topics of interest and relevance, and choice of vocabulary words. Reading content that is authentic, relevant, and reflects language usage beyond the classroom can help children to make links between the reading materials and language they have already been exposed to (Mukundan, Zarifi, & Kalajahi, 2016). For example, the term 'hawker centre', is frequently used in the Singapore context to refer to a food complex with many stalls selling local affordable food, could be included as a vocabulary word on the topic of food.

Fourth, the use of local developed materials needs more research in order to investigate whether such materials improve reading outcomes. High quality intervention research, as described in Chapter 2, must be incorporated in order for results to be interpretable in meaningful ways. Such research provides additional opportunities for educational researchers to collaborate with schools and educators.

Recommendations for educators

First, the education faculty collectively needs to recognise that there are substantial gaps in teachers' knowledge about the science of reading and how to support students using systematic and explicit phonics. This applies to many countries, such as the US and of course, Singapore. Teacher professional training is critical especially for teachers working with early readers at primary levels, as well as those supporting students with reading difficulties. It has long been recognised that teacher education does not sufficiently equip teachers with necessary knowledge about how to teach reading (Mukundan, Zarifi, & Kalajahi, 2016). In order for teachers to effectively teach reading, they would need to first understand topics within language and literacy, such as linguistic knowledge about how children acquire language, reading development and instruction. An educational reform in Singapore informed by science of reading would be necessary to overcome possible resistance so that beginning readers can all benefit from systematic explicit phonics.

Second, with consideration of the linguistic context of Singapore English, educators would need to understand how pronunciations of vowels and consonants in Singapore can affect instruction on teaching reading, and why

programmes designed for other varieties of English might be poorly suited for local use. For example, the word “bed” is usually pronounced [bɜd] in Singapore and not [bed]. Hence, this word is not suitable to be included when teaching children to blend sounds with vowel sound /e/ as the resulting decoded word would not resemble how a child in Singapore would pronounce the word. Such professional development might require discussions about Singapore English, how it differs from other Standard varieties of English, professional development resources, such as courses and textbooks to be developed. This suggests that close collaborations between education researchers and educators would be necessary for such long-term endeavours.

Given the importance of the local language landscape in phonics-based instruction, Chapter 4 will attempt to bridge the gap between what we know about Singapore English and phonics interventions.

Chapter 4: Challenges in designing a phonics scope and sequence appropriate for teaching English in Singapore

Chapter 1 introduced how explicit and systematic instruction in teaching grapheme-phoneme correspondence is critical in early reading instruction. Chapter 2 presented a systematic review of small-group reading interventions for struggling readers, and found there is more evidence for phonics-based interventions, suggesting that future reading interventions in Singapore for children with reading difficulties should be conducted using a phonics-based approach. At the same time, we found no bilingual intervention studies with sufficient quality, meaning that the application of phonics to the Singapore context may need to be treated with caution. In Chapter 3, we showed that skilled readers' intuitions about the vowels in written words suggest that Singapore English has a different vowel system compared to other Standard varieties of English such as the US and UK, and discussed how this may have implications for reading instruction in Singapore. National curriculum materials in Singapore suggest that phonics is included in a balanced literacy approach, but do not provide any guidance on how to achieve this in the local English language context. Several challenges therefore remain for the development of phonics-based reading materials that would be suitable for Singapore English. This chapter concludes by addressing these challenges, and discussing steps that would be required to provide a robust, evidence-based approach to phonics materials for Singapore.

Implications of the Singapore English context on the use of phonics

Many researchers have noted the discrepancy in the vowel system of Singapore compared to other varieties of English. However, in their work aimed at educators, Jones & Deterding (2007) take the view that a reduced inventory for vowels in Singapore or Malaysia would not be advisable and maintain the perspective that all 44 phonemes in British English should be taught. In line with this view, they propose that reading instruction should highlight differences between pairs of vowels that may well have merged. For example, while they acknowledge that students and teachers might not distinguish between /æ/ and /e/ in their speech, or in their hearing, they recommend that teachers teach the sounds “explicitly” and explain the distinction between these sounds. They suggest using a simple rule for teachers and students to tell them apart, that the presence of ‘a’ in spelling followed by one or more consonants mean that it is /æ/. From the author’s own experience, the way educators in Singapore achieve a distinction between the sounds is to artificially lengthen the tense vowels in tense-lax pairs, and to create a wholly novel pronunciation for the / æ / vowel: a long /ε/ sound. This means that pronunciation of English words in the reading context is different from the way that words are typically pronounced in Singaporean homes, and is also different from the ‘target’ varieties of English. It should be noted that teaching new pronunciations at the same time as teaching children to read is not aligned with the core principles of phonics instruction.

This perspective of following the standard British English linguistic system is consistent with many phonics programmes available in Singapore. For example, most phonics instruction programmes in Singapore are sold by

international publishing companies and have been developed in the US or UK. Even programmes that are adapted or developed locally, have not assessed how vowel sounds of Singapore affect grapheme-phoneme correspondences. Educators are not explicitly trained on how the vowels operate differently in Singapore, compared to pronunciations that can be found in dictionaries of UK or US English.

Within the first year of life, a child's auditory perception starts to tune in to speech sounds in languages that the child is exposed to in one's environment (Kuhl, 2011). For a child who cannot hear the difference between two sounds due to the fact that the environment simply does not discriminate between the two phonetic contrasts, it would be confusing and possibly frustrating to learn otherwise. Jones & Deterding (2007)'s approach to highlight the differences between pairs of vowels that might have merged (for e.g., /æ/ and /e/), suggests teaching children new pronunciations at the same time as teaching spelling. Applying the concept of attentional load theory from cognitive science of learning suggests that this is not an optimal way to learn, as we are typically only able to selectively focus on one message at a time (Weinstein, Sumeracki, & Caviglioli, 2019). By contrast, teaching novel spellings and novel pronunciations at the same time requires the child to learn two difficult concepts simultaneously. In addition, learning might be further reduced due to the fact that the sounds may perceptually sound the same to the child, even if the educator can pronounce /æ/ and /e/ accurately.

Linguistic differences in pronunciation impact how grapheme-phoneme mappings are formed. As discussed in Chapter 1, oral language skills form the

foundation for learning to read. Learning to read from a perspective of a non-reader is a different process from how a skilled adult reader would approach reading. A beginning reader is more likely to decode a printed string of letters successfully if the resulting sequence of sounds corresponds to a word that already exists in their mental lexicon (Weinstein, Sumeracki, & Caviglioli, 2019). Each word is represented and stored as a structured sequence of phonemes. Following the above reasoning, if the phonological categories encoded by a written word do not correspond to the representation in a child's mental lexicon, learning might not be effective. International literature has suggested that the differences between a child's dialect and English instruction in the classroom may present a source of confusion. Brown and colleagues (2015) found that the task of learning to read can be more challenging when African American children were exposed to alternative pronunciations in the classroom. Specifically, decoding in African American children was affected by their knowledge of two dialects for words that are pronounced differently in Mainstream American and African American English (AAE). The learning experiences and challenges faced by children in Singapore are potentially similar, where home and school dialects are likely to be different. For example, imagine a child has heard all their life that a 'bat' is called a [bet] (a common Singaporean pronunciation), rather than a [bæt] (the British English pronunciation), then it is possible to confuse the child when he/she is exposed to instruction and is taught that this winged mammal is [bæt] and spelt with letter A and not [bet] spelt with letter E even though they sound the same to the child (assuming the educator was able to pronounce [bæt] accurately if the educator is a speaker of Singapore English).

The other alternative is to localise teaching materials so that they match the phonetic inventory, by developing a truly Singaporean way of learning about how sounds go with letters. In this section, we consider what would be necessary to develop such a ‘scope and sequence’ for early readers in Singapore.

Developing a scope and sequence for phonics instruction

Researchers have proposed evidence-based guidelines for the content and instructional plan of phonics instruction, also known as ‘scope and sequence’ for phoneme-grapheme correspondence (Chard & Osborn, 1999; Earle & Sayeski, 2017). To explore challenges in developing a ‘scope and sequence’ in Singapore, application of guidelines developed by these authors will be discussed. One of the key guidelines for such a scope and sequence would be that frequent grapheme-phoneme correspondences should be taught directly and explicitly (See Chapter 1 for discussion on explicit systematic phonics). The first challenge for establishing frequent grapheme-phoneme correspondences in Singapore English is that there is no agreed, standard inventory for vowels and consonants in Singapore English, even though many researchers have recognised some possible mergers (See Chapter 3 for possible vowel mergers). Without an agreed inventory, it is not possible to teach common grapheme-phoneme correspondences in a direct and explicit manner. Future research would need to look into the development of an agreed phoneme inventory for Singapore English, as well as dictionaries with Singaporean pronunciations listed.

One guideline for the order in which grapheme-phoneme correspondences should be taught (the sequence) is to teach tense vowel sounds

before lax vowel sounds (Chard & Osborn, 1999; Earle & Sayeski, 2017). This presents a challenge as the status of the tense/lax distinction in Singapore English is questionable. As reported in Chapter 3, some people reported minimal pairs with tense/lax distinctions as ‘different’ 100% of the time while some people reported minimal pairs with tense/lax distinctions as ‘same’ 100% of the time. There was therefore large variability in people’s intuitions about the tense/lax distinction. Moreover, we cannot assume that what we have found about adult English speakers in Singapore would apply to children who are starting to learn reading. Just as how skilled adult reading differs from children learning to read, we need to find out more about the developmental trajectory of Singapore children’s vowel space, and when they start to acquire the tense/lax distinction.

The lack of a tense/lax distinction may introduce further complexities for children learning to read multiple graphemes mapping to the same phoneme. Though this is speculative at the moment, we think it may be possible to teach tense vowel sounds first along with the graphemes, then introduce lax spellings later, just as is the case with alternative spellings like the high-frequency ‘F’ correspondence is taught before the alternative ‘GH’ spelling in ‘enough’. In order to determine whether this guideline applies in Singapore, we need clarity regarding the status of the tense/lax distinctions in both Singapore adults and children. Following which, intervention research can be conducted to determine whether such sequencing make a difference in how children learn the mappings.

Another guideline for developing a scope and sequence is to teach a number of high-utility grapheme-phoneme correspondences first then add low-

utility grapheme-phoneme correspondences later (Earle & Sayeski, 2017). High-utility correspondences are those that are encountered frequently in simple words that would be known to children. An ideal corpus for reference would be children-centric, with a large collection of both local and overseas children's literature that reflects children's language exposure and usage in Singapore, along with children's original writing. This allows for investigations into words that children encounter frequently in literature, and types of context. There is no corpus of Singaporean children's English that could be used to establish a) word frequencies, and b) pronunciations of words in early vocabulary used in Singapore English, which allows for documentation of the frequency of grapheme-phoneme correspondences. Apart from not having a corpus of Singaporean children's English, we do not even have a suitable adult corpus that would provide us documentation of the frequency of grapheme-phoneme correspondences in typical spoken English. Table 4.1 shows a list of current databases of Singapore English. As systematic analysis of the mapping of graphemes to phonemes in Singapore would require phonological transcriptions of the recordings, and such data are not currently available for existing databases. In order to follow the guideline of 'high-utility' correspondences, we would need a large corpus of spoken Singaporean English along with transcriptions or a pronouncing dictionary, so that the frequency of grapheme-to-phoneme mapping can be established.

Table 4.1. List of databases currently on Singapore English.

Databases	Language	Description	Date	Phon. Trans (1)	Text Trans (2)
National Speech Corpus	English	2,000 hours of locally accented audio and corresponding text transcriptions	n.d.	N	Y
NIE Spoken Corpus of English in Asia	English	Audio and corresponding text transcriptions of 10 speakers of Singapore English reading short passages, single words, and a 5-minute interview	2015	N	Y
Singapore Corpus of Research in Education	English, Mandarin, Malay and Tamil	Audio/ video recording of classroom interactions in various subjects in Singapore schools; along with transcripts	n.d.	N	Y
International Corpus of English - Singapore	English	Spoken and written components, with 500 texts (total of one million words)	2011	N	Y
Singapore Bilingual Corpus	English, Mandarin	Audio recording of 55 children in 2 childcare centres in Singapore speaking a mixture of English and Mandarin	2013	N	Y

(1) Phonological transcription is not available for all the databases.

(2) Text transcription is available for all the databases.

Another guideline for developing a scope and sequence is to teach a single mapping for each grapheme before introducing alternate grapheme-phoneme correspondences (e.g., ‘c’ = /k/ as in “cat”, “cook” “cup”, before teaching ‘c’ = /s/ as in “city,” “circus”). Essentially, this guideline is about teaching one-to-one correspondences, which are considered simple, before proceeding to teach complex many-to-one correspondences. Without a corpus of Singaporean children’s English, it is hard to establish individual grapheme-phoneme correspondences, and the relative frequency of which correspondence is more frequent. While this is likely to be similar for most consonants, it could be problematic due to possible vowel mergers (See Chapter 2). For example, if /æ/

and /e/ have merged, then ‘met’ and ‘mat’ would be homophones, making it hard to decide when to introduce this many-to-one correspondence.

Due to the lack of a suitable corpus in Singapore for grapheme-phoneme analysis, information from (Earle & Sayeski, 2017) was extracted to present grapheme-phoneme correspondences for the /æ/ and /e/ vowels suspected of merger in Table 4.2 (See Appendix 10 for information on other vowel pairs suspected of merger). The table shows that /æ/ and /e/ are represented by distinct spellings in other varieties of English (e.g., UK, US): /æ/ is most frequently spelt with 'a' while /e/ is often spelt as 'e' or 'ea'. For a speaker of Singapore English who might not distinguish between /æ/ and /e/, misspelling might be more common as 'a', 'e' and 'ea' become possible spellings for a word. In order to establish which order to introduce the letter ‘a’ and the letter ‘e’, we would need to have a corpus of spoken Singapore English along with a pronouncing dictionary for Singapore English, or phonological transcriptions of the speech, so that the frequency of the different mappings of graphemes to phonemes can be established.

Table 4.2. Phoneme/grapheme correspondences for /æ/ and /e/

Grapheme	/æ/	/e/
a	4192	-
a-e	147	-
e	-	3316
ea	-	139
e-e	-	79

The numbers from the table were extracted from Hanna et al. 1966, in which they analysed phoneme/grapheme correspondences in a large US corpus. The numbers reflect the number of words that were found to match the respective grapheme-phoneme categories.

Another guideline for developing a scope and sequence is to introduce consonants and vowels in a sequence that permits children to learn how to read, starting with letters that allow children to sound out slowly, and blend while reading. For example, in a possible scope and sequence proposed in Earle & Sayeski (2017), the first set of consonants and short vowel sounds to be introduced would be “*m, s, a, t, r, i, f*” followed by “*n, d, o, h, l, g*”. Most of these graphemes are unambiguous in simple, short words in Singapore. However, while the letter ‘a’ unambiguously represents the phoneme /æ/ in short words such as ‘cat’ in British English, the letter ‘a’ may be problematic in Singapore for two reasons: /æ/ may be merged with /e/, making the phoneme ambiguous (as described above); The letter ‘a’ unambiguously represents the sound /ʌ/ in Malay and Hanyu Pinyin – two of the prevalent writing systems Singaporean children are familiar with – potentially making it hard or confusing for children to remember which sound they should be applying to an unfamiliar word when learning to read in English.

Besides requiring a corpus and pronouncing dictionary (or phonological transcriptions), further research is needed to determine whether possible interference from children’s other languages might present additional challenges for children learning to read English in Singapore.

Lastly, another important consideration for developing a scope and sequence is to provide ample opportunities for practice using simple VC and CVC words after phoneme-grapheme correspondences have been mastered (V refers to vowel, C refers to consonant). Most sources agree that the voiceless word-final stop consonant (such as /t/ in “put”) is optional in Singapore English, and often

replaced with a glottal stop, or deleted (Deterding, 2007; Deterding & Hvitfeldt, 1994). This also applies to syllable-final consonant clusters (Deterding, 2007; Deterding & Hvitfeldt, 1994). A lot of simple VC and CVC words in English end in these sounds. It is unclear if children's representations of these words contain a silent /t/, or no /t/ at all. One option for developing a scope and sequence for Singapore is to avoid these words in early reading instruction to avoid ambiguity. However, the number of VC and CVC words will be greatly reduced if we avoid words ending with /t/, /k/ and /p/ for example. Simple words such as "mat, rat, fat, sat, fit, it" are all words that early readers would encounter first in a typical programme, leaving few remaining words for use in early reading materials. Besides requiring a corpus along with a pronouncing dictionary or transcriptions, we would need research into children's representation of word-final stop consonants, in order to figure out whether delaying the teaching of words ending with stop consonants would be the best approach for instruction.

Taken together, Table 4.3 summarises information that we need to know about Singapore English before a meaningful and relevant scope and sequence can be developed for phonics instruction in Singapore.

Table 4.3. What do we need to know about Singapore English

Guidelines for developing scope and sequence for phonics instruction (1)	What is needed?
Explicit and direct phonics instruction	An agreed phoneme inventory
Tense/ lax vowel order	Clarity on tense lax/distinctions in adults and children; Instructions for teachers on how to explain spellings of merged vowels to children
Teach high-frequency correspondences first	A representative Singapore English corpus; Pronouncing dictionary OR transcriptions
One to one before many-to-one.	A representative Singapore English corpus; Pronouncing dictionary OR transcriptions
Start with letters that allow children to sound out slowly, and blend while reading.	A representative Singapore English corpus; Pronouncing dictionary OR transcriptions; Research to find out if sound/letter differences between scripts cause confusion.
Practice using VC and CVC words.	A representative Singapore English corpus; Pronouncing dictionary OR transcriptions; Research into children's representation of word-final stop consonant

(1) Guidelines adapted from (Low & Brown, 2005); Earle & Sayeski (2017)

Recommendations for researchers

We clearly need more research into Singapore English in order to inform us about effective phonics instruction that takes characteristics of Singapore English into consideration. One of the top priorities, judging from its frequent mention, would be the development of a corpus Singapore English that would allow for mapping of graphemes to phonemes. This could come in the form of either a pronouncing dictionary or phonological transcriptions. Ideally, the corpus would comprise of substantial data collected from both adults and children.

Recommendations for educators

Although there is a relatively long list of research necessary before phonics instruction can be customised to the linguistic context of Singapore English, it is our belief that phonics-based instruction would be relevant and

important for children in Singapore learning to read. This approach needs to be tested through local intervention studies, by comparing phonics instruction customised to Singapore English, against phonics instruction designed for other varieties of English.

In this chapter, the discussion has centered on how a scope and sequence developed to suit the linguistic context of Singapore is needed for phonics instruction. However, educators need to bear in mind that the development of such a scope and sequence based on research would just be a start for more meaningful work in the development of a full-fledged intervention programme. For example, decodable texts are reading materials that allow beginning readers to practice grapheme-phoneme correspondences that has been taught (Castles et al., 2018). This provides reading practice in a manner that gradually builds up word exposure based on the number of correspondence already learnt. Ideally, decodable texts follow the scope and sequence closely. Hence, such resources should be locally developed, and use words that are of high frequency in the local corpus.

Importance of collaborations between researchers and educators cannot be emphasized enough. For example, building a corpus of children's speech, literature and original writing would take a huge amount of data collection. One such ongoing project would be the Oxford Corpus, which is a joint project by researchers from Oxford University (ReadOxford lab) and Oxford University Press. The Oxford Corpus benefits Oxford University Press through the development of children's dictionaries and learning resources, while ReadOxford lab uses the corpus for research, e.g. how children learn new words (Read

Oxford, n.d.). Similarly, educators and researchers can work together in a mutually beneficial manner. Educators have access to huge amount of children's books and original writing, which can be tapped upon to build a local corpus. While the emphasis for discussion in the paper is on phonics instruction, the resulting corpus can be used for wide educational purposes such as curriculum planning according to children's developmental trajectory in vocabulary and even grammar.

We have highlighted that the poor suitability of using materials developed for Inner circle English in Singapore as suitability depends on the similarities between linguistic contexts, and whether simple adaptations can be done to factor in differences in pronunciation between different linguistics contexts. We have also explored the challenges in designing a scope and sequence suitable for use in Singapore, taking phoneme-grapheme mapping and available information about Singapore English into considerations.

Conclusions

We have shown through our systematic review, there is more evidence for group-based phonics interventions for struggling readers, than there is for group-based interventions without phonics, and that there are no bilingual studies with sufficient quality to provide evidence of intervention effectiveness. We also found that English-speaking adults in Singapore respond differently to vowel pairs suspected of merger compared to vowel pairs not suspected of merger. There is large variability in judgments across all vowel pairs suspected of merger. Presently, in Singapore, there is no agreed phoneme inventory that

could be used to guide instruction on how to teach English using systematic phonics nor is there corpus with phonological transcription that would allow for research into grapheme-phoneme correspondences. Hence, further research will be necessary before a truly local approach to phonics instruction can be developed. We have detailed various complications arising from the suitability of materials developed for other varieties of English in Singapore. We then propose designing a scope and sequence suited to the Singapore linguistic context and explored challenges we need to overcome, from the perspectives of both researchers and educators.

As highlighted in Chapter 1, literacy is critical. There are grave and long-lasting consequences of not being able to read. Children who fall behind in their literacy will lose out as other peers learn increasingly more information through reading. Such widening gaps become increasingly apparent when children get older and content knowledge gets more complex and require increasing amount of reading to get access to information. The use of instructional practices validated by research can ensure better rates of literacy success for more children. As discussed throughout the thesis, the use of phonics instruction to teach beginning readers is critical in ensuring that they get the best opportunity at learning to read, through a pre-determined set of phoneme-grapheme correspondences. Systematic explicit phonics instruction thus provides beginning readers with an essential foundation to subsequently master other critical components of literacy, such as reading comprehension.

References

- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., ... Treiman, R. (2007). The english lexicon project. *Behavior Research Methods*, 39(3), 445–459. <https://doi.org/10.3758/BF03193014>
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1). <https://doi.org/10.18637/jss.v067.i01>
- Bishop, D. (2008). Criteria for evaluating behavioural interventions for neurodevelopmental disorders. *Journal of Paediatrics and Child Health*, 44, 520–530. <https://doi.org/10.1111/j.1440-1754.2008.01358-01359>
- Bishop, D. V. M. (2013). Research Review: Emanuel Miller Memorial Lecture 2012 - Neuroscientific studies of intervention for language impairment in children: Interpretive and methodological problems. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 54(3), 247–259. <https://doi.org/10.1111/jcpp.12034>
- Brown, M. C., Sibley, D. E., Washington, J. A., Rogers, T. T., Edwards, J. R., Macdonald, M. C., & Seidenberg, M. S. (2015). Impact of dialect use on a basic component of learning to read. *Frontiers in Psychology*, 6(196). <https://doi.org/10.3389/fpsyg.2015.00196>
- Buckingham, J., Wheldall, K., & Beaman-Wheldall, R. (2013). Why poor children are more likely to become poor readers: The school years. *Australian Journal of Education*, 57(3), 190–213. <https://doi.org/10.1177/0004944113495500>
- Byrne, B. (2005). Theories of learning to read. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 104–119). Malden, MA: Blackwell.
- Castles, A., Rastle, K., & Nation, K. (2018). Ending the Reading Wars: Reading Acquisition From Novice to Expert. *Psychological Science in the Public Interest*, 19(1), 5–51. <https://doi.org/10.1177/1529100618772271>
- Chard, D. J., & Osborn, J. (1999). Phonics and word recognition instruction in early reading programs: Guidelines for accessibility. *Learning Disabilities Research & Practice*, 14(2), 107–117.
- Clarke, P. J., Snowling, M. J., Truelove, E., & Hulme, C. (2010). Ameliorating children's reading-comprehension difficulties: A randomized controlled trial. *Psychological Science*, 21(8), 1106–1116. <https://doi.org/10.1177/0956797610375449>
- de Graaff, S., Bosman, A. M. T., Hasselman, F., & Verhoeven, L. (2009). Benefits of systematic phonics instruction. *Scientific Studies of Reading*, 13(4), 318–333. <https://doi.org/10.1080/10888430903001308>
- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, 47(1), 1–12. <https://doi.org/10.3758/s13428-014-0458-y>
- Dehaene, S. (2011). The massive impact of literacy on the brain and its consequences for education. *Human Neuroplasticity and Education*, (October 2010), 19–32. Retrieved from <http://www.pas.va/content/dam/accademia/pdf/sv117.pdf#page=17>
- Denton, C. A., Fletcher, J. M., Taylor, W. P., Barth, A. E., & Vaughn, S. (2014). An Experimental Evaluation of Guided Reading and Explicit Interventions for

- Primary-Grade Students At-Risk for Reading Difficulties. *Journal of Research on Educational Effectiveness*, 7(3), 268–293.
<https://doi.org/10.1080/19345747.2014.906010>
- Deterding, D. (2003). An instrumental study of the monophthong vowels of Singapore English. *English World-Wide*, 24(1), 1–16. <https://doi.org/10.1075/eww.24.1.02det>
- Deterding, D. (2005). Emergent patterns in the vowels of Singapore English. *English World-Wide*, 26(2), 179–197.
<https://doi.org/10.1075/eww.26.2.04det>
- Deterding, D. (2007). *Singapore English*. Edinburgh University Press.
- Deterding, D., & Hvitfeldt, R. (1994). The Features of Singapore English Pronunciation : Implications for Teachers. *English*, 15(1), 98–107.
- Deterding, D., & Poedjosoedarmo, G. (2000). To what extent can the ethnic group of young Singaporeans be identified from their speech? *The English Language in Singapore: Research on Pronunciation*, (eds), 1–9.
- Duff, F. J., & Clarke, P. J. (2011). Practitioner Review: Reading disorders: What are the effective interventions and how should they be implemented and evaluated? *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 52(1), 3–12. <https://doi.org/10.1111/j.1469-7610.2010.02310.x>
- Duff, F. J., Hayiou-Thomas, M. E., & Hulme, C. (2012). Evaluating the effectiveness of a phonologically based reading intervention for struggling readers with varying language profiles. *Reading and Writing*, 25(3), 621–640.
<https://doi.org/10.1007/s11145-010-9291-6>
- Duff, F. J., Reen, G., Plunkett, K., & Nation, K. (2015). Do infant vocabulary skills predict school-age language and literacy outcomes? *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 56(8), 848–856.
<https://doi.org/10.1111/jcpp.12378>
- Earle, G. A., & Sayeski, K. L. (2017). Systematic Instruction in Phoneme-Grapheme Correspondence for Students With Reading Disabilities. *Intervention in School and Clinic*, 52(5), 262–269.
<https://doi.org/10.1177/1053451216676798>
- Ehri, L. C., Nunes, S. R., Stahl, S. A., Willows, D. M., Ehri, L. C., Nunes, S. R., & Stahl, S. A. (2001). Systematic Phonics Instruction Helps Students Learn to Read : Evidence from the National Reading Panel ' s Meta-Analysis Published by : American Educational Research Association Stable URL :
<http://www.jstor.org/stable/3516004> Systematic Phonics Instructi. *Review of Educational Research*, 71(3), 393–447.
- Feldman, P., Cosman, B. C., Feldman, N., & Quiros, A. (n.d.). List of English Homophones. Retrieved February 11, 2019, from
<http://phillipmfeldman.org/English/Homophones.html>
- Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly*, 41(1), 93–99.
<https://doi.org/10.1598/rrq.41.1.4>
- Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality Indicators for Group Experimental and Quasi-Experimental Research in Special Education. *Exceptional Children*, 71(2), 149–164.
<https://doi.org/10.1177/001440290507100202>
- Gordon, M. J. (2004). Investigating Chain Shifts and Mergers. In J. K. Chambers, P.

- Trudgill, & N. Schilling-Estes (Eds.), *The handbook of language variation and change* (pp. 244–266). Blackwell Publishing.
- Gough, P., & Tunmer, W. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, 7(1), 6–10.
- Hagan-Burke, S., Coyne, M. D., Kwok, O. man, Simmons, D. C., Kim, M., Simmons, L. E., ... McSparran Ruby, M. (2013). The Effects and Interactions of Student, Teacher, and Setting Variables on Reading Outcomes for Kindergarteners Receiving Supplemental Reading Intervention. *Journal of Learning Disabilities*, 46(3), 260–277. <https://doi.org/10.1177/0022219411420571>
- Hagan-Burke, S., Kwok, O., Zou, Y., Johnson, C., Simmons, D., & Coyne, M. D. (2011). An examination of problem behaviors and reading outcomes in kindergarten students. *The Journal of Special Education*, 45(3), 131–148. <https://doi.org/10.1177/0022466909359425>
- Hanford, E. (2018). Hard Words: Why aren't kids being taught to read? Retrieved August 8, 2019, from <https://www.apmreports.org/story/2018/09/10/hard-words-why-american-kids-arent-being-taught-to-read>
- Hanna, P. R., Hanna, J. S., Hodges, R. E., & Rudorf, E. H. (1966). *Phoneme-grapheme correspondences as cues to spelling improvement*. Washington, DC.
- Heuven, W. J. B. Van, Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, 67(6), 1176–1190. <https://doi.org/10.1080/17470218.2013.850521>
- Higgins, John. (n.d.). Minimal pairs for English RP. Retrieved February 11, 2019, from <https://minimal-pairs.000webhostapp.com/minimal.html>
- Higgins, JPT, & Green, S. (Eds.). (2011). *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0*. The Cochrane Collaboration. Retrieved from Available from www.handbook.cochrane.org.
- Hulme, C., & Snowling, M. J. (2016). Reading disorders and dyslexia. *Current Opinion in Pediatrics*, 28(6), 731–735. <https://doi.org/10.1097/MOP.0000000000000411>
- Hwee, N. C. K., & Houghton, S. (2011). The effectiveness of Orton-Gillingham-based instruction with Singaporean children with specific reading disability (dyslexia). *British Journal of Special Education*, 38(3), 143–149. <https://doi.org/10.1111/j.1467-8578.2011.00510.x>
- Jones, S. A., & Deterding, D. (2007). *Phonics and Beginning Reading: A practical guide for teachers in Southeast Asia*. Singapore: McGraw-Hill.
- 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(2), 101–114. <https://doi.org/10.1177/0022219407313412>
- Kuhl, P. K. (2011). Early Language Learning and Literacy: Neuroscience Implications for Education. *Mind Brain Educ*, 5(3), 128–142. <https://doi.org/10.1111/j.1751-228X.2011.01121.x>
- Kuperman, V., Drieghe, D., Keuleers, E., & Brysbaert, M. (2013). How strongly do word reading times and lexical decision times correlate? Combining data from eye movement corpora and megastudies, 66(3), 563–580.

- <https://doi.org/10.1080/17470218.2012.658820>
- Kyle, F., Kujala, J., Richardson, U., Lyytinen, H., & Goswami, U. (2013). Assessing the effectiveness of two theoretically motivated computer-assisted reading interventions in the United Kingdom: GG Rime and GG Phoneme. *Reading Research Quarterly*, 48(1), 61–76. <https://doi.org/10.1002/rrq.038>
- Lee, S. H., & Scanlon, D. M. (2015). The effects of the Interactive Strategies Approach on at-risk kindergartners' spelling. *Reading and Writing*, 28(3), 313–346. <https://doi.org/10.1007/s11145-014-9526-z>
- Leung, W. T., & Styles, S. J. (2018). *Reading Intervention Quality: Preregistration Docs*. DR-NTU (Data), V1. <https://doi.org/https://doi.org/10.21979/N9/RAQM8H>
- Leung, W. T., & Styles, S. J. (2019). *Preregistration Documents: What do words sound like in Singapore?* DR-NTU (Data), V1. Retrieved from <https://doi.org/10.21979/N9/UAZ7AS>
- Low, E. (2012). Singapore English. In E. Low & A. Hashim (Eds.), *English in Southeast Asia: Features, policy and language in use* (pp. 35–54). Amsterdam: John Benjamins Publishing Company.
- Low, E. (2014). Research on English in Singapore. *World Englishes*, 33(4), 439–457. <https://doi.org/10.1111/weng.12106>
- Low, E., & Brown, A. (2005). *English in Singapore: An Introduction*. Singapore: McGraw Hill.
- Lysenko, L. V., Abrami, P. C., Bernard, R. M., Dagenais, C., & Janosz, M. (2014). Educational Research in Educational Practice: Predictors of Use. *Canadian Journal of Education*, 37(2), 1–26.
- Ministry of Education. (2010). *English Language Syllabus 2010 Primary & Secondary (Express/ Normal [Academic])*. Retrieved from <https://www.moe.gov.sg/docs/default-source/document/education/syllabuses/english-language-and-literature/files/english-primary-secondary-express-normal-academic.pdf>
- Ministry of Education. (2014). Sool-based Dyslexia Remediation (SDR) Programme. Retrieved July 17, 2019, from [https://www.schoolbag.sg/story/school-based-dyslexia-remediation-\(sdr\)-programme](https://www.schoolbag.sg/story/school-based-dyslexia-remediation-(sdr)-programme)
- Ministry of Education. (2018). Support programmes available for small groups of students with dyslexia. Retrieved July 17, 2019, from <https://www.moe.gov.sg/news/forum-letter-replies/support-programmes-available-for-small-groups-of-students-with-dyslexia>
- Moats, L. (2007). Whole-Language High Jinks. *Thomas B. Fordham Institute*, 1–33. Retrieved from <http://www.k12.wa.us/Reading/ReadingFirst/pubdocs/Moats2007.pdf>
- Moats, Louisa. (2014). What teachers don't know and why they aren't learning it: addressing the need for content and pedagogy in teacher education. *Australian Journal of Learning Difficulties*, 19(2), 75–91. <https://doi.org/10.1080/19404158.2014.941093>
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Journal of Clinical Epidemiology*, 62(10), 1006–1012. <https://doi.org/10.1016/j.jclinepi.2009.06.005>

- Moir, T. (2018). Why Is Implementation Science Important for Intervention Design and Evaluation Within Educational Settings? *Frontiers in Education*, 3(July), 1–9. <https://doi.org/10.3389/feduc.2018.00061>
- Morris, R. D., Lovett, M. W., Wolf, M., Sevcik, R. A., Steinbach, K. A., Frijters, J. C., & Shapiro, M. B. (2012). Multiple-component remediation for developmental reading disabilities: Iq, socioeconomic status, and race as factors in remedial outcome. *Journal of Learning Disabilities*, 45(2), 99–127. <https://doi.org/10.1177/0022219409355472>
- Mukundan, J., Zarifi, A., & Kalajahi, S. A. R. (2016). Developing Reading Materials for ESL Learners. *Issues in Materials Development*, 65–73. https://doi.org/10.1007/978-94-6300-432-9_6
- NAEP. (2019). NAEP Reading 2019 Highlights. Retrieved January 6, 2020, from <https://www.nationsreportcard.gov/highlights/reading/2019/>
- Nation, K. (2019). Children's Reading Difficulties, Language, and Reflections on the Simple View of Reading. *Australian Journal of Learning Difficulties*, 1–50. <https://doi.org/10.1080/19404158.2019.1609272>
- National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. National Institute of Child Health & Development.
- O'Connor, R. E., Bocian, K. M., Sanchez, V., & Beach, K. D. (2014). Access to a Responsiveness to Intervention Model: Does Beginning Intervention in Kindergarten Matter? *Journal of Learning Disabilities*, 47(4), 307–328. <https://doi.org/10.1177/0022219412459354>
- OASIS. (n.d.). Open Accessible Summaries In Language Studies. Retrieved August 9, 2019, from <https://oasis-database.org>
- Pakir, A. (2001). Bilingual education with English as an official language: Sociocultural implications. In J. E. Alatis & A.-H. Tan (Eds.), *Language in Our Time: Bilingual Education and Official English, Ebonics and Standard English, Immigration and the Unz Initiative* (pp. 341–349). Washington, D.C.: Georgetown University Press.
- Pang, E. S., Lim, F. V., Choe, K. C., Peters, C., & Chua, L. C. (2015). System Scaling in Singapore: The STELLAR Story. In W. O. Lee, D. W. L. Hung, L. W. Teh, & C.-K. Looi (Eds.), *Scaling Educational Innovations* (pp. 105–122). Springer Singapore. <https://doi.org/10.1007/978-981-287-537-2>
- Poedjosoedarmo, G. (2000). A description of the English pronunciation of young educated Singaporeans: a study in multidimensional variation. In Adam Brown, D. Deterding, & E. L. Low (Eds.), *The English Language in Singapore: Research on Pronunciation* (pp. 66–75).
- Qualtrics. (2005). *Qualtrics*. Provo, Utah.
- R Core Team. (2018). R: A Language and Environment for Statistical Computing. Retrieved from <https://www.r-project.org>
- Rastle, K. (2019a). EPS mid-career prize lecture 2017: Writing systems, reading, and language. *Quarterly Journal of Experimental Psychology*, 72(4), 677–692. <https://doi.org/10.1177/1747021819829696>
- Rastle, K. (2019b). How Research in Psychology Can Help Us End the “Reading Wars.” In *Talk given at an International Meeting for Cognitive Science in the Classroom at UNESCO, Paris*.
- Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., Seidenberg, M. S., Rayner,

- K., ... Seidenberg, M. S. (2001). How Psychological Science Informs the Teaching of Reading. *Psychological Science in the Public Interest*, 2, 31–74. <https://doi.org/10.1111/1529-1006.00004>
- Read Oxford. (n.d.). The Oxford Corpus. Retrieved August 14, 2019, from <https://readoxford.org/our-research/leverhulme>
- Roehrig, A. D., Soper, D., Cox, B. E., & Colvin, G. P. (2018). Changing the Default to Support Open Access to Education Research. *Educational Researcher*, 47(7), 465–473. <https://doi.org/10.3102/0013189x18782974>
- Rose, J. (2006). *Independent review of the teaching of early reading final report*. U.K. Department for Education and Skills.
- Rowe, K. (2005). *Teaching reading: National inquiry into the teaching of literacy*. Department of Education, Science and Training, Australian Council for Educational Research.
- Seidenberg, M. (2017). *Language at the Speed of Sight: How we Read, Why so Many Can't, and what can be done about it*. Basic Books.
- Sim, C. (2016). Bilingual policy. Retrieved August 12, 2019, from http://eresources.nlb.gov.sg/infopedia/articles/SIP_2016-09-01_093402.html
- Simmons, D. C., Taylor, A. B., Oslund, E. L., Simmons, L. E., Coyne, M. D., Little, M. E., ... Kim, M. (2014). Predictors of at-risk kindergarteners' later reading difficulty: Examining learner-by-intervention interactions. *Reading and Writing*, 27(3), 451–479. <https://doi.org/10.1007/s11145-013-9452-5>
- Snow, C., Griffin, P., & Burns, M. S. (Eds.). (2007). *Knowledge to support the teaching of reading: Preparing teachers for a changing world*. John Wiley & Sons.
- Snow, P. C. (2019). Dear Parents: Welcome to the Confusing World of Reading Instruction. Retrieved August 8, 2019, from <https://pamelasnow.blogspot.com/2019/08/dear-parents-welcome-to-confusing-world.html>
- Snow, P., & Clarke, A. (2015). The way we teach most children to read sets them up to fail. Retrieved July 24, 2019, from <https://theconversation.com/the-way-we-teach-most-children-to-read-sets-them-up-to-fail-36946>
- Snow, Pamela C. (2016). Elizabeth Usher Memorial Lecture: Language is literacy is language - Positioning speech-language pathology in education policy, practice, paradigms and polemics. *International Journal of Speech-Language Pathology*, 18(3), 216–228. <https://doi.org/10.3109/17549507.2015.1112837>
- Snowling, M. J. (2014). Dyslexia: A language learning impairment. *Journal of the British Academy*, 2(June), 43–58. <https://doi.org/10.5871/jba/002.043>
- Snowling, M. J., & Hulme, C. (2011). Evidence-based interventions for reading and language difficulties: Creating a virtuous circle. *British Journal of Educational Psychology*, 81(1), 1–23. <https://doi.org/10.1111/j.2044-8279.2010.02014.x>
- Stuart, M., & Stainthorp, R. (2015). *Reading Development and Teaching*. Reading Development and Teaching. Thousand Oaks, CA: Sage. <https://doi.org/10.4135/9781473920170>
- Styles, S. J. (2018). *PEBBLES: Perceptual Experience Battery - BaseLine Experiment Screener*. DR-NTU (Data), V2.

- <https://doi.org/https://doi.org/10.21979/N9/XSCH9X>
- Styles, S. J. (2019a). How language mixes contribute to effective bilingualism and effective biliteracy in Singapore: Surveying the Language Landscape. In *OER Knowledge Bites* (Vol. 10, pp. 12–13). Singapore: National Institute of Singapore.
- Styles, S. J. (2019b). *Singapore's languages at a glance ... Changes over time*. <https://doi.org/http://10.6084/m9.figshare.7781660>
- Suber, P. (2012). *Open Access*. The MIT Press. The MIT Press.
- Suzanna, H., & Brown, A. (2000). The [e] and [a] vowels in Singapore English. In A. Brown, D. Deterding, & E. L. Low (Eds.), *The English language in Singapore: Research on pronunciation* (pp. 93–99). Singapore: Singapore Association for Applied Linguistics.
- Test, D. W., Kemp-Inman, A., Diegelmann, K., Hitt, S. B., & Bethune, L. (2015). Are online sources for identifying evidence-based practices trustworthy? An evaluation. *Exceptional Children*, 82(1), 58–80. <https://doi.org/10.1177/0014402915585477>
- Torgerson, C., Brooks, G., Gascoine, L., & Higgins, S. (2018). Phonics: reading policy and the evidence of effectiveness from a systematic 'tertiary' review. *Research Papers in Education*, 34(2), 208–238. <https://doi.org/10.1080/02671522.2017.1420816>
- Torgerson, C. J., & Torgerson, D. J. (2003). The design and conduct of randomised controlled trials in education: Lessons from health care. *Oxford Review of Education*, 29(1), 67–80. <https://doi.org/10.1080/03054980307434>
- Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Herron, J., & Lindamood, P. (2010). Computer-assisted instruction to prevent early reading difficulties in students at risk for dyslexia: Outcomes from two instructional approaches. *Annals of Dyslexia*, 60(1), 40–56. <https://doi.org/10.1007/s11881-009-0032-y>
- Treiman, R. (2018). What research tells us about reading instruction. *Psychological Science in the Public Interest*, 19(1), 1–4. <https://doi.org/10.1177/1529100618772272>
- Vadasy, P. F., & Sanders, E. A. (2008). Repeated Reading Intervention: Outcomes and Interactions With Readers' Skills and Classroom Instruction. *Journal of Educational Psychology*, 100(2), 272–290. <https://doi.org/10.1037/0022-0663.100.2.272>
- Vadasy, P. F., & Sanders, E. A. (2009). Supplemental fluency intervention and determinants of reading outcomes. *Scientific Studies of Reading*, 13(5), 383–425. <https://doi.org/10.1080/10888430903162894>
- Wanzek, J., Vaughn, S., Scammacca, N., Gatlin, B., Walker, M. A., & Capin, P. (2016). Meta-Analyses of the Effects of Tier 2 Type Reading Interventions in Grades K-3. *Educational Psychology Review*, 28(3), 551–576. <https://doi.org/10.1007/s10648-015-9321-7>
- Weinstein, Y., Sumeracki, M., & Caviglioli, O. (2019). *Understanding how we learn: A visual guide*. New York: Routledge.
- Willingham, D. T. (2017). *The reading mind: A cognitive approach to understanding how the mind reads*. John Wiley & Sons.
- Winter, B. (2013). *Linear models and linear mixed effects models in R with linguistic applications*. Retrieved from <http://arxiv.org/abs/1308.5499>

World Literacy Foundation. (2018). *The Economic & Social Cost of Illiteracy: A snapshot of illiteracy in a global context*. Retrieved from <https://worldliteracyfoundation.org/wp-content/uploads/2019/06/TheEconomicSocialCostofIlliteracy-2.pdf>

Appendix 1: Preregistration codebook for Systematic review

Leung WT & Styles SJ (2018). 'Reading Intervention Quality: Preregistration Docs', DR-NTU (Data)
<http://dx.doi.org/10.21979/N9/RAQM8H>

Reading Intervention Quality: Preregistration Codebook (Items from PRISMA-P)

Leung, Wai Tung
Styles, Suzy J

Psychology, School of Social Sciences, Nanyang Technological University
14 Nanyang Drive, 637332
Singapore

This Codebook gives detailed instructions on how to fill in the items in the Data Extraction Table for the preregistered projects on Reading Intervention Quality, and explains how the Quality Score will be computed.

The majority of the quality indicators listed here were taken from Bishop's (2013) methodological criteria for evaluating intervention studies. Studies will need to achieve a score of five essential quality indicators before the intervention's findings can be analyzed meaningfully. The first five quality indicators will therefore be administered as a ladder, e.g., if a study does not fulfill the first quality indicator (study participants are appropriately selected), it will get a score of zero, and will not be assessed on any subsequent criteria. If a study fulfills the first three quality indicators, but not the fourth, it will receive a score of three, and will not be evaluated on the fifth, or any subsequent criteria. Interventions receiving a score below five will be classed as reported with 'insufficient' evidence. Interventions receiving a score of five or more, will be classed as 'sufficient' evidence for further evaluation. Interventions with a minimum score of five will then be assessed for the additional quality indicators, which will be implemented as additional 'bonus points' on top of their score of five, to indicate the presence of the other quality indicators. A study can be awarded up to a total of 18 points.

Essential quality indicators

The criteria for quality indicators 1 to 13, and descriptors for Quality Indicator 1, 3-13 are lifted from Bishop's (2013) discussion, with mild editing. Example was lifted from Bishop (2008) as reference for coding.

1) Study participants appropriately selected

Y: Study participants should be comparable with those for whom the intervention is intended.

N: Inconsistencies between intervention intended use and participants characteristics.

The Dore Programme is promoted for dyslexia, dyspraxia or ADHD. DST cut-off used to select participants fell below the level for 'mild risk' of dyslexia (0.6); Out of a sample of 35 children, six had diagnosis of dyslexia, two had diagnosis of dyspraxia, and one had diagnosis of ADHD.

Bishop (2008) p.520

2) Study participants selected by objective criteria (for inclusion in the study)

Y: Identified by school using national or state criteria for dyslexia diagnosis or at-risk of reading difficulties, teacher nomination, standardised assessment, curriculum-based measurement probes, appropriate screening tools with reported sensitivity and specificity, doctor diagnosis of dyslexia.

N: No information provided; lack of clear guidelines on selection criteria.

Risk quotient score of at least 0.4 on the DST; Children selected from three classes of a mainstream school in the UK

Bishop (2008) p.520

3) Sample size selected to give adequate power (to detect an intervention effect)

Bishop (2013) cautioned that underpowered studies might lead to misleading conclusions that there is no effect reported although there might be a clinically meaningful effect.

Y: Power analysis was provided with sufficient explanation / Effect sizes and sample sizes reported, allowing for calculation of power.

N: No power analysis reported; insufficient sample size, resulting in low power to detect an effect size.

Total sample size of 35; no evidence that power was considered.

Bishop (2008) p.520

4) Intervention adequately described

Y: Clear descriptions of intervention should be provided, allowing for replications of study. The study described intervention procedures/components, materials and critical elements, or provided citation for sources of information.

Not fully described because of commercial sensitivity.

Bishop (2008) p.520

5) Appropriate control group included

Some authors assume that a treatment effect can be demonstrated with comparison of pre and post intervention scores of only a treated group, although studies have shown otherwise.

Y –Control group was included, and tested at pre, post and follow-up for all assessments.

N– Some information/ assessments were not reported, although control group might be included; no control group.

Control group included, but only for a subset of assessments, and not for follow up.

Bishop (2008) p.520

Type of control group (See Styles, 2018)

- 1: Treat/ wait – one group gets treatment first and the other group will only get treatment later.
- 2: Placebo control/ no treatment for control group
- 3: Best-practice as control/ alternative treatment for control group

(Additional info on intervention)

T0: Additional assessments conducted prior to baseline

T1: Baseline (pre-intervention) assessments

T2: Assessments mid-way of intervention (if applicable)

T3: Post-tests after intervention has ended

F1: Follow-up post-intervention

F2: Follow-up post-intervention

Additional quality indicators (proceed only if study has obtained a score of 5 on essential quality indicators)

6. Random/matched clinical controls (additional score 1)

Y – Presence of random allocation of participants to treatment and control group (i.e. participants first meet selection criteria before randomly assigned to either treatment or control group); or presence of group matching on relevant variables.

N – No random allocation; or matching of treatment and control group on relevant variables; no information on randomization or method available.

Random allocation mentioned, though no details of randomisation method.

Bishop (2008) p.520

7. Information on dropouts (additional score 1)

Y: N reported with stated reason for dropping out, and provided additional information such as test scores and treatment status.

N: No or inadequate information provided e.g. only N is reported.

Six children lost to follow-up phase of study; no information provided on their test scores or treatment status.

Bishop (2008) p.520

8. Primary outcome measures specified (additional score 1)

Y: Clear distinction between primary and secondary outcomes.

In phase I, all measures treated the same, though on a priori grounds the key measures are literacy-related tests and tests of cerebellar function; SATS data

specified as secondary in (uncontrolled) follow-up study.

Bishop (2008) p.520

9. Reliable, standardized assessment methods (additional score 1)

Y: Use of standardized assessments that are reliable, sensitive and valid. They should be age-appropriate when possible.

DST is a standardised screening test; authors used unpublished normative data to convert to decile scores. No information on test sensitivity. Raw scores only on cerebellar tests. Other measures, with no control data, included one standardised test (NFER Reading test – but standard scores not reported on this), teacher assessments and parent ratings of attention.

Bishop (2008) p.520

List of assessments:

BPVS, CELF, CTOPP, GORT, NARA, PPVT, TOWRE, TROG, WISC,W-J, WPPSI, YARC

- Score 1 if the full assessment battery was used.
- Score 2 if only a / some specific sub-test(s) of the assessment battery was/were used

10. Blind assessment (additional score 1)

In order to minimize bias, people who conduct assessments should not be aware of whether participants came from intervention or control group.

Y: Reported blind assessment with description

11. All key data reported (additional score 1)

Y: Reported *n*, mean, *SD* for pre and post for both treatment and control group for all outcomes.

Means only presented.

Incomplete presentation (means only). Missing data on over 50% for electronystagmography. Data on treated group only (subset) for NFER reading test, SATS and attention ratings.

No data reported for an assessment at a 3rd time point that would have provided opportunity to see whether results from initial phase replicated.

Bishop (2008) p.520

12. Intervention effect appropriately analysed (additional score 1)

Y: Effects of intervention were assessed through comparison between intervention and control groups. Studies cannot achieve this if there was no appropriate control group. [Refer to **Appropriate control group included**]

ANOVA (described as 'one-tailed') for DST only; Main focus on gains in treated children only.

Bishop (2008) p.520

13. Correction for multiple comparisons (additional score 1)

Studies are susceptible to spurious findings if large numbers of comparisons are conducted in an unconstrained way.

Y: Bonferroni corrections were appropriately applied for post hoc tests.

14. Selective outcome reporting bias minimized (additional score 1)

Y: Low risk of outcome reporting bias; no missing data or outcomes, which were measured but not reported.

15. Fair and appropriate interpretation of results (additional score 1 only when 1 is achieved)

Bishop (2008) noted that a study should take into account “study hypotheses, sources of potential bias or imprecision, and the dangers associated with a multiplicity of analyses and outcome” (p. 520).

1= appropriately mentioned weakness with good analysis; interpretation takes into weaknesses (best practice)

2= mentioned weaknesses, but data still interpreted as if there were no weakness

3= no weakness reported. No discussion of statistical issues arising from multiple analyses

No discussion of statistical issues arising from multiple analyses.

Bishop (2008) p.520

16. Open access to article (additional score 1)

Y: Article was available to all members of the public, as archived on an open access database. (Examples include: Gold Access publication in a traditional journal, publication in a fully non-traditional open access journal, archived on a pre-print server or institutional repository). To achieve this score, the anonymized data have to be available without special requests or registration.

17. Open access to intervention data (additional score 1)

Y: Additional information and data were available on open access databases.

Examples include Open Access supplementary materials for the journal article, Open Science Framework repositories, or institutional archives). To achieve this score, the anonymized data have to be available without special requests or registration.

18. Open access to intervention materials (additional score 1)

Y: Materials used for intervention were available on open access databases.

Examples include links to independent publication in an Open Access journal/repository; Open Access supplementary materials for the journal article; Open Science Framework repositories, or institutional archives). To achieve this score, the anonymized data have to be available without special requests or registration.

References

- Bishop, D.V.M. (2008). Criteria for evaluating behavioural interventions for neurodevelopmental disorders (Letter). *Journal of Paediatrics and Child Health*, 44, 520–521.
- Bishop, D. V. M. (2013). Research Review: Emanuel Miller Memorial Lecture 2012 - Neuroscientific studies of intervention for language impairment in children: Interpretive and methodological problems. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 54(3), 247–259.
<https://doi.org/10.1111/jcpp.12034>
- Styles (2018). Ethics in Practice: A Workbook for Junior Investigators in Psychology, Medicine and the Behavioural and Social Sciences. *Dr-NTU (Data)*
<http://dx.doi.org/10.21979/N9/RNASRX>

Appendix 2: Preregistration report for Systematic review

Leung WT & Styles SJ (2018). 'Reading Intervention Quality: Preregistration Docs', DR-NTU (Data)
<http://dx.doi.org/10.21979/N9/RAQM8H>

Reading Intervention Quality: Preregistration for Systematic review

Questions from AsPredicted.org

Leung, Wai Tung
Styles, Suzy J

Psychology, School of Social Sciences, Nanyang Technological University
14 Nanyang Drive, 637332
Singapore

1) Data collection.

Has any data been collected for this study already?

No data have been collected for this study yet.

2) Hypothesis.

What's the main question being asked or hypothesis being tested in this study?

What is the quality of evidence for different interventions for children aged 6 to 9 years old with reading difficulties: a comparison between interventions with and without phonics? This question will be addressed using a qualitative systematic review.

3) Dependent variable.

Describe the key dependent variable(s) specifying how they will be measured.

A scoring checklist (see Data Extraction Table) will be used to compute a quality "score" for each paper included, using criteria adapted from Bishop (2013). Description of the quality indicator checklist, and score computation can be found in the Reading Intervention Quality: Preregistration Codebook. A score of 5 is treated as a quality benchmark: Interventions that score less than 5 will be classed as having 'insufficient' evidence. Interventions that score 5 or more will be classed as having 'sufficient' evidence for further investigation.

4) Conditions.

How many and which conditions will participants be assigned to?

Two conditions: Reading interventions with phonics vs without phonics.

Quality of evidence between interventions with phonics and without phonics will be compared.

Additional data will be collected on interventions conducted in bilingual communities for exploratory analysis.

5) Analyses.

Specify exactly which analyses you will conduct to examine the main question/hypothesis.

The primary analysis is a chi-square test comparing how many interventions with and without phonics are considered to have 'sufficient' or 'insufficient' evidence. In addition, for those interventions with 'sufficient' evidence, we will also conduct a t-test comparing the quality scores of interventions with and without phonics, which will indicate how many of the quality criteria each intervention meets. These analyses will be used in a qualitative synthesis of the state of evidence for different intervention types conducted over the past 10 years.

6) Outliers and Exclusions.

Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

Inclusion criteria will be peer-reviewed studies on reading intervention for children aged 6 to 9 with reading difficulties; conducted in school settings; administered in groups; with intervention duration of at least 10 weeks. Studies will be included if they were published from 2008 to 2017. Exclusion criteria will be mainstream classroom instructional approaches; single-subject studies; one-to-one tutoring.

'Reading difficulties' are defined as affecting children "whose reading is at least one SD, or one grade below the expected level with no reported problems that could explain their impaired ability to learn reading" (Mcarthur, 2011, p.6).

For further details, see the PRISMA Protocol (Checklist item 11) and the Preregistration Codebook.

7) Sample Size.

How many observations will be collected or what will determine sample size?

The number of articles included in the systematic review pipeline will be determined by a search of peer-reviewed Journal article databases. See PRISMA Protocol Checklist items 8-10. Eligibility criteria for inclusion in analysis is based on the pre-decided search strategies and exclusion criteria outlined in the PRISMA Protocol: Checklist items 8-12.

8) Other.

Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

For studies on children growing up in bilingual communities (e.g. Singapore, Hong Kong), additional data will be collected on the location of the intervention and the status of bilingualism in mainstream society in that location (i.e, majority/minority). This subset will be considered for further descriptive analysis. Any statistical analysis will be considered to be exploratory.

One additional exploratory investigation will be conducted, looking at the trajectory of the various quality indicators over time.

9) Name.

Give a title for this AsPredicted pre-registration. Suggestion: use the name of the project, followed by study description.

Evaluating quality of evidence in reading interventions with phonics training and those without for children aged 6 to 9 years old with reading difficulties: Protocol for a systematic review

Appendix 3: PRISMA Protocol for Systematic review

Leung WT & Styles SJ (2018). 'Reading Intervention Quality: Preregistration Docs', DR-NTU (Data)
<http://dx.doi.org/10.21979/N9/RAQM8H>

Reading Intervention Quality:

PRISMA Protocol

(Items from PRISMA-P)

Leung, Wai Tung

Styles, Suzy J

Psychology, School of Social Sciences, Nanyang Technological University 14 Nanyang Drive,
637332 Singapore

Administrative information

1. Title

Evaluating quality of evidence in reading interventions with phonics training and those without, for children aged 6 to 9 years old with reading difficulties: Protocol for a systematic review

2. Registration

Preregistration documents lodged with timestamp on institutional data repository with version control: Leung WT & Styles SJ (2018). 'Reading Intervention Quality: Preregistration Docs', DR-NTU (Data) <http://dx.doi.org/10.21979/N9/RAQM8H>

Introduction

3. Rationale

Reading difficulties affect many children. This is significant as reading is intertwined with language, attention and even academic achievement later in life. Early intervention is critical in remediating the difficulties experienced by primary school children.

Phonics, which refers to the linkages between written symbols and phonemes, is an essential component of reading instruction during the early years and lower primary levels. Several well-known national inquiries in the UK, US and Australia have recommended young children to be provided with phonics instruction (Rose, 2006; Rowe, 2005; NICHD, 2000). Successful interventions that address reading difficulties typically involve phonics, although research for the past decade has included interventions with and without phonics.

Poorly designed educational interventions could potentially compromise the conclusions of a study. Due to the varying degree of quality of evidence in reading research, teachers and educational administrators might not be sufficiently well informed to decipher biases present in the literature. This has potentially grave consequences, as children might not be taught using high quality evidence-based practices.

In this systematic review, I will explore the proportion of literature that reflects well-designed reading interventions for primary school children aged 6 to 9. In addition, I will also explore the proportion of studies from bilingual populations where English Language is one of the formal languages of instruction in school (e.g. Hong Kong and Singapore), beyond predominantly monolingual English populations where reading research is heavily relied on.

4. Objectives

The aim of this systematic review is to evaluate the quality of evidence of interventions for children with reading difficulties. The proposed systematic review will answer the following questions:

1. What is the quality of evidence in interventions with phonics training compared to those without?
2. What is quality of evidence for reading interventions in bilingual communities?

Methods

5. Eligibility criteria

Studies will be selected according to the criteria outlined below.

Report characteristics

- Years of publication considered: 2008 – 2017. Research in the last 10 years will be included, providing a reasonable time range to investigate the quality of evidence in reading interventions. In addition, the time frame of study follows after the publication of Gersten et al. (2005), which was an influential paper cited 832 times so far, calling for more high quality research in education.
- Articles published in English
- Peer-reviewed journal articles only

Study characteristics

This review will include all peer-reviewed studies in reading interventions targeting children aged 6 to 9 years old with reading difficulties. This review will only include studies conducted in school settings, administered in groups, with intervention duration of at least 10 weeks. I will exclude (i) mainstream classroom instructional approaches, which do not specifically target only children with reading difficulties, (ii) single-subject studies which have limited generalizability, (iii) one-to-one tutoring, which requires intensive resourcing to be carried out in schools effectively.

Participants

Children with reading difficulties are defined as those “whose reading is at least one SD, or one grade below the expected level with no reported problems that could explain their impaired ability to learn reading” (Mcarthur, 2011, p.6). Children with a diagnosis of learning difficulties (other than dyslexia) or medical problems will not be included in the review. This review will be restricted to English-speaking poor readers, including both monolinguals and bilinguals.

6. Information Sources

Digital searches for articles published 2008-2017, in the following repositories:

1. PubMed
2. PsycINFO
3. ERIC
4. Scopus
5. Social Science Citation Index

7. Search Strategy

Title, Abstract and Keywords will be searched using the following conceptual structure: “reading difficulty” AND “intervention”

“reading difficulty” will be defined as any of the following:

read* difficult*
read* delay
read* defic*
read* impair*
read* disorder*
poor read*
weak read*
slow read*
dyslex*

“intervention” will be defined as any of the following:

intervention
remedial read*
remedial teach*
read* program*
remedia*
read* assist*

8. Study Records

Results from the literature search will be uploaded to Distiller Systematic Review (DSR) Software, an Internet based software program that facilitates tracking of articles and removal of duplicates during the study selection process only.

I will independently screen the titles and abstracts obtained from the search against the inclusion criteria. Full reports will be obtained for potentially relevant papers and assessed against the inclusion criteria. The full text reports will be assessed for eligibility and to decide whether these meet the inclusion criteria.

Through the use of the Data Extraction Form, one single data coder will extract data from the included studies based on characteristics of participants, specified quality indicators, type of interventions, and outcome measures. Refer to Data Extraction Form for details. Piloting of the exclusionary criteria and Date Extraction Form has been trialed on 5 articles known to the review author. Further refinement will be described in subsequent reports.

9. Data Items

I will collect data from each study to investigate the quality of evidence reported, based on pre-decided quality indicators for evaluating intervention studies, adapted from Bishop (2013), as follows:

1. Did the samples give adequate power?
2. Were study participants appropriately selected?
3. Were study participants selected by objective criteria?
4. Were there random/matched clinical controls?
5. Was there an appropriate control group?
6. Did the article provide information on dropouts?
7. Was the intervention adequately described?
8. Were the primary outcome measures specified?

9. Were reliable and standardized instruments used to measure participants' abilities?
10. Was the measurement blind to group?
11. Was all key data reported?
12. Was the intervention effect appropriately analyzed?
13. Was there correction for multiple comparisons?

Additional quality indicators

14. Was the risk of selective reporting bias minimized? (c.f. Cochrane Collaboration)
15. Was the interpretation of results fair? (Bishop, 2008) (i.e. takes into account study hypotheses, sources of potential bias or imprecision, and the dangers associated with a multiplicity of analyses and outcomes)
16. Open access to journal article (c.f. Open Access Mandate)
17. Open access to data? (c.f. Open Access Mandate)
18. Open access to materials? (c.f. Open Access Mandate)

These quality indicators can be found in the Data Extraction Table. A scoring checklist will be used to obtain a quality "score" for each paper included. For details of how the score is computed, see the Preregistration Codebook.

10. Outcomes

For analysis plan see 'Preregistration for Systematic Review'

11. Assessment of risk of bias in included studies

In this review, risk of bias will be reflected in the Quality Score for each included study. See item 12, above. This allows for the assessment of risk of bias for the five items outlined in Cochrane risk of bias tool.

12. Data synthesis

Due to the heterogeneity of studies on reading difficulties, no meta-analysis has been planned. Instead, a qualitative synthesis will be provided with information presented in the text and tables to summarize and explain the characteristics and findings of the included studies. For analysis plan, see Item 5 and 8 of the Preregistration for Systematic Review.

13. Meta-biases

N.A.

14. Confidence in Cumulative evidence

Following analysis, the quality of evidence for different types of interventions will be discussed in a qualitative synthesis.

References

- Bishop, D. V. M. (2013). Research Review: Emanuel Miller Memorial Lecture 2012 - Neuroscientific studies of intervention for language impairment in children: Interpretive and methodological problems. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 54(3), 247–259.
<https://doi.org/10.1111/jcpp.12034>
- Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional children*, 71(2), 149-164.
- National Institute of Child Health and Human Development (NICHD). (2000). Report of the National Reading Panel: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: reports of the subgroups. NIH Publication No.00-4754. Washington, DC: U.S. Government Printing Office.
- Rose, J. (2006) Independent Review of the Teaching of Early Reading. Nottingham: DfES Publications.
- Rowe, K. (2005). Teaching reading: Report of the National Inquiry into the Teaching of Literacy. *Canberra: Department*

Appendix 4: Differences between protocol and review

Differences between protocol and review

1. Limiters on publication date, type, language and age group were included when possible during the search on the specified digital repositories. See Appendix 1 for details.
2. Inclusion criteria was modified to include kindergarten children as kindergarten was considered to be the first year of formal schooling in some countries. Some studies targeted students with reading difficulties ranging from kindergarten to first graders without distinguishing between them.
3. Computerised programs were included as it does not require intensive resourcing (e.g. attention from teachers/ paraeducators). Exclusion criteria for other forms of one-to-one tutoring remained unchanged.
4. When there is a combination of both group and one-to-one tutoring, it was included when one-to-one tutoring made up no more than half of intervention time.
5. We had planned to gather more information on studies targeting children growing up in bilingual communities. However, out of 32 studies in the qualitative synthesis, only one study was from a bilingual community. Chia & Houghton (2011) was conducted in Singapore with children diagnosed with dyslexia. However, the article was found not to have sufficient evidence and was excluded from further analysis. Thus, no further descriptive or statistical analysis was possible.
6. Calculations using G*power was not specified in the pre-registration codebook. The following method was used:

Using G*Power, with an effect size of 0.5, alpha error probability of 0.05 and power of 0.85, it was determined that a sample size of 59 per group will be sufficient to detect a medium- sized effect (see below). Hence, studies with 59 participants or more per group will be regarded as having sufficient power to detect a difference. For studies with less than 59 participants, a sensitivity power analysis will be used to determine the minimal detectable effect. However, if no effect sizes were reported in the study for outcome measures, a post-hoc power analysis will be used to test whether the study had sufficient subjects to detect the statistics reported, using means and *SD*.

Studies with at least one outcome measure with sufficient power are accepted.

t tests - Means: Difference between two independent means (two groups)

Analysis: A priori: Compute required sample size

Input:	Tail(s)	=	One
	Effect size d	=	0.5
	α err prob	=	0.05
	Power (1- β err prob)	=	0.85
	Allocation ratio N2/N1	=	1
Output:	Noncentrality parameter δ	=	2.7156951
	Critical t	=	1.6580957
	Df	=	116

Sample size group 1	=	59
Sample size group 2	=	59
Total sample size	=	118
Actual power	=	0.8542756

Appendix 5: Preregistration report for ‘What do words sound like in Singapore?’

Preregistered Design for ‘What do words sound like in Singapore?’

Wai Tung Leung & Suzy J Styles
Psychology, School of Social Sciences, Nanyang Technological University
14 Nanyang Drive, 637332 Singapore

To cite this work:

Leung WT & Styles SJ (2019). Preregistration for What do words sound like in Singapore? *DR-NTU (Data)*.
<https://doi.org/10.21979/N9/UAZ7AS> Archive date: (20 Mar 2019)

Preregistration questions using template from Open Science Framework (2019)

Study information

Title

What do words sound like in Singapore?

Introduction

Phonics refers to the relationship between graphemes and phonemes in the English Language. (Read Oxford, n.d.) defined systematic phonics instruction as “reading instruction programs that teach pupils the relationship between graphemes and phonemes in an alphabetic writing system” (p. 12). The word ‘systematic’ refers to the teaching of “grapheme-phoneme correspondences in an ordered manner” (p. 12). Over the years, national inquiries by the Australian, UK, and US education systems (Rowe, 2005; Rose, 2006; NICHD, 2000) have specified the importance of explicit and systematic phonics in teaching children how to read. Various prominent researchers have reiterated these points. For example, Willingham (2017) highlighted that for a reader to successfully decode, one needs to visually discriminate one letter from another, distinguish speech sounds, and learn the mappings between letters and corresponding sounds. Seidenberg (2018) also elaborated extensively on research that has demonstrated the importance of the phonological pathway in beginning reading, citing evidence from linguistic, psychological and computational modeling research.

Pronunciation differences between Singapore English and other varieties of English, such as British English, have been documented for vowels and consonants (Low & Brown, 2005; Deterding, 2007). However, phonics instruction in Singapore is based on materials developed for other varieties of English, and therefore does not take into consideration how the local pronunciations of English words differ from other varieties, for example, in the pronunciation of vowels. Given the lack of literature exploring the implications of pronunciation differences on the teaching of phonics in the Singapore context, the lack of

localization could affect students' ability to understand the regularities of the English writing system, and to gain mastery of Singaporean grapheme-phoneme correspondences. In this study, we focus on the intuitions of adult Singaporeans who are skilled readers of English, about the identity of individual sounds, specifically investigating monophthong vowels (see below). This study will help to provide a baseline describing skilled adult readers' linguistic beliefs about spoken English in Singapore, which can help to drive future development of reading instruction and intervention materials.

Research Questions

The purpose of this project is to find out about Singaporean English speakers' intuitions about vowel pairs that are contrastive in other standard varieties of English. For example, in Standard British English, and standard Australian English, the words "pitch" and "peach" differ in the pronunciation of the medial vowel, making them a 'minimal pair'. Previous research on the pronunciation of Singaporean English vowels suggests that certain vowel pairs have 'merged' into a single phonemic representation (e.g., tense/lax pairs like in "pitch" and "peach"), due to substantial overlap in where they are articulated in vowel space (as described by F1/F2). Whether or not the vowels differ in an alternative feature of articulation (e.g., vowel duration etc) is largely undescribed.

By investigating speakers' intuitions about the articulation of different vowels in Singaporean English, this study will allow us to refine a phonemic model of Singaporean English vowel-space that is aligned to speaker intuitions.

Vowel pairs suspected of merger (test vowel pairs)

Vowel pair	Examples
i:/ɪ	beet / bit
æ/e	bat / bet
ɑ:/ʌ	Bart / but
ɔ:/o	bought / bot
u:/ʊ	suit / soot

In order to test whether these vowel categories have merged, we will ask people to make decisions about whether they think 2 words (a minimal pair) are pronounced the same or different. Participants' responses to these word pairs will be compared against their responses for a set of control words containing vowels not suspected of merger (control word pairs). If participants respond differently to the test words and the control words (i.e., with higher 'same' responses for test words than for control words), this will suggest that the vowels have undergone a merger in the Singapore variety of English.

Vowels not suspected of merger (control vowel pairs)

Vowel pair	Examples
ʊ/o	look / lock
æ/ʌ	hat / hut
ɪ / e	bit / bet
ɔ:/ɑ:	bought / Bart
u:/i:	boot / beet

To ensure participants are aware that both “same” and “different” responses can be correct answers in different trials, homophone-pairs are included as filler trials. Homophones, such as ‘bye’ and ‘buy’, sound identical but are spelt differently. The inclusion of homophones therefore helps to mask the purpose of the experiment. Homophone trials will not be included in the primary hypothesis testing.

Stimulus lists will be constructed so that one third of the trials are vowel pairs not suspected of merger (expected response: “different”); one third of the trials are homophones (expected response: “same”), and one third of the trials are vowel pairs suspected of merger (expected response: unknown). For this third type of trial, participants might respond with either “same” or “different”, depending on the structure of their vowel categories.

Hypotheses

Primary hypothesis.

H1a. We predict people will give more ‘same’ responses for minimal pairs containing vowels suspected of merger (test vowel pairs) than they will for minimal pairs containing vowels not suspected of merger (control vowel pairs).

H1b. We also predict that different vowel pairs may show different degrees of merger. This prediction is exploratory, and we will report the percentage of ‘different’ responses for each vowel pair, as a way of describing the degree of merger.

Secondary hypothesis.

H2. We predict that people will be slower to respond for minimal pairs containing vowels suspected of merger than they will be for minimal pairs containing vowels not suspected of merger.

We have a secondary research question about whether participants’ articulations match their intuitions (with more ‘same’ responses for vowel pairs that overlap in their speech more). For each Singaporean participant, we will measure the acoustic formants of vowels used in the study, and compute the difference between the positions in vowel space for pairs suspected of merger. We will compare this distance to the participant’s percentage of ‘different’ responses in the same/different task. As the acoustic analysis of formants is labour intensive, we intend to explore this in a separate paper, where acoustic analysis methods will be preregistered in more detail before audio is processed.

H3. If participants’ intuitions about vowels match up with their own articulations, then the closer two vowels are in vowel space, the more likely a participant is to rate those vowels as the same (i.e., a positive correlation between distance and difference). On the other hand, if participants’ intuitions about vowel sounds are not linked to their own articulation of those vowels, then the two will be uncorrelated. This hypothesis will tell us about the mental model Singaporeans have about the sound system of their language (their meta-linguistic awareness).

We have a supplementary research question about the æ/e contrast. Based on the intuitions of two speakers of Singaporean English, some participants may articulate the /æ/ phoneme differently when followed by a voiced alveolar stop (e.g., ‘head’). As such, the main analysis

uses æ/e words that do not end with /d/ after the vowel. A supplementary set of /-æd/ and /-ed/ words have been added as fillers for more detailed investigation. Differences in pronunciation may further influence same/different judgments for these tokens, i.e., participants might report a higher proportion of 'different' scores for æ/e pairs ending with /d/ (e.g., 'head had'), as compared to æ/e pairs ending with other consonants (e.g., 'trek track').

H4a. The distribution of /æ/ articulations in acoustic space will differ when it precedes a /d/, such that it will be less similar to /e/.

H4b. Participants will judge more æ/e minimal pairs as 'different' when the vowel precedes a voiced alveolar stop /d/ than when it precedes another consonant.

Since we know of no previous analysis of this kind, these analyses will be primarily descriptive, and exploratory in nature.

Sampling Plan

Data collection procedures

For this study, participants will be recruited on the campus of the host University. Collection of data for the study is expected to end by July 2019.

Inclusion criteria: As our study investigates how words sound in Singaporean English, only participants who are Singaporeans or have mostly grown up in Singapore with typical hearing, will be recruited for participation.

Sample size

The researchers aim to collect 60 participants, depending on the availability of participants on campus during the testing period.

Sample size rationale

We used G*Power (Faul, Buchner & Lang, 2009) to conduct a power analysis. In order to achieve .95 power to detect a medium effect size of .5 at the standard .05 alpha error probability, using a paired t-test, it was determined that a minimum sample size of 45 will be sufficient to detect a medium-sized effect. To balance the needs of high power against participant availability, we aim to recruit a minimum of 50, and a maximum of 60 adult participants from the University community (Refer to power analysis document for more information).

Stop rule

We plan to collect data until a maximum of 60 participants have completed the procedure, or until 31 July 2019 at the latest (whichever comes first).

Variables

All participants belong to the same group.

Measured variables

All participants will take part in two primary tasks.

Task 1 (Pronunciation intuition task)

Participants will be asked to share their intuitions about the pronunciation for written English words. Participants will be asked whether a pair of words are pronounced the same or different (e.g., 'source' vs 'sauce'). The primary measured variable in this task is participants' binary choice of whether the words in a given pair sound the same or different. When the words are rated as 'same', they are given a score of 0 and when the words are rated as 'different', they are given a score of 1. The total difference score for a contrast will be the sum of all difference values.

The secondary measured variable is time taken for participants to respond, measured in milliseconds. Treatment of reaction time data is described under 'Transformations'.

Task 2 (Audio production task)

Participants will be asked to read a list of single words, comprised of test words used in Task 1. Audio recording will acoustically analysed. Acoustic measures will include F0, F1, F2, F3 and duration.

Other rating or measurements

Participants will also complete the Vision and Hearing sections of the PEBBLES (The Perceptual Experience Battery - Baseline Experiment Screener <https://doi.org/10.21979/N9/XSCH9X>), and a brief demographic survey about age, gender and language background, to provide context about our participants.

For each participant, the binary decisions about whether minimal pairs sound the same or different may be converted to a composite score for visualization purposes (computed as % of answers rated as 'different').

Word pairs for minimal pairs were selected from Higgins (n.d.) while homophones were obtained from Feldman, Cosman, Feldman, & Quiros (n.d.). Word pairs were evaluated using word frequency on the Zipf scale, based on the SUBTLEX-UK word frequencies (Heuven, Mandera, Keuleers, & Brysbaert, 2014). According to the authors' estimate, a mean Zipf value of 2.5 and below can be considered low-frequency words. As such, word pairs selected have a minimum of Zipf value of 2.5 to avoid the use of words with low-frequency words. Words with no Zipf value are usually uncommon words with inflections, for e.g. 'parches', and are not considered for selection. Based on the available pool of word pairs to choose from, 10 word pairs with higher Zipf values were then chosen for each contrast, with the exception of contrast 'ʊ/u' and 'æ/e' due to insufficient words of sufficiently high frequency. Word pairs were selected to avoid repetition.

Word naming times measure the amount of time taken to read aloud a single word presented on a computer screen, and is known to be an index of the speed of visual word recognition (Heuven, Mandera, Keuleers, & Brysbaert, 2014). As word naming times reflect time taken in processing isolated words, it is a useful measure for modelling the variability that might come from accessing the phonological forms of words in a pair. Word naming times (see attached) were obtained from (Kuperman, Drieghe, Keuleers, & Brysbaert, 2013). As words will be presented as a pair in the experiment, each pair will be represented by the

sum of the two word naming times Words without word naming times were excluded. This also resulted in the reduction of word pairs contrast 'ʊ/u' from 10 to 7.

Design plan

Study type

Experiment

Blinding

No blinding is involved in this study.

Study design

Repeated measures (within-subjects)

We have a within-subjects design with 1 factor (contrast) with 2 levels (merger; no merger) in Hypothesis 1a; and 1 factor (contrast) with 5 levels in Hypothesis 1b (different vowel pairs).

Randomization

All participants will go through the same procedure. The order of stimulus presentation, and orientation of word presentation (left/right) will be randomized within JSPsych, which is a web-based JavaScript library for running the experiment. Participants will indicate "same" or "different" using either 'a' or 'l' keys.

To address the potential problem of handedness of participants having an effect on reaction time (for e.g., right-handed participants possibly responding faster when pressing the key using their right hand), participants will swap response keys half way through the procedure. Starting side will be counterbalanced between participants.

Data handling and Analysis Plan

Statistical models

Hypothesis 1a

We plan to use the lme4 package (Bates et al., 2015) in R (R Core Team, 2018) to perform a generalized linear mixed-effects model of the relationship between merger (merged, unmerged), vowel pair, and participants' response. Since the variable is binary, the model will be fitted to a logistic function.

- 1) Fixed effects: merger (2 levels: merged, unmerged)
- 2) Random effects: intercepts for subjects and items, as well as by-subject and by-item random slopes for the effect of merger

Hypothesis 1b

We plan to use the lme4 package (Bates et al., 2015) in R (R Core Team, 2018) to perform a generalized linear mixed-effects model of the relationship between different vowel contrasts and participants' response. Since the variable is binary, the model will be fitted to a logistic function.

- 1) Fixed effects: contrast (5 levels: i:/ɪ, æ/e, ɑ:/ʌ, ɔ:/o, u:/ʊ)

- 2) Random effects: intercepts for subjects and items, as well as by-subject and by-item random slopes for the effect of contrast

For both hypothesis 1a and 1b, we will remove non-significant factors to identify the best-fitting model with merger (1a) or contrast (1b) as a fixed effect. To establish whether the respective fixed effect is a significant predictor of participants' % difference score, the Likelihood Ratio Test will be used to obtain p -values by comparing the final model against the matching model without the critical experimental variable (1a: merger; 1b: contrast), as recommended by Winter(2013).

Hypothesis 2

We plan to use the lme4 package (Bates et al., 2015) in R (R Core Team, 2018) to perform a linear mixed effects analysis of the relationship between merger (merged, unmerged) and reaction time.

Predictor variables entered into the model will be:

- 1) Fixed effects: merger (merged, unmerged), and word naming times of word pairs
- 2) Random effects: intercepts for subjects and items, as well as by-subject and by-item random slopes for the effect of merger
- 3) Interaction effect: merger*word naming times

We will remove non-significant factors/interactions to identify the best-fitting one with merger as a fixed effect or interaction. To establish whether merger is a significant predictor of RT, the Likelihood Ratio Test will be used to obtain p -values by comparing the final model against the matching model without the critical experimental variable (merger), as recommended by Winter (2013). See Transformations for details about RT handling.

Hypothesis 3

In order to find out whether participants' intuitions about vowels match up with their articulations, we plan to perform a correlation analysis between distance in vowel space and participants' response (measured through % difference score). If the distribution of data is normal, Pearson's correlation will be used. But if data is not normal, Spearman's correlation will be used. Acoustic analysis methods will be specified in more detail separately before the audio recordings are processed.

Hypothesis 4

We intend to explore the acoustics of /-æd/ words in a separate paper, where acoustic analysis methods will be preregistered before audio is processed (i.e., before it is known whether people are articulating these tokens differently). The analysis will be primarily descriptive (exploratory) in nature.

Transformations

As reaction time distributions are rarely normal (i.e. non-Gaussian), we will check for normality by plotting histograms of reaction time by person. If the histogram of reaction time data is positively skewed with a long tail, then the data will be handled using either log or inverse transformation. First, using the method of visual inspection, the transformation with a histogram that best approximates normality for the majority of the dataset will be

selected. The histogram will also be checked visually for extreme outliers (e.g., those separated from the majority by a large gap). Extreme outliers will be trimmed. Following trimming, skewness and kurtosis using both absolute values as well as z-scores will be checked. For absolute values, skewness and kurtosis should be close to zero in a normal distribution. For z-scores calculated for both parameters, values greater than 2.58 or lesser than -2.58 are regarded as significant at $p < 0.01$. While z-scores are typically preferred as it allows for comparisons of scores from different distributions, Field, Miles & Field (2012) noted that z-scores for skewness and kurtosis need to be interpreted with caution with large sample sizes as such significance tests might yield significant values even with small deviations from normality. Field et al. (2012) recommends that a combination of methods should be used to determine whether data is normally distributed. Specifically, normality tests should be interpreted along with histograms and values of skew and kurtosis. As such, z-scores are only used as a point of reference but will not be applied as a strict exclusionary criteria.

If data is determined to be sufficiently close to normal using the reported methods above, the parametric tests described above will be used for analysis. However, if data is reported to be non-normal, individual RT data greater than 3 standard deviations from the mean will be trimmed, and the resulting distributions checked again for normality. But if the data still fails to approximate a normal distribution, non-parametric tests will be substituted.

Follow-up analyses

NA

Inference criteria

Using G*Power (Bates et al., 2015), we determined that for an alpha error probability of 0.05 and power of .95, a minimum sample size of 45 per group would be needed to detect a medium-sized effect of 0.5. Increasing alpha error probability from 0.05 to 0.01 would require a minimum sample of 66 participants. To align with standards within our field, we will use $p < .05$ as the threshold for statistical significance for all analysis. However, given recent developments in the interpretation of statistics, p -values of .005 and below will be treated as providing stronger statistical evidence.

Data exclusion

Participants who do not complete 50% of the trials will be excluded. Participants who indicate that a substantial proportion of homophone word pairs sound different will also be excluded, as this suggests the participant might not have understood the task requirements, or they may speak a different variety of English. Participants whose response pattern suggests that they failed to understand/did not follow task instructions will be excluded (e.g., 100% "same" response). The decision will be based on visual inspection of histograms for homophones. If a participant's responses were significantly different from the responses of the majority, then the participant's data will be excluded from analysis. Decisions on data exclusion will be made prior to inspection/analysis of the critical experimental conditions (i.e., merged/unmerged).

Missing data

Participants who do not complete 50% of the trials will be excluded. Refer to data exclusion for more information.

Appendix 6: Demographics and language questionnaire

Demographics and language

Brain, Language and Intersensory Perception at NTU

Code:

About You

What is your age (in years): _____

What is your gender: M / F / Prefer not to say/Other

What is your nationality?

What country did you mostly grow up in during pre-school (0-5 years)

What country did you mostly go to school in during primary school? (6-12 years)

What country did you mostly go to school in for high-school? (13-18 years)

If you have lived in another country for more than one year, please give details in the space below:

Language self-evaluation

Identity

How would you describe yourself:

Monolingual	Bilingual	Multilingual
<i>I have one language</i>	<i>I have two languages</i>	<i>I have more than two languages</i>

If you would like to add more information about this question, you please add in the space below:

If you speak more than one language, please list your languages in the space below:

Proficiency

Evaluate your skills for different communication tasks, in each of your languages...

Understanding when people speak

For each of your languages, write the name of the language, and give yourself a score

from 1-7, where 7 indicates native-level proficiency in understanding people speak, and 1 indicates understanding just a few words. A score of 0 means you don't have proficiency in un understanding. (e.g., I can say some words, but don't understand what others say).

	Proficiency Score (tick)							
Language Name	7	6	5	4	3	2	1	0

Speaking to others

For each of your languages, write the name of the language, and give yourself a score

from 1-7, where 7 indicates native-level proficiency in understanding people speak, and 1 indicates understanding just a few words. A score of 0 means you don't have proficiency in un understanding. (e.g., I can say some words, but don't understand what others say).

	Proficiency Score (tick)							
Language Name	7	6	5	4	3	2	1	0

Existing conditions

Do you have an existing diagnosis of dyslexia or related language condition?

Appendix 7: PEBBLES – The Perceptual Experience Battery

PEBBLES

PERCEPTUAL EXPERIENCE BATTERY
BASELINE EXPERIMENT SCREENER

DATE:

Vision

Q1. Which of these best describes your day-to-day sight?

(Circle one option. Answer for when you are wearing glasses or other vision aids you normally use)

- a. I have no light perception.
- b. I can tell by the light where the windows are.
- c. I can see the shapes of furniture in the room.
- d. I can recognize a friend by sight alone if I am close to their face.
- e. I can recognize a friend by sight alone if I am at arms' length away.
- f. I can recognize a friend by sight alone if I am on the other side of the room.
- g. I can recognize a friend by sight alone if I am on the other side of the street.

Questions adapted from Network 1000 Project by Douglas, Corcoran and Pavey (2006)

Q2. If you have a diagnosed vision condition, what is it called? _____

2a. From what age have you had this diagnosis? _____

Q3. If you use glasses or contact lenses, what is your prescription? _____

3a. From what age have you had this diagnosis? _____

3b. On a typical day, how many hours do you wear your glasses/contacts? _____

Q4. If you have had surgery to 'correct' an eyesight condition, what was the procedure? _____

4a. How old were you when you had this procedure? _____

In Question 1, if you answered 1a, 1b, or 1c, skip to Q8; If you answered 1d or above, please continue to Q5

Q5. Without glasses/contacts, what is your score in a standard vision test (e.g., 20:20)? _____

Q7. Are you aware of any differences in your colour vision, compared to other people? Y/N

Q7a. If yes, please describe _____

Q8. Are you aware of any other differences in your vision, compared to other people? Y/N

Q8a. If yes, please describe _____

PEBBLES

PERCEPTUAL EXPERIENCE BATTERY
BASELINE EXPERIMENT SCREENER

Hearing

Q1. Which of these best describes your day-to-day hearing?

(Circle one option. Answer for when you are using hearing aids, if you normally use them)

- a. I have no perception of sound.
- b. I can hear loud sounds and rhythms. I can't always tell if the sound I am hearing is a person talking, or another sound.
- c. I usually know when a sound is a person is talking, but I usually can't understand the words.
- d. I usually understand what someone says when they speaking loudly, and close to my ear.
- e. I usually understand what someone says when they are arm's length away, and the room is quiet.
- f. I usually understand what someone says when they speaking on the other side of the room, and the room is quiet.
- g. I usually understand what someone says, even when other people are talking in the room.

Q3. If you have a diagnosed hearing condition, what is it called? _____

3a. From what age has your hearing been like this? _____

Q4. If you use any hearing aids or cochlear implants, what device do you use? _____

4a. From what age have you used this device? _____

4b. On a typical day, how many hours do you wear your hearing devices? _____

4c. On a typical day, how many hours do you switch on your hearing devices? _____

Q5. Are you aware of any other differences in your hearing, compared to other people? Y/N

Q7a. If yes, please describe _____

Appendix 8: List of words used in ‘What do words sound like in Singapore?’

	Contrast	Word1	Word2
1	ʊ/u	full	fool
2	ʊ/u	hood	who'd
3	ʊ/u	look	Luke
4	ʊ/u	pull	pool
5	ʊ/u	pulling	pooling
6	ʊ/u	soot	suit
7	ʊ/u	wood	wooded
8	æ/e	and	end
9	æ/e	cattle	kettle
10	æ/e	land	lend
11	æ/e	bag	beg
12	æ/e	than	then
13	æ/e	lass	less
14	æ/e	sat	set
15	æ/e	bland	blend
16	æ/e	pan	pen
17	æ/e	gas	guess
18	i/ɪ	cheek	chick
19	i/ɪ	wean	win
20	i/ɪ	heat	hit
21	i/ɪ	peak	pick
22	i/ɪ	deep	dip
23	i/ɪ	steel	still
24	i/ɪ	cheap	chip
25	i/ɪ	least	list
26	i/ɪ	reach	rich
27	i/ɪ	sleep	slip
28	o/ɔ:	knotty	naughty
29	o/ɔ:	Don	dawn
30	o/ɔ:	stock	stalk
31	o/ɔ:	chock	chalk
32	o/ɔ:	cot	caught
33	o/ɔ:	odd	awed
34	o/ɔ:	popper	pauper
35	ɑ:/ʌ	calm	come
36	ɑ:/ʌ	staff	stuff
37	ʊ / o	shock	shook
38	ʊ / o	rocky	rookie
39	ʊ / o	folly	fully
40	i/e	will	well
41	i/e	listen	lesson
42	i/e	since	sense
43	i/e	middle	medal
44	i/e	gym	gem
45	i/e	built	belt

	Contrast	Word1	Word2
46	i/e	trick	trek
47	i/e	lift	left
48	i/e	miss	mess
49	i/e	wrist	rest
50	ɑ:/ʌ	calf	cuff
51	ɑ:/ʌ	branch	brunch
52	ɑ:/ʌ	task	tusk
53	ɑ:/ʌ	last	lust
54	ɑ:/ʌ	grant	grunt
55	ɑ:/ʌ	half	huff
56	ɑ:/ʌ	pass	pus
57	ɑ:/ʌ	graph	gruff
58	æ / ʌ	bat	but
59	æ / ʌ	as	us
60	æ / ʌ	rash	rush
61	æ / ʌ	lack	luck
62	æ / ʌ	lamp	lump
63	æ / ʌ	mad	mud
64	æ / ʌ	ankle	uncle
65	æ / ʌ	tack	tuck
66	æ / ʌ	track	truck
67	æ / ʌ	carriage	courage
68	ɑ: / ɔ:	farm	form
69	ɑ: / ɔ:	card	chord
70	ɑ: / ɔ:	fart	fort
71	ɑ: / ɔ:	park	pork
72	ɑ: / ɔ:	cart	court
73	ɑ: / ɔ:	starry	story
74	ɑ: / ɔ:	star	store
75	ɑ: / ɔ:	hard	hoard
76	ɑ: / ɔ:	scar	score
77	ɑ: / ɔ:	impart	import
78	i:/ u:	tea	to
79	i:/ u:	she	shoe
80	i:/ u:	he	who
81	i:/ u:	feed	food
82	i:/ u:	scene	soon
83	i:/ u:	mean	moon
84	i:/ u:	steal	stool
85	i:/ u:	cheese	choose
86	i:/ u:	beat	boot
87	i:/ u:	fleet	flute
88	ʊ / o	god	good
89	ʊ / o	cod	could
90	ʊ / o	knock	nook

	Contrast	Word1	Word2
91	ʊ / o	cocky	cookie
92	ʊ / o	posh	push
93	ʊ / o	box	books
94	ʊ / o	locking	looking
95	homophone	meat	meet
96	homophone	peace	piece
97	homophone	blew	blue
98	homophone	son	sun
99	homophone	suite	sweet
100	homophone	some	sum
101	homophone	cell	sell
102	homophone	muscle	mussel
103	homophone	beach	beech
104	homophone	him	hymn
105	homophone	flew	flu
106	homophone	cue	queue
107	homophone	which	witch
108	homophone	cent	scent
109	homophone	rap	wrap
110	homophone	laps	lapse
111	homophone	know	no
112	homophone	grate	great
113	homophone	knight	night
114	homophone	brake	break
115	homophone	right	write
116	homophone	main	mane
117	homophone	hall	haul
118	homophone	sight	site
119	homophone	mail	male
120	homophone	made	maid
121	homophone	toe	tow
122	homophone	dye	die
123	homophone	hairy	harry
124	homophone	roll	role
125	homophone	poll	pole
126	homophone	raise	rays
127	homophone	sail	sale
128	homophone	groan	grown
129	homophone	waist	waste
130	homophone	loan	lone
131	homophone	we	wee
132	homophone	might	mite
133	homophone	I	eye
134	homophone	throne	thrown
135	homophone	vain	vein

	Contrast	Word1	Word2
136	homophone	dew	due
137	homophone	pail	pale
138	homophone	pause	paws
139	homophone	doe	dough
140	homophone	plane	plain
141	homophone	seem	seam
142	homophone	wait	weight
143	homophone	hole	whole
144	homophone	pray	prey
145	homophone	slay	sleigh
146	homophone	wring	ring
147	homophone	waive	wave
148	homophone	seam	seem
149	homophone	bell	belle
150	homophone	cent	scent

Appendix 9: Data analysis for “What do words sound like in Singapore?”

Hypothesis 1a

Models:

m2: same ~ (1 + merger | participant) + (1 | stimulus)

m1: same ~ merger + (1 + merger | participant) + (1 | stimulus)

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
m2	5	3278.7	3311.8	-1634.3	3268.7			
m1	6	3167.5	3207.2	-1577.8	3155.5	113.18	1	<2.2e-16***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Hypothesis 1b

m2b: same ~ (1 + mcontrast | participant) + (1 | stimulus)

m1b: same ~ mcontrast + (1 + mcontrast | participant) + (1 | stimulus)

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
m2b	5	2871.2	2900.6	-1430.6	2861.2			
m1b	6	2872.6	2907.9	-1430.3	2860.6	0.6295	1	0.4275

Hypothesis 2

mod5.null: inv.rt ~ scale(NMG_T) + (1 + merger | participant) + (1 | stimulus)

mod5: inv.rt ~ merger + scale(NMG_T) + (1 + merger | participant) + (1 | stimulus)

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
m2	6	12	51.714	-1.2292e-12	2.4585e-12			
m1	7	14	60.333	-1.2292e-12	2.4585e-12	0	1	1

Appendix 10: Grapheme-phoneme correspondences (for vowels suspected of merger in Singapore)

Grapheme	/æ/	/e/
a	4192	-
a-e	147	-
e	-	3316
ea	-	139
e-e	-	79

Grapheme	/i:/	/ɪ/
i	38	5346
e	1765	-
y	1801	100
ee	249	-
ea	245	-
e-e	62	-
ie	62	-
i-e	44	339
ey	40	-
ea-e	30	-
ie-e	23	-
ei	16	11
ai	-	15
ui	-	16

Grapheme	/u:/	/ʊ/
u	907	200
oo	173	114
o	-	17
u-e	290	11
ew	60	-
ou	29	-
eu	28	-
ue	27	-
oo-e	12	-
o-e	12	-

Grapheme	/o/	/ɔ:/
o	1558	123
a	80	165
o-e	20	17
or	-	312
au	-	146
aw	-	75
ough	-	15
augh	-	12

Grapheme	/ʌ/	/ɑ:/
ar		474
a		44
ar-e		31
ear		18
u	1212	
o	112	
ou	30	
o-e	26	
u-e	23	
oo	7	

The numbers from the table were taken from Hanna et al. 1966, in which they analysed phoneme/grapheme correspondences in a large US corpus. The numbers reflect the number of words that were found to match the grapheme-phoneme category.

Appendix 11: Full-text articles excluded with reasons (For Chapter 3)

Source 1 PsycINFO		Exclusion reason
1	Zentall & Lee (2012)	Intervention occurred right after pre-test. One-time use of motivational encouragement.
2	Wolgemuth, Savage, Helmer, Harper, Lea, Abrami & Loudon (2013)	Pupils in the intervention were not identified or screened for reading difficulties.
3	González-Castro, Rodríguez, Núñez, Vallejo, González-Pienda (2014)	Not English-speaking population
4	Grills, Fletcher, Vaughn, Barth, Denton & Stuebing (2014)	Insufficient data on intervention outcomes; anxiety was the focus of the article. Study used data from Denton et al (2011)
5	Clemens, Oslund, Simmons & Simmons (2014)	Comparison of spelling skills with other indicators at the end of intervention- focus of the paper. Not focused on intervention in this paper.
6	Sanchez & O'connor (2015)	one to one instruction; Tier 3
7	Wanzek (2014)	Observational data from teachers
8	Cho, Roberts, Capin, Roberts, Miciak & Vaughn (2015)	Fourth grade students
9	Cartwright, Coppage, Lane, Singleton, Marshall & Bentivegna (2017)	Intervention lasted 5 weeks for study 2. No intervention in study 1.
10	Donovan & Marshall (2016)	No intervention was conducted
11	Fälth, Gustafson, Tjus, Heimann & Svensson (2013)	Not English-speaking population
12	Hagans & Good (2013)	First grade students from low SES backgrounds; not with reading difficulties
13	Law & Ghesquière (2017)	No intervention was conducted. Article used a longitudinal study design.
14	Kaminski & Powell-Smith (2017)	One-to-one instruction; multiple baseline design
15	Walcott, Marett & Hessel (2014)	Intervention lasted 9 weeks; participants had reading and attention difficulties
16	Baker, Park, Baker, Basaraba, Kame'enui & Beck (2012)	No intervention was conducted. Comparison between 2 different approaches in schools.
17	Goldstein, Olszewski, Haring, Greenwood, McCune, Carta & Kelley (2017)	Preschoolers; Mean age < 5
18	Gorp, Segers & Verhoeven (2017)	Not English-speaking population
19	Brimo (2016)	Some children had co-morbidity e.g. dysgraphia.
20	Wolgemuth, Abrami, Helmer, Savage, Harper & Lea (2014)	Reading difficulties not assessed.
21	Rodgers, D'Agostino, Harmey, Kelly & Brownfield (2016)	Study investigated nature of scaffolding; outcomes of students' literacy not investigated
22	Kim, Kim, Koh, Lee, Shin & Kim (2014)	Not English-speaking population
23	Good & Gillon (2014)	Children did not have reading difficulties.
24	Reutzel, Petscher & Spichtig (2012)	Three interventions were grouped in the control condition. Unable to distinguish the interventions.
25	Lawrence, Rolland, Branum-Martin & Snow (2014)	School-wide intervention

26	Ojanen, Ronimus, Ahonen, Chansa-Kabali, February, Jere-Folotiya & Puhakka (2015)	Not English-speaking population
27	Krafnick, Flowers, Napoliello & Eden (2011)	Intervention lasted 8 weeks.
28	Breteler, Arns, Peters, Giepmans & Verhoeven (2010)	average age = 10.33; range 8.0–15.98 (wide range)
29	Tyler, Hughes, Beverley & Hastings (2015)	Pupils identified were not screened for reading difficulties.
30	Lawton (2011)	Three subjects used in the study; single case studies
31	Steenbeek-Planting, van Bon & Schreuder (2012)	Not English-speaking population
32	Blachman, Schatschneider, Fletcher, Murray, Munger & Vaughn (2014)	One-to-one tutoring
33	Brinchmann, Hjetland & Lyster (2016)	Not English-speaking population
34	Chong, Moore, Nonis, Tang, Koh & Wee (2014)	One- to-one sessions
35	Toste, Capin, Vaughn, Roberts & Kearns (2017)	Intervention lasted 8 weeks.
36	Flaugnacco, Lopez, Terribili, Montico, Zoia & Schön	Mean age = 10 years
37	Farris, Ring, Black, Lyon & Odegard (2016)	Participants ranged from 6 to 14 years (wide range)
38	Bach, Richardson, Brandeis, Martin & Brem (2013)	Intervention lasted 8 weeks.
39	Phillips (2014)	Screening for inclusion used receptive language and syntax-related measures, not on reading.
40	O'connor, Bocian, Beach, Sanchez & Flynn (2013)	Does not compare between interventions explicitly; different tier support for students in different tiers
41	Le Roux, Swartz & Swart (2014)	Participants between 7 and 13 years of age
42	Nevo, Brande & Shaul (2016)	Not English-speaking population
43	Taub & Szente (2012)	No intervention was conducted
44	Edwards & Taub (2016)	No intervention was conducted
45	Hasko, Groth, Bruder, Bartling & Schulte-Körne (2014)	Intervention conducted in individual setting
46	Müller, Richter, Križan, Hecht & Ennemoser (2015)	Not English-speaking population
47	Foy (2009)	Comparison between 1:1 vs small group only
48	Vadasy & Sanders (2008)	fourth and fifth grader with reading fluency issues
49	Ritter & Saxon (2011)	Children did not have reading difficulties.
50	Menzies, Mahdavi & Lewis (2008)	All children received intervention (both at risk and those not risk)
51	Kerins, Trotter & Schoenbrodt (2010)	SLP does not typically operate within school-based interventions.
52	Guzel Ozmen (2011)	Not English-speaking population
53	Ross & Begeny (2011)	Comparison between 1:1 vs small group only
54	Keilmann & Wintermeyer (2008)	Not English-speaking population
55	Alt (2011)	participants had SLI not reading difficulties
56	Chard, Stoolmiller, Harn, Wanzek, Vaughn, Linan-Thompson & Kame'enui (2008)	Investigated the tiered system of support and corresponding interventions
57	Scheltinga, van der Leij & Struiksma (2010)	Not English-speaking population
58	Lipka & Siegel (2010)	Study not limited to those with reading difficulties.

59	Forgeard, Schlaug, Norton, Rosam, Iyengar & Winner (2008)	Insufficient information to determine inclusion and exclusion criteria
60	Edl, Humphreys & Martinez (2009)	Article provided discussion but no intervention.
61	Van den Bos (2008)	Not English-speaking population
Source 2 ERIC		
62	Filippini, Gerber & Leafstedt (2012)	Intervention lasted 8 weeks.
63	Therrien & Budin (2008)	two students; case study approach
64	Hornery, Seaton, Tracey, Craven & Yeung (2014)	Insufficient information to determine inclusion and exclusion criteria
65	Chapman & Tunmer (2016)	Critique on reading recovery
66	Pfenninger (2015)	English as Foreign language
67	Pfenninger (2016)	English as Foreign language
Source 3 Scopus		
68	Saine, Lerkkanen, Ahonen, Tolvanen & Lyytinen (2013)	Not English-speaking population
69	Chambers, Abrami, Slavin & Madden (2011)	Discussion of proposed reading intervention but no study was conducted.
70	Loeb, Gillam, Hoffman, Brandel & Marquis (2009)	Children in the study had both "language impairment and poor reading skills"
71	Deault, Savage & Abrami (2009)	Study not limited to those with reading difficulties.
72	Given, Wasserman, Chari, Beattie & Eden (2008)	mean age= 12
Source 4 SSCI		
73	Ritter, Colson & Park (2013)	Children in the study had "co-occurring language and reading impairments"
74	Thomson, Leong & Goswami (2013)	Intervention lasted for 6 weeks
75	Gebauer, Fink, Kargl, Reishofer, Koschutnig, Purgstaller & Enzinger (2012)	Not English-speaking population
76	Wanzek & Roberts (2012)	Fourth graders
77	Davidson & Hobbs (2013)	Study not limited to those with reading difficulties.
78	Bhide, Power & Goswami (2013)	intervention lasted for approximately 2 months; one-to-one instruction for musical intervention and either seen one-on-one or in pairs for GraphoGame
79	Ecalte, Kleinsz & Magnan (2013)	Not English-speaking population
80	Björn & Leppänen (2013)	English as Foreign language
81	Regtvoort, Zijlstra & Leij (2013)	Not English-speaking population
82	Ritter, Park, Saxon & Colson (2013)	Language impairment
83	Duff, Hulme, Grainger, Hardwick, Miles & Snowling	One-to-one instruction
84	McCartney, Boyle & Ellis (2015)	Language impairment
85	Helland, Tjus, Hovden, Ofte, Heimann (2011)	Not English-speaking population
86	Steady, Ellemann, Lovett & Compton (2016)	Students from Grade 3 to 6
87	Tilanus, Segers & Verhoeven (2016)	Not English-speaking population
88	González-Valenzuela & Martín-Ruiz (2017)	Not English-speaking population
89	Rosas, Escobar, Ramírez, Meneses & Guajardo (2017)	Not English-speaking population
90	Christodoulou, Cyr, Murtagh, Chang, Lin, Guarino, A. J. & Gabrieli (2017)	6 weeks; Participants attended the program 5 days per week for 4 hr each day, totaling 100 to 120 hr of instruction.

91	Wanzek, Petscher, Otaiba, Rivas, Jones, Kent, S. C. & Mehta (2017)	Fourth-grade students
92	Müller, Richter, Karageorgos, Krawietz & Ennemoser (2017)	Not English-speaking population
93	Mayseless (2011)	university students
94	Lane, Pullen, Hudson & Konold (2009)	One-on-one tutoring
95	O'Connor, Bocian, Beebe-Frankenberger & Linklater (2010)	English as Foreign language
96	Struiksmā, van der Leij & Stoel (2009)	Not English-speaking population
97	Duff, Fieldsend, Bowyer-Crane, Hulme, Smith, Gibbs & Snowling (2008)	9-week reading intervention incorporating reading, phonological and vocabulary training was implemented
98	Silva & Crenitte (2016)	Article not in English Language.
99	Lovett, De Palma, Frijters, Steinbach, Temple, Benson & Lacerenza (2008)	Wide age range in the intervention
Source 5 Pubmed		
100	O'callaghan, McIvor, McVeigh & Rushe (2016)	Intervention lasted 8 weeks.
101	Poulsen, Nielsen, Juul & Elbro (2017)	Not English-speaking population
102	Burchfield, Hua, Noyes & Waal (2017)	Not English-speaking population
103	Ardoin, Binder, Foster & Zawoyski (2016)	Children did not have reading difficulties.
104	Barth & Elleman (2017)	Students from Grades 6–8
105	Roberts, Rane, Fall, Denton, Fletcher & Vaughn (2015)	fifth-grade students at beginning of intervention
106	Wolter & Dilworth (2014)	Intervention did not last for 10 weeks.
107	Kim, Apel & Al Otaiba (2013)	No intervention was specifically provided; intervention occurred within tiers in the system
108	Connor, Morrison, Fishman, Crowe, Al Otaiba & Schatschneider (2013)	Intervention was not targeted at children with reading difficulties.
109	Cogo-Moreira, de Avila, Ploubidis & de Jesus Mari (2013)	Not English-speaking population
110	Decker & Buggley (2014)	Intervention only lasted 2 weeks.
111	Hulme, Bowyer-Crane, Carroll, Duff & Snowling (2012)	Students recruited on the basis of their poor verbal ability not poor reading
112	Compton, Gilbert, Jenkins, Fuchs, Fuchs, Cho, & Bouton (2012)	No intervention was conducted; authors analysed the data available to explore how to identify potential students requiring Tier 3 support
113	Allen, Ukrainetz & Carswell (2012)	No intervention was conducted.
114	Saine, Lerkkanen, Ahonen, Tolvanen & Lyytinen (2011)	Not English-speaking population
115	Frijters, Lovett, Steinbach, Wolf, Sevcik & Morris (2011)	In this study, the authors reanalyzed data from Morris et al (2012) using a mediation analysis.
116	Gustafson, Fälth, Svensson, Tjus & Heimann (2011)	Not English-speaking population
117	Luft Baker, Park & Baker (2010)	No intervention was conducted
118	O'connor, Arnott, McIntosh & Dodd (2009)	children received intervention 2 years ago in preschool; whole-class intervention
119	Bunn (2008)	Different literacy interventions were used in the schools.

Source 6 Additional records through other sources		
120	Denton, Solari, Ciancio, Hecht & Swank (2010)	Intervention did not last for 10 weeks.
121	Nielsen & Friesen (2012)	Students recruited on the basis of their poor language and narrative ability not poor reading
122	Puhalla (2011)	Intervention involved only 5 sessions.
123	Zvoch & Stevens (2013)	intervention lasted for 5 weeks.
124	Lee, Morrow-Howell, Jonson-Reid & McCrary (2011)	One-to-one tutoring
125	Vadasy & Sanders (2008a)	One-to-one tutoring. Study did not investigate the effects of no tutoring group vs dyads separately.
126	Vadasy & Sanders (2011)	One-to-one intervention
127	Vernon-Feagans, Kainz, Ginsberg, Wood & Bock (2012)	One-to-one tutoring
128	Gilbert, Compton, Fuchs, Fuchs, Bouton, Barquero & Cho (2013)	Unable to isolate the interventions that did not involve one-to-one tutoring