"An investment in knowledge pays the best interest."

—Benjamin Franklin

Exploring the Relationships Between Executive Functions and Reading Intervention Responsiveness in Children with Reading Disability

Benjamin Johnson, B.A.

Child and Youth Studies

Submitted in partial fulfillment of the requirements for the degree of

Master of Arts

Faculty of Social Sciences, Brock University
St. Catharines, Ontario

Abstract

Background: Notwithstanding the emergence of rigorous, evidence-based interventions, some students with reading disability (RD) still fall behind in reading skill. Copious amounts of evidence support a link between children's executive functions (EF) and learning, including acquisition of reading skill specifically. Nonetheless, the relationship between EF and reading intervention outcomes have seldom been studied.

Objectives: The purpose was to understand the extent to which EF predict reading intervention outcomes for students with RD.

Methods: In the present study, 115 children with RD in Grade 3 and 4 completed an intensive, evidence-based 70-hour reading intervention for one hour per day over the course of fourteen weeks. EF in both neuropsychological (e.g., D-KEFS) and behavioural (e.g., BRIEF) measurements were observed at pretest, categorized as inhibition, shifting, working memory, emotional control, and reasoning/metacognition. Reading skills in the form of word recognition, decoding, fluency, and passage comprehension were measured before and after intervention. Standard scores were computed for EF and reading measures. Reading gains were converted into reliable change scores using the reliable change index (RCI).

Results: Approximately 27% of correlations between EF and pretest reading were statistically significant. Analyzing reading gains, most correlations with EF were non-significant. The few significant coefficients lost significance following Bonferroni correction for multiple comparisons. Next, a composite EF deficit category was generated to differentiate participants who did or did not have below average EF across multiple measures, which 29 participants met and 86 did not. While *t*-tests revealed that the multi-deficit group had significantly lower gains in fluency than the group without multiple deficits (*p*s<.05), differences were not significant following Bonferroni correction. Finally, a chi-square analysis using the multi-deficit grouping and categorical reliable change in reading did not yield any significant results.

Implications: While children's EF relates to reading skill before the reading intervention, EF does not relate to gains in reading skill from the beginning to the end of intervention. This finding suggests that small group, structured interventions can mitigate learning barriers posed by executive dysfunction. However, the patterns of relationships between EF and reading fluency suggest that more research is needed.

Keywords: executive functions, reading disability, reading intervention, reliable change, response to intervention.

Acknowledgements

Ben's thesis is supported in part by funding from the Social Sciences and Humanities Research Council (SSHRC).

La thèse de Ben est soutenue en partie par un financement du Conseil de recherches en sciences humaines du Canada (CRSH).

Social Sciences and Humanities Research Council of Canada
Conseil de recherches en sciences humaines du Canada

Canada

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ADHD	Attention-deficit/hyperactivity disorder
BRIEF	Behavioural Rating Inventory of Executive Functioning
CS1	Free Sorting, Card Set 1 (from D-KEFS)
CS2	Free Sorting, Card Set 2 (from D-KEFS)
CWI	Colour-Word Interference (from D-KEFS)
CWI-I	Colour-Word Interference, Inhibition subtest
CWI-S	Colour-Word Interference, Shifting subtest
CWT	Challenge Word Test
D-KEFS	Delis-Kaplan Executive Function System
EF	Executive functions
LW	Letter-Word Identification (from WJ III)
MTSS	Multi-tier systems of support
PC	Passage Comprehension
PDE	Phonemic Decoding Efficiency (from TOWRE)
PHAST	Phonological and Strategy Training
RC	Reading comprehension
RD	Reading disability
RF	Reading Fluency (from WJ III)
RTI	Response to intervention
SRI-2	Standardized Reading Inventory, 2nd Edition
SWAN	Strengths and Weaknesses of ADHD-Symptoms and Normal-Behaviour Scale
SWE	Sight Word Efficiency (from TOWRE)
TM	Trail-Making (from D-KEFS)
TOWRE	Test of Word Reading Efficiency
WA	Word Attack (from WJ III)
WASI	Weschler’s Abbreviated Scale of Intelligence
WJ III	Woodcock-Johnson III Tests of Academic Achievement
WRA	Word Recognition Accuracy (from SRI-2)

Introduction

Learning how to read is a fundamental process in a child's educational journey, preparing them to become literate members of society. Imagine you are a teacher of a Grade 3 class with 20 students, four of whom are identified as substantially struggling with reading. Throughout a couple of months, those four students took part in an intensive evidence-based reading intervention, after which one student did not make noticeable gains in reading skill as they are still distinctively error-prone when reading aloud words with multiple consonants in a row, such as the word *strong*, compared to the other students. As an educator, you do your best to provide them with opportunities to continue practicing their reading, but there is only so much that you can do, and you are concerned because of how much time remediation is taking away from other subjects.

This enigma is where educational research comes into the picture. No educator or caregiver wants a child to be in this situation. When left untreated, reading disability (RD) predicts above-average rates of mental illness, trouble with the criminal justice system, and other negative life outcomes (Willcutt et al., 2007). Fortunately, the science of reading has informed numerous best practices, including tiered systems of support and emphasis on rigorously researched evidence-based instruction in the classroom (Lovett et al., 2014; Swanson et al., 2017). Despite drastic improvement in the scientific understanding of acquiring reading skill, the aforementioned situation of lack of intervention responsiveness still occurs (Lovett et al., 2017; Wanzek et al., 2016). Delays in the acquisition of reading skill are problematic, because after Grade 2, it becomes increasingly difficult for struggling readers to catch up to their peers, which can have lasting impacts on their academic trajectories and other outcomes (Swanson et al., 2017). Tiered intervention systems, such as response to intervention (RTI), involve early screening and small group interventions (Fletcher & Vaughn, 2009). Identifying predictors of intervention responsiveness could enable greater supports earlier on, maximizing remediation.

Students with RD often exhibit difficulties with executive functions (EF; Daucourt et al., 2018; Horowitz-Krauss, 2023; Soares et al., 2020). EF are goal-directed processes such as filtering distractions, switching from one task to the next, and problem-solving (Barnes, 2023; Miyake & Friedman, 2012). EF are often strong predictors of learning (Center on the Developing Child at Harvard University, 2011; Serpell & Esposito, 2016; Yeniad et al., 2013). Nevertheless, few studies have examined reading intervention responsiveness using EF as predictors, and even fewer have used intensive programs designed specifically for students with RD. Building on previous research on predictors of intervention responsiveness, the present study offers an opportunity to examine the capacity of EF to predict reading gains made in small group evidence-based intervention.

Pause and Ponder

Do you remember learning how to read?
What strategies were most effective?

Literature Review

Reading skill, as with other language skills, involves assembling multiple pieces of information together, like assembling a puzzle. Fortunately, there is an incredibly vast foundation of existing research upon which one can build a study to investigate the aforementioned problem of what supports are needed by students who struggle with reading. This section will canvass the scholarly literature of the core topics within the present study: how reading skill is learned, why students with learning disabilities (LD) struggle with developing reading skill, the role of EF in reading, and the relationship between EF and reading intervention. To understand the final topic, the question of how EF relates to reading intervention, a structured review was undertaken to analyze key information about existing evidence to determine patterns and gaps in order to set the stage for the present study.

Development of Reading Skill

Experts far and wide agree that a foundational component of reading skill is phonological awareness, which comprises the ability to understand the sounds that are associated with language (Al Otaiba & Fuchs, 2002; Hogan et al., 2005; Melby-Lervåg et al., 2012). The most prominent component of phonological awareness is phonemic awareness, the knowledge of and ability to manipulate individual phonemes, which are the most basic units of sound within a language (for a review, see Chapman, 2003). There are 44 distinct phonemes in the English language. The word *cake* is comprised of three phonemes: /k/ /eɪ/ /k/. In fact, in this example, the first and third phonemes are identical (/k/) despite being represented by different letters ("c" and "k"). There are other components of phonological awareness, such as understanding syllables and rhyming, but phonemic awareness is the most fundamental. The provision of phonemic awareness training typically reaps a multitude of benefits for beginning readers (Lovett et al., 2000a; Lovett et al., 2000b; for an earlier study, see Wagner & Torgesen, 1987). Phonemic awareness for a beginning reader would be equivalent to a beginning musician learning the distinct sound of individual types of musical notes: it is the most basic and raw unit used to build more complex ideas.

You may remember learning how to read in kindergarten and spending a lot of time on learning letters of the alphabet. While visual letter recognition and orthography (i.e., spelling) are important, reading skill also necessitates the integration of how letters, both individually and in groups, are associated with specific sounds, known as letter-sound knowledge, which largely draws on students' phonemic awareness (Hulme & Snowling, 2012). Most letters, especially vowels, can represent multiple phonemes depending on the word. For example, the letter "a" does not make the same sound in the word *cat* as in the word *cake* – one is short, and the other is long, thus they are distinct phonemes. Yet the phoneme represented by "a" in the word *cake* is the same phoneme made by the letters "ey" in the word *prey* despite being represented by different letters.[1] Multiple letters may combine with each other to make unique phonemes. Two letters combining into one sound is called a digraph, such as when the letters "t" and "h" often come together to form two unique phoneme as in the word *thick* and the slightly different "th" sound in *then*. Three letters combining into one sound is a trigraph, such as the "tch" in *catch*. Evidently, there is a myriad of language rules that children must learn to become literate members of society. The more one practices, the more effectively and automatically one can apply phonological awareness to decode new words (Ardoin et al., 2016; Ardoin et al., 2018).

Within research and clinical practice, phonological awareness tests are an effective predictor of future reading skill and reading disability, especially in kindergarten (Hogan et al., 2005; Hulme & Snowling, 2012; Lyon et al., 2003; Mercugliano et al., 2024). One research instrument to determine phonological awareness is the Comprehensive Tests of Phonological Processing (CTOPP; Wagner et al. 1999). Key components of this test are Phoneme Segmentation and Blending subtests, which involve separating and combining phonemes respectively. The Phoneme Segmentation subtest asks participants to break down a word spoken by the test administrator into its individual phonemes in the proper order. For example, if the test administrator says the word *my*, the participant would break it down into the phonemes of "m"

---

[1] Another phonemic match is with "eh" as we Canadians say.

and “y” separately. In the Blending subtest, participants are asked to combine a series of phonemes enunciated by the test administrator to form a whole word – essentially the reverse of the Phoneme Segmentation subtest.

For decades, researchers have provided consistent evidence for the substantial role of phonics instruction in cultivating literacy skills (Al Otaiba & Fuchs, 2002; Hogan et al., 2005; Lovett et al., 2000b; Melby-Lervåg et al., 2012; for earlier studies, see also Foorman et al., 1998; Wagner & Torgesen, 1987). Since then, researchers have expanded on this by identifying the most effective ways of engaging students in activities that help them practice these skills, such as comparing words with similar phonemic composition (e.g., *dog* and *fog*). These activities practice phonological awareness by addressing various ways phonemes are manipulated in speech. The most basic strategy that students learn is to sound out words, where students breakdown a word into each constituent phoneme. Building upon phonemic awareness are two similar reading skills of word recognition and vocabulary. Word recognition takes places when children can automatically recall the phonemic composition of a given word when reading the without needing to decode it in smaller pieces. Vocabulary refers to words of which students have a semantic understanding – that is, they know the meaning of the word, not solely how it is pronounced. While a child may possess various words in their vocabulary acquired through oral communication, that does not necessarily mean that they know how to spell or recognize those words in written form. Over time, students automatically recognize words they practice, which becomes reading fluency. In research and clinical practice, single-word decoding fluency is often measured by capturing how many words one can recognize through timed assessment, such as the Test of Word Reading Efficiency (TOWRE; Torgesen et al., 2012). TOWRE has subtests for both Word Recognition Accuracy (WRA), which consists of real words, and Phonemic Decoding

Efficiency (PDE), which involves smoothly decoding nonwords (e.g., “efarir”). Overall, phonemic awareness is only a stepping stone to reading words fluently (Compton et al., 2014).

Despite a core “code” for the associations between letters into sounds, there are nuances and irregularities to letter-sound associations in the English language which make it difficult to rely solely on a bottom-up, letter-for-letter approach to reading new words. For example, when “t” is followed by “h,” it does do not produce the same phoneme every time. Sometimes they represent their original separation of the hard “t” and the sound of “h.” The word *pothole* is an example of when “t” and “h” are adjacent, but they do not produce the same phoneme as in the word *thick*. Many words that are adopted from other languages may also deviate from the typical letter-sound ruleset. For example, the words *pizza* and *mozzarella*, adopted from the Italian language, use unconventional associations for the letters “zz” which make the sound /ts/ instead of the typical /z/ sound like in the word *zebra*. Therefore, one has to also memorize an assortment of words that are exceptions to the default convention, as well as strategies to recognize those patterns in new words. In fact, some scholars suggest that English has greater number of irregularities to letter-sound patterns compared to other languages such as Spanish and German (Melby-Lervåg et al., 2012).

Aside from sounding out words, there are additional decoding strategies that assist beginning readers by identifying the structure (morphology) and meaning (semantics) of words. Morphemes are the smallest units of meaning within a word. Understanding morphology has been shown to play an important role in understanding text (Levesque et al., 2019). One decoding strategy is to recognize familiar morphemes hidden in larger words (a strategy sometimes called “spy”; Lovett et al., 2000a). For instance, once a student learns the word *cake*, they can more easily recognize the compound word *cupcake*. Not all morphemes are full words, however. Bound morphemes such as prefixes and suffixes are common letter patterns that children may recognize from other words. An existing decoding strategy is to “peel off” prefixes or suffixes to reveal the root word, making the word recognition process easier. An example is separating the “pre” in the word *preview* to then identify the root word *view*. Further, the “pre” morpheme carries meaning by signifying something taking place beforehand (such as a preview being advertised before viewing the full movie). Just as phoneme rules have exceptions, morphemes are also dependent on the word of interest; consider the word *press* which contains “pre” but here it does not act as a morpheme denoting that something is occurring beforehand – the word meaning is completely different. To advanced readers, these rules and exceptions are self-evident and processed unconsciously. Yet taking a closer look, decoding skills require emerging readers to learn a variety of phonological, morphological, and semantic attributes of words that must be learned for children to read words accurately and efficiently.

Beyond single word recognition, the goal of reading is for students to learn to read and comprehend full texts. Analyzing multiple dimensions of a passage, students must draw from vocabulary, background knowledge, language conventions (e.g., grammar and punctuation), literacy knowledge (e.g., genres), and verbal reasoning (e.g., inference and metaphor) to create an effective mental representation of the described situation (Scarborough, 2001). While word recognition and vocabulary are important precursors to reading comprehension (Denton et al., 2022; García & Cain, 2014), there are reading comprehension strategies that are explicitly taught, as it charts new territory beyond the phonology and semantics of individual words as targeted by the decoding and vocabulary components of reading interventions (Silverman et al., 2020). In sum, the diversity of skills involved in the reading process, in addition to the assortment of deviations to language conventions, contribute to the perplexity of fostering literacy skills among emerging readers.

Reading Disability

Reading disability (RD) is a type of learning disability where students struggle with one or more of the reading skills described above. Remediating RD is a serious matter not only because of the axiomatic importance of a literate society, but more specifically, numerous studies have found that reading difficulties predict various important life outcomes such as poorer academic and career success (Eloranta et al., 2019; Gibson & Kendall, 2010; Hulme & Snowling, 2016; Swanson et al., 2017; Willcutt et al., 2007).

RD versus Dyslexia

While RD encompasses all students who fail to read at their expected developmental level, dyslexia refers to children with a genetic predisposition to phonological awareness deficiencies, which excludes those who have these deficiencies exclusively due to environmental factors such as poor instruction. The board of directors of the International Dyslexia Association (2002) adopted the following formal definition:

> Dyslexia is a specific learning disability that is neurobiological in origin. It is characterized by difficulties with accurate and/or fluent word recognition and by poor spelling and decoding abilities. These difficulties typically result from a deficit in the phonological component of language that is often unexpected in relation to other cognitive abilities and the provision of effective classroom instruction. Secondary consequences may include problems in reading comprehension and reduced reading experience that can impede growth of vocabulary and background knowledge.

This definition of dyslexia is beneficial in understanding this particular risk factor for difficulties with phonological awareness and reading. Nevertheless, such difficulties are influenced by many other factors, both environmental and genetic (Christopher et al., 2015). Many with dyslexia go on to have RD, but many with RD do not have dyslexia. In the latter case, functional impairment in reading derives from environmental circumstances, namely poor reading instruction, poor engagement in school (e.g., due to behavioural issues), lack of access to reading support at home, or a mixture thereof (Al Otaiba & Fuchs, 2002). Recognizing the prominence of how inadequate instruction can impact reading skill, scholars have posited that, in these cases, students are not reading disabled, but rather "instructionally disabled" (Vellutino et al., 1996). Reading interventions should target all students who need help. Therefore, in the context of reading interventions, a broader definition of RD that focuses on reading difficulties irrespective of origin is warranted.

The disability dimension to this issue is important to understand in context. Traditionally, the education system has taken an approach where a reading disability diagnosis is not officially given until it can be determined that there is a discrepancy between the student's reading ability and their other cognitive abilities such as intelligence (Fletcher et al., 2018). The implications of this system were that students were not provided with special education services such as interventions to reading difficulties until later years, when remediation was more difficult for the developing brain (Perdue et al., 2022). Multi-tiered systems of support (MTSS), response to intervention (RTI), and other educational strategies have sought to use early screening for risk of RD to then prompt preventative interventions that cultivate core reading skills before RD develops (Coyne et al., 2016). Nonetheless, the question of when does reading difficulty turn into a disability is still often split into two camps, one of which evaluates performance relative to other variables, such as intelligence, and the other evaluating reading skill using absolute performance based on a fixed criterion (Wagner et al., 2020). The absolute performance perspective is to evaluate whether a student falls substantially below average, typically less than 85 on standardized tests where the mean is 100 and standard deviation is 15. This simple approach does not depend on any other variables, making it efficient. From clinical and educational standpoints, this measurement approach is beneficial in determining the important decision as to whether a student requires further intervention. The approach of the other camp commonly involves calculating the discrepancy between intelligence and reading achievement, often called the IQ-achievement discrepancy model (Fletcher et al., 2018). From a research standpoint, such measurements are important in order to ascertain associations between progress in reading and other cognitive factors (Fletcher et al., 2018; Frijters et al., 2013) as well as understand why some students are unexpectedly poor readers.

Predictors of RD and Intervention Responsiveness

Previous research has uncovered that children with RD struggle with the following early reading skills: phonological awareness, decoding, word identification, fluency, and comprehension (Foorman et al., 1998; Lovett et al., 2000b; Lovett et al., 2014). By targeting these skills, schools can remediate and even prevent substantial reading difficulties in their students, which will later support their reading fluency and comprehension skills when they get older (Eloranta et al., 2019). Foorman et al. (1998) found that explicit letter-sound instruction can prevent reading failure in the first and second grade. Using this approach to reading instruction, researchers and educators have developed various interventions that address the crucial early reading skills of phonological awareness, decoding, and word identification. One example is the PHAST (Phonological and Strategy Training) program[2] that focuses on explicit instruction in letter-sound knowledge and word reading strategies in order to improve early literacy skills (Lovett et al., 2000a; Lovett et al., 2013; Morris et al., 2012). For example, PHAST participants might practice the “peeling off” strategy by breaking down words with prefixes and suffixes such as *empowering* in order to help figure out how the word sounds and what it means (i.e., “em,” “power,” and “ing”). Like having multiple tools in a toolbox, students learn various strategies in combination with other strategies such as rhyming strategy. For example, in *empowering*, the part “pow” is spelled like “cow,” a word they may already know how to say. Lovett et al. (2017) studied the delivery of a hybrid version of PHAST combined with another intervention in 219 students with RD, finding that two thirds of Grade 1 students who completed the intervention no longer met criteria for RD.

Although interventions have consistently proven to be highly effective, research has investigated certain factors that are associated with variation in responsiveness to reading interventions (e.g., Al Otaiba & Fuchs, 2006; Frijters et al., 2011). Here, a key problem to resolve is why not all students make significant reading gains through intervention. Understanding risk factors for low responsiveness may help deliver individualized interventions or otherwise supplement them. Compton et al. (2014) raised concern around the pattern of many reading interventions focusing on word decoding with minimal emphasis on making gains in reading fluency and comprehension as skills crucial for everyday tasks involving reading.

Scammacca et al. (2015) conducted a meta-analysis of studies on reading comprehension interventions for students in Grade 4-12 from 1980 to 2011, finding that effect sizes of these interventions in recent years have shrunk compared to further in the past (e.g., from 0.42 to 0.21

---

[2] This intervention has since been developed into a commercially available program as “Empower Reading.”

on standard testing), but that this pattern is likely explained by the rising effectiveness of general education (e.g., proliferation of RTI strategies) and increasing use of standardized testing rather than weaker intervention potency. They noted how recent interventions have increased in rigour and complexity. However, this meta-analysis highlights the need for continual improvement of these interventions due to the unignorable rates of students who fall short of grade level standards in reading comprehension skill. For example, in Ontario, 2023 results on standard testing reveal that 29% of Grade 3 students are below the standard provincial reading level.[3]

Traits of the students receiving an intervention are known to affect the degree of reading gains made therein (Stuebing et al., 2015). One crucial discovery identified in numerous studies (e.g., Wanzek et al., 2018; Lovett et al., 2014; Lovett et al., 2017) is that the younger the age when intervention occurs, the greater the long-term gains in reading ability. Given how a large number of studies demonstrated the effect of age at intervention, a meta-analysis by Wanzek et al. (2018; an update from the 2007 meta-analysis by Wanzek & Vaughn) showed that intervention studies have found larger effect sizes (i.e., larger impact of interventions) when the reading intervention is completed in kindergarten or Grade 1 compared to Grade 2 or 3. In this meta-analysis, 25 studies were included, and the sample criterion of students with substantial reading difficulties in kindergarten to Grade 3, which closely resembles the sample used for the present study. Concerning pre-intervention reading ability, Tran et al. (2011) conducted a meta-analysis using 13 studies situated in an RTI framework and found that higher reading ability prior to the intervention predicts higher reading gains. Compared to non-responders, responders had substantially higher standardized pretest scores in receptive vocabulary, comprehension, word attack, phonological awareness, and real word reading.

---

[3] According to the Education Quality and Accountability Office (EQAO).

Notably, there is a growing body of research on the cognitive predictors of reading intervention outcomes. Frijters et al. (2011) studied the prediction of intervention responsiveness using a sample of 278 students with RD, classifying them as poor, average, or good responders using a growth curve model of reading outcomes such as word recognition, fluency, and word attack (i.e., nonword decoding). They found that visual processing, phonological memory, and other cognitive abilities were identified as predicting which students respond well to intensive reading intervention. In particular, this predictive capacity persisted above and beyond common predictors such as phonological awareness and rapid naming. Nelson et al. (2003) conducted a meta-analysis with 30 studies, identifying a pattern where the primary learner characteristics most influential on the intervention responsiveness were, in order of magnitude, rapid naming, problem behaviour, phonological awareness, alphabetic principle, memory, IQ, and demographics. There are various non-reading factors whose association with RD have been explored in multiple studies. Lam and McMaster (2014) conducted a review and found that, while core skills like phonological awareness consistently predicted RD symptoms, intelligence and memory capacity were inconsistent predictors. Vocabulary, language, and demographics were not observed as useful predictors. This growing body of research helps researchers and educators better understand the cognitive profiles of students resistant to intervention to potentially offer additional supports tailored to their learning needs. The consistent evidence suggests that supports should not place excessive demand on memory, and should account for factors such as behavioural issues.

Another important observation is that RD is linked to domain-general attentional issues. For example, RD is often comorbid with attention-deficit/hyperactivity disorder (ADHD; American Psychiatric Association, 2022), estimated around 30-50%, drastically more than the typical prevalence rates for either disorder (Willcutt et al., 2007). Further, it is beneficial to consider attentional problems beyond a clinical outlook. Many studies have found that RD habitually involves subclinical deficits in attention (Lovett et al., 2017; Walda et al., 2022). Rabiner et al. (2004) demonstrated that attention problems identified in kindergarten predicted poor reading skill in Grade 1. Furthermore, they provided an intervention in Grade 1 and found that it was beneficial for students with attention problems but without RD as well as those with RD but no attention problems; however, it was ineffective for those with both RD and attention problems. These results show that not only are children with RD more likely than children without RD to have attention regulation problems unrelated to reading, but these problems have been shown to be relevant to the implementation of interventions.

Translating Evidence-Based Instruction to Practice

Given the importance of reading to personal and academic success, it is important for the plethora of evidence on reading interventions to be maximally leveraged by implementing them in schools, which has been a historical challenge. Ontario, Canada is an example where the implementation of evidence-based reading instruction has been flawed. In 2022, the Ontario Human Rights Commission released the results of the *Right to Read Inquiry*, concluding that the province’s public education system was discriminatory toward students with RD by substantially neglecting decades of research that produced evidence for efficacious methods of reading instruction that would have either prevented or reduced the severity of RD in the vast majority of students. While evidence-based interventions have been put into practice for students who struggled with reading, a key issue was that evidence-based instruction was the not the default instructional method for all students.

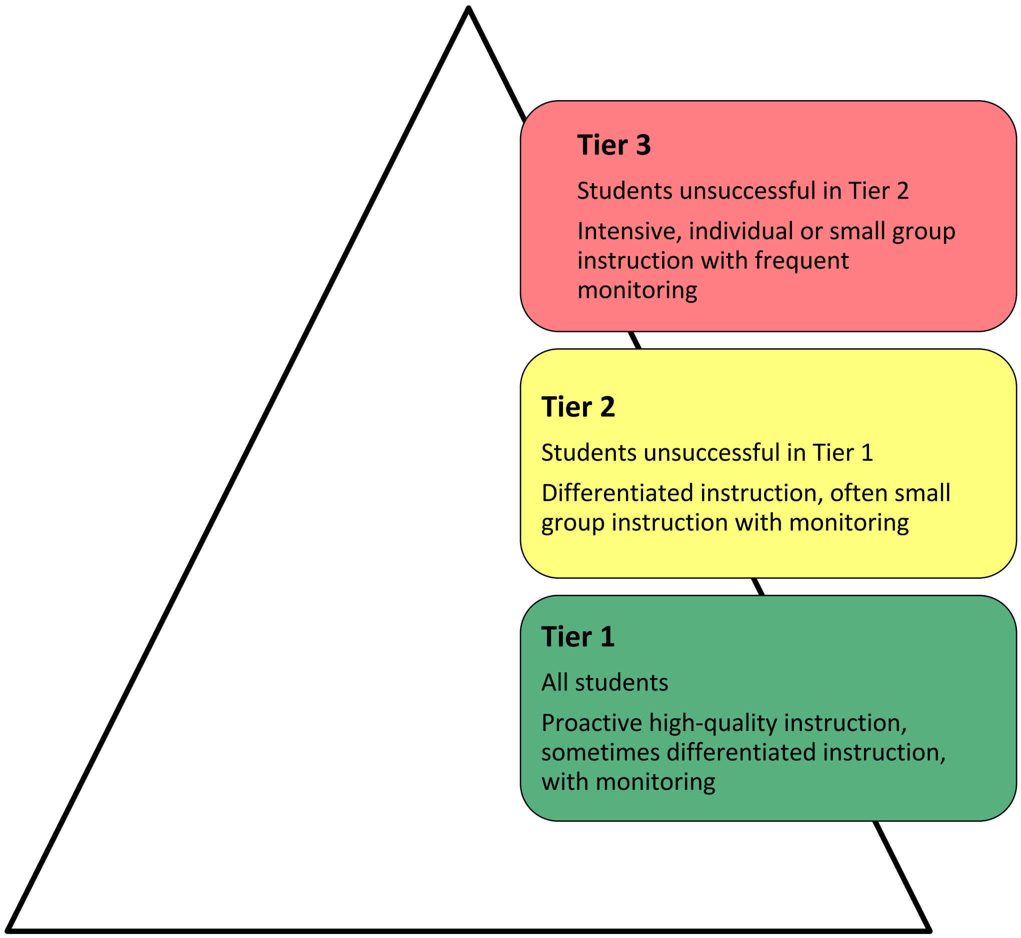
Tiered Intervention

As highlighted by the Right to Read Inquiry (Ontario Human Rights Commission, 2022), while delivering an effective intervention for struggling students is important, an arguably more important endeavour is to introduce preventative, evidence-based instruction for all students as the default approach, which educational policymakers have not always prioritized. Throughout the late twentieth century, as research produced more and more robust academic interventions, educators and researchers encountered perplexing questions regarding how to implement these evidence-based interventions in an optimal way (Fletcher et al., 2018). On one hand, intensive interventions delivered separate from the main classroom often proved more effective in cultivating academic skills like reading and mathematics compared to integrating special education support within the general education classroom without any separate intervention (Burns et al., 2005). On the other hand, pulling students out of the main classroom had its own repercussions, such as social ramifications based on separation from peers. Arising from this conflict was the concept of MTSS where schools deliver multiple tiers of intervention of increasing intensity. Quality instruction and monitoring are prioritized from the start, and the higher tiers are reserved for students only when necessary (i.e., when a student is unsuccessful in the previous tier).

Response to intervention (RTI) is a widely adopted MTSS framework that delivers comprehensive support to students struggling with learning difficulties, whether or not these difficulties are disability related (Fletcher & Vaughn, 2009). Just like MTSS, inspiration for RTI arose in part to improve literacy instruction given its importance within education and beyond. One prominent mechanism is the prioritization of quality, evidence-based teaching being implemented at the beginning with all students such that the reason any students require subsequent intervention would not be due to the preventable circumstance of inadequate teaching. Preventing academic difficulties lessens the burden on special education resources which can then be concentrated on students who are resistant to initial instruction, requiring differentiated instruction and intensive intervention. To accomplish this, RTI identifies three tiers of intervention (Fletcher & Vaughn, 2009; Nilvius & Svensson, 2022). Tier 1 is initial evidence-based classroom instruction for all students and includes initial monitoring of progress (i.e., early screening). Those who continue to struggle are referred to Tier 2, which is differentiated instruction, sometimes in the form of small group intervention. Students who fail to show significant gains after this stage are referred to Tier 3, which continues differentiated instruction, but often this is delivered as individualized intervention. Identifying factors that predict intervention responsiveness within small groups could allow educators to bypass the small group stage when necessary, skipping directly to individualized instruction and maximizing the resources available to help students attain reading skill within the optimal time period in kindergarten and Grade 1.

Figure 1

The Three Tiers of the Response to Intervention Framework



Note. Adapted from Nilvius & Svensson (2022).

Executive Functions

One factor that is less understood in the literature on reading intervention responsiveness is executive functions (EF), referring to cognitive abilities that drive goal-oriented behaviours, such as suppressing distractions and switching from one task to the next (Anderson, 2002; Friedman & Miyake, 2004). The prefrontal cortex and basal ganglia are the brain regions primarily responsible for EF (Monchi et al., 2006). EF as well as similar domain-general cognitive faculties such as self-regulation have largely been an inherent factor in educational settings, which is why they are highly useful in educational research (Gerst et al., 2017; Miyake & Friedman, 2012). EF is conceptualized in multiple paradigms (Packwood et al., 2011). One such conceptualization is the unitary model, which posits that various functions such as updating information in working memory and planning are all controlled by a single system, the "central executive" (e.g., Niendam et al., 2012). Alternatively, many scholars adopt the diversity model envisioning EF as a set of multiple but related functions (Centre on the Developing Child at Harvard University, 2011). EF has been found to be associated with intelligence (Brydges et al., 2012), although the relationship is viewed to be weak as there are many children who exhibit high intelligence and executive dysfunction concurrently (American Psychological Association, 2022; for a depiction of how EF is considered a possible area impaired by specific learning disabilities, see Learning Disability Association of Canada, 2015).

Despite originating as a purely cognitive faculty, EF has been conceptualized in scenarios with involving other psychological phenomena such as affect. For example, the Behaviour Rating Inventory of Executive Function (BRIEF; Gioia et al., 2000; Baron, 2000) includes a dimension called "Emotional Control," referring to one's self-control capacity in regulating emotional reactions. Recognizing the distinction between scenarios in which emotions are involved in EF, some researchers refer to emotion-infused EF as "hot EF" and their non-emotion-infused counterparts "cool EF." A study conducted in Hong Kong by Poon (2018) found that hot EF correlated among one another, cool EF correlated among one another, but the hot categories did not correlate with the cool ones. The diversity in EF conceptualizations can be imagined lending the poetic style of Dr. Seuss:

One function, three functions,

Hot function, cool function

Some functions are high, some functions are low

Some things about these functions we just don't know (yet)!

In the context of children's cognitive development, the principal EF paradigm follows the diversity model and encompasses three 'archetypal' functions: inhibition, shifting, and working memory (e.g., DeBruin-Parecki & Cartwright, 2023; Centre on the Developing Child at Harvard University, 2011). Inhibition refers to the ability to suppress automatic impulses and is typically assessed where a participant completes an otherwise automatic task in a different way, requiring exertion of effortful control. Shifting refers to the flexibility of switching and adapting to different rules, tasks, or situations. Working memory refers to the capacity to manipulate information in short-term memory. The literature often refers to two main components of working memory, a visual component, known as the visuo-spatial sketchpad, and an auditory component, known as the phonological loop. In tandem, these functions help children complete daily tasks in school and beyond. The Centre on the Developing Child at Harvard University (2011) shares a brilliant metaphor for this mainstream EF triad, maintaining "just as an air traffic control system requires the interaction of multiple people—pilots, navigators, controllers, weather forecasters—our human executive functioning system requires that each type of skill utilize elements of the others" (p. 9). This symbolism portrays how students' brains are constantly processing a breadth of information at any given time, meaning that a system of controlling information and motor response flow is necessary for goal-driven behaviours, including learning in school.

Notably, students with RD tend to struggle with EF (Horowitz-Krauss, 2023; Soares et al., 2020; see also Altemeier, Abbot, & Berninger, 2008). Furthermore, RD is highly comorbid with ADHD, a condition for which EF impairment is a key diagnostic marker (American Psychiatric Association, 2022). Daucourt et al. (2018) found that the core executive functions of inhibition, working memory, and shifting were able to predict RD symptoms in a sample consisting of children in kindergarten through Grade 3. Moreover, Walda et al. (2022) found that all participants in the RD group failed a sustained attention task despite the fact that it did not involve any kind of reading of words or letters. Additionally, the majority of children without RD did not fail the attention task. Furthermore, children with RD did not make significant improvements on the attention task when given simple sustained attention training. Collectively, these studies suggest issues in EF and attention appear to be related to impaired development of reading skill. Notwithstanding the similarity of symptomology in ADHD and RD, there are limits on the extent to which these two neurodevelopmental disorders overlap (Fletcher et al., 1999; American Psychiatric Association, 2022). Studies have found that EF does not always correlate with RD (e.g., Doyle et al., 2018). Nevertheless, EF remains an important piece of the puzzle in seeking to understand what cognitive factors can predict gains in reading skill for children with RD.

Executive Functions and Reading

The pervasive nature of EF manifests in several ways within students' academic lives. Thus, there is an abundance of empirical evidence that EF plays a role in academic performance (Barnes, 2023; Gerst et al., 2017) and reading performance more specifically (Arrington et al., 2019; Carretti et al., 2009; Cartwright et al., 2017; Church et al., 2019; Norton & Wolf, 2012). A meta-analysis by Yeniad et al. (2013) found that EF predicted reading performance across 25 different studies. The relationship between EF and reading skill has been examined extensively in both RD and typically developing populations (for a systematic review, see Follmer, 2018).

One important nuance to EF is the distinction of domain-general and domain-specific cognitive abilities. While the typical EF paradigm is domain-general, some researchers have looked at reading-specific EF (e.g., Gosselin & Sabourin, 2024). Vellutino et al. (1996) found that phonological-specific coding deficits were seen to impact working memory capacity during reading. In other words, for children with phonological coding deficits, their working memory is taxed greater during the reading process, revealing a reciprocal nexus between reading and EF capacities. Additionally, Cartwright et al. (2017) compared whether reading-specific (i.e., letters and words) neuropsychological shifting performance predicted reading comprehension compared to a shifting measure revolving around colours and shapes. They found that the reading-specific shifting measure significantly predicted reading comprehension, whereas the domain-shifting general task did not. These results suggest that certain components of EF may be more directly involved in reading than others. Shifting may play a pronounced role because reading switches between rules based on context. While fluent readers may be more flexible in employing various reading rules, beginning readers may rely more on shifting capacity to apply the various sets of rules involved in reading words and text.

Arrington et al. (2014) sought to understand how certain EF traits related to decoding and comprehension. Their results showed that sustained attention, working memory, and cognitive inhibition significantly related to reading comprehension, whereas only working memory and response inhibition significantly related to decoding. One limitation is that, because the study did not take place in an intervention context, the results do not tell us whether this connection involves the learning of reading skill or merely the performance of reading. Studies have also examined how EF relates to progress in reading ability over time. For example, Swanson and Jerman (2007) conducted a study that looked at how working memory, a component of EF, predicted children's reading skill gains made across a three-year time span without intervention. They found that working memory was correlated with reading ability, and reading ability increased over time, but they did not find that growth in working memory could uniquely predict reading gains over time. Similar conclusions regarding general EF were found by Dolean et al. (2021). Nonetheless, these studies only looked at natural improvement in reading over time did not involve special intervention, whereas the present study does.

Executive Functions as Predicting Reading Intervention Outcomes

Due to the fact that EF predicts academic abilities such as reading (Yeniad et al., 2013), how EF plays a role in moderating the learning process (Center on the Developing Child at Harvard University, 2011; Serpell & Esposito, 2016), as well as how some children are less responsive to evidence-based interventions (e.g., Tran et al., 2011), researchers have investigated how EF may be linked to outcomes of reading interventions. In other words, there has been recent attention on whether there is a relationship between children's EF and the reading gains they make through participating in an intervention, although results have been mixed (Peng, 2023). Because intervention responsiveness isolates effects over time by examining the different between pretest and posttest, there are two main ways that EF may play a role in intervention effectiveness. First, EF may relate to how well they participate and learn within the intervention session. Second, EF may relate to the practicing of reading skill outside of the intervention context but within the intervention timeline. The role that EF plays in reading performance is not at the forefront because as underlying cognitive mechanisms they are assumed to affect pretest and posttest performance in a similar way. Rather, the role of EF in intervention responsiveness entails the latent function of reading gains over time. In pursuit of this question, two studies looked at a reading intervention for typically developing students rather than students with RD.

A Chinese study by Gao et al. (2018) used a sample of 60 typically developing students aged 11 to 12 years old, finding that performance on shifting, also known as cognitive flexibility, as measured by the Wisconsin Card Sorting Task, predicted rapid reading gains made from a 9-hour intervention. Reading Speed was captured by the number of characters per minute and reading comprehension was measured via a paper and pencil multiple choice test after reading a reading passage. Effective Reading Rate was a combination of the two reading measures. This was a randomized control trial where the intervention had a significant effect by increasing reading skill in the experimental condition compared to the control condition. Within the experimental group, students with higher shifting scores had significantly higher increase in reading skill compared to those with lower shifting scores. Notably, the Chinese language is substantially different than English and the intervention was distinct from PHAST in that it was neither intensive nor targeted at students with RD.

Using a whole class intervention without looking at RD specifically, DeBruin-Parecki and Cartwright (2023) conducted a study to understand how preschoolers' developing cognitive skills may support their narrative comprehension development. Taking place in the US, 47 preschoolers participated from three preschool classes– each one assigned to either the inferential, vocabulary-based, or business-as-usual instructional conditions. Shifting was assessed by having participants sort sets of images by colour and shape into a 2×2 grid. Inhibition was assessed using the day/night task in which participants were shown images of the sun and moon and were told to say "day" when a picture of the moon was shown and "night" to a picture of the sun. Working memory was measured by task performance on digit span forward and backward. Participants who received structured inferential comprehension instruction showed significant gains in narrative comprehension, whereas those in the other two conditions did not; this effect prevailed even after accounting for receptive vocabulary. Moreover, participants in the inferential comprehension condition made greater gains if they had high cognitive flexibility or theory of mind, but not other EF scores, suggesting that these two preschool cognitive skills may be important for developing narrative comprehension through instructional practice. The similarities to the present thesis are the use of inhibition/working memory/shifting as the three core EF and narrative comprehension as a construct that overlaps with reading comprehension. Key distinctions in this study are that the method of measuring EF was not standardised, the sample was younger (preschool instead of Grade 3 or 4), the lack of behavioural rating scales for EF, the lack of ADHD or RD diagnoses, the lower intensity of the intervention, as well as the lack of decoding or fluency measures.

A review by Horowitz-Kraus (2023) detailed various studies that either trained EF simultaneous to reading, or trained EF prior to intervention. Highlights from these results are how inhibition, shifting, and working memory were consistently found to be related to decoding and comprehension. However, these studies found that EF abilities at pretest could not predict intervention gains. Most of these studies were conducted in the Hebrew language, posing a generalizability issue. Among other differences, Hebrew involves different morphological rules than English, the language concerned in the present study. For example, one important early literacy skill is word recognition, which involves knowledge of irregular verb tenses, and irregular conjugation differs in English and Hebrew, meaning that their results may not extrapolate to English speakers.

Table 1 includes five articles I reviewed in-depth, focusing on studies intervention responsiveness using samples of students with RD to compare methods and results.

Table 1

Key Evidence Outcome Analysis for Studies Predicting RD Intervention with EF

	Walda et al., 2014	**Miciak et al., (2019)**	**Wilson et al., 2021**	**Walda et al., 2022**	**Walda et al., 2023**
Region/ Language	Netherlands/Dutch	USA/English	UK/English	Netherlands/Dutch	Netherlands/Dutch
Sample	**Sampling**: Referral and purposive sampling, targeting primary school children below the 10th percentile in decoding or spelling who were referred to services at a clinic. **N =** originally 229 (ranged from 62 to 169, depending on EF due to pairwise exclusion) **Age:** 7 to 11 years	**Sampling:** Referral sampling (inferred), random assignment to intervention condition from larger study. **N =** 231 **Age:** M = 9.7	**Sampling:** none reported. **N =** 391 **Age:** 6-7 years	**Sampling:** Referral and purposive sampling, targeting primary school children below the 10th percentile in decoding or spelling who were referred to services at a clinic. **N =** 137 **Age:** M = 9.1 years Range = 7 to 11	*Sample and sampling technique varied depending on study.
EF Measures	**Planning/Reasoning:** *Tower of London* (experimental computerised version). **Executive control:** *Berg's card sort task*. **Attentional control/Inhibition:** *Go-Nogo task*. *Flanker task arrows*. *Flanker task letters*. **Working memory:** *Digit span forward and backward*. *Block recall*. **Sustained/divided attention:** *TEA-Ch score double task*. **Inhibition:** *Opposite worlds*. Standardisation is inferred as the paper indicates that only the reading and spelling measures were not standardised.	**Working memory as recall span with planning:** *Listening Span* subtest from the *Working Memory Test Battery for Children.* *Tower of London*. *WJ III Planning subtest*. **Working memory as updating in active memory:** *N-back task*. **Initiation/Generative fluency:** D-KEFS Lexical, Categorical, and Design Fluency tasks. **Self-regulated learning:** *Contextual Learning Scale* (researcher-created). **Metacognition:** *BRIEF Metacognition* (teacher rating). *SWAN Inattention subscale*.	**Working memory**: Digit span backwards; **Inhibition/Shifting:** Heart and flower task [when shown a heart you press a button on the same side as the heart, when shown a flower you press a button on the opposite side of the flower]. Raw scores were used.	**Attention:** *Attention-Concentration Task*. No standardisation reported.	**Study 1 Predictors:** *Nonword repetition;* *Naming speed;* *Phoneme segmentation;* *Phoneme deletion*; **Study 2 Predictors:** *Grapheme phoneme identification;* *Grapheme phoneme discrimination;* *Naming speed;* *Vocabulary;* *Nonverbal reasoning;* *Digit recall;* *Block recall;* *Word recall;* **Study 3 Predictors:** *Rhyme;* *Rhyme prime;* *Auditory synthesis;* *Phoneme deletion;* *Letter naming*;
Reading Measures	**Decoding/Fluency:** *Word decoding 1*. *Word decoding 2*.	**Decoding:** *WJ III LWI and Spelling subtests.* **Fluency:** *TOWRE SWE*.	**Fluency:** TOWRE PDE and SWE.	**Decoding/Fluency:** *Word decoding 1*.	**Outcome for all 3 studies:** *Word decoding efficiency*.

	Word decoding 3. *Word decoding 4.* *Text reading.* **Spelling**: *Word spelling test.* No reading or spelling measures were standardised. No composite scores were calculated.	**Comprehension:** *Gates–MacGinitie Reading Test*. *WJ III Passage Comprehension.*	**Spelling:** Single Word Spelling Test. Raw scores were used.	*Word decoding 2*; reliability of summed 1 and 2 = 0.92-0.97. *Word decoding 3*; reliability of summed 1, 2, and 3 = 0.94-0.97; convergent validity = 0.76-0.92 (reading aloud text), 0.70-0.85 (silent reading), 0.12-0.36 (vocabulary), 0.45-0.48 (spelling). *Word decoding 4*; parallel test reliability = 0.90-0.96; test-retest reliability = 0.80-0.92; sufficient construct validity. *Text reading*; reliability = 0.85-0.91. **Spelling**: *Word spelling test*; reliability = 0.90-0.93. No reading or spelling measures were standardised. No composite scores were calculated.	
Intervention Details	**Length:** weekly 45min session plus homework for 9 months. **Targeted skills:** spelling and decoding. **Intervention:** standard remediation program.	**Length:** 30 min 5 times per week for ~16 weeks. **Targeted skills:** vocabulary, decoding, passage comprehension. **Intervention:** standardized, researcher-delivered. **Group size:** four to five students.	**Length:** 10-15min per day for 12 weeks; 10 hours total. **Intervention**: GraphoGame Rime	**Length:** weekly 45min session plus homework for 9 months (27 hours excluding homework). **Targeted skills:** spelling and decoding. **Intervention:** standard remediation program.	*Intervention varied depending on study.
Statistics/ Outcome	80% of correlations were significant between EF and either pretest or posttest reading and spelling. However, when calculated as gains, only 8.2% of correlations were significant between EF and spelling or decoding/fluency. Repeated measures ANOVA, with EF as a predictor and age as a covariate indicated that only 13.2% of results changed when EF was added into the model.	Using a 3 x 6 ANOVA, only general EF and metacognition, not the other 4 EF constructs, significantly varied between the non-RD participants and inadequate responders with RD. No significant differences were detected between adequate and inadequate responders with RD.	Higher working memory and inhibition/shifting performance at pretest predicted higher progression made in the game and higher posttest reading scores. This held true after controlling for other cognitive factors such as phonological awareness.	Majority of participants had trouble completing the ACT without making errors, contrary to a non-dyslexic sample that had all completed the ACT error-free aside from one participant. Higher working speed was significantly correlated with higher word decoding, text decoding, and spelling, but distraction time was not. Main effect of time on reading gains remained significant using age as a covariate, even when adding working speed and distraction time as covariates.	In study 1, the ML model performed similar predictions to the traditional GLM as both yielded acceptable predictive ability. In study 2, the ML model performed similar predictions to the traditional GLM, as both yielded poor predictive ability. In study 3, the ML model performed similar predictions to the traditional GLM once again, as both yielded poor predictive ability.

As seen in Table 1, predicting intervention gains with EF has been looked at through various approaches by Walda et al. in 2014, 2022, and 2023). Their first analysis used a sample of 229 Dutch children, ranging in age from seven to 11 years old, who participated in a reading intervention provided by a major clinic. The intervention, targeting the skills of spelling and decoding, provided involved special instruction as well as home reading with a parent. Intervention length was a standard remediation program provided weekly in 45min sessions, plus homework for 9 months. EF measures were shifting, four different inhibition tasks, digit span forward and backward plus block recall for working memory. Standardisation of EF is inferred as the paper indicates only that the reading and spelling measures were not standardised. Decoding/fluency, text reading, and spelling were measured. Pearson correlations revealed 80% of relationships were significant between EF and either pretest or posttest reading and spelling. However, when calculated as reading gains, only 8.2% of correlations were significant between EF and spelling or decoding/fluency. With age as a covariate in a regression model, most results were non-significant. Therefore, they concluded that EF was not related to intervention responsiveness.

Walda et al. (2022) examined how performance on an attention task related to gains made in a reading intervention in a sample of 137 participants with RD, ages ranging from 7 to 11 years old. Attention was measured via the Attention-Concentration Task (ACT) broken down by working speed and distraction time. Reading measures resembled those in their 2014 study. A surprising latent finding was that the majority of participants had trouble completing the ACT without making errors, contrary to a non-dyslexic sample where all but one completed the ACT error-free. This supports previous findings where RD often involves cognitive deficits unrelated to reading. Higher working speed was significantly correlated with higher word decoding, text decoding, and spelling, but distraction time was not. The intervention was effective by showing significant progress over time, but the amount of progress did not significantly differ as a function of attention. Thus, students with RD frequently show attention problems, but the variance in attention did not explain the variance in reading intervention responsiveness.

Most recently, Walda et al. (2023) analyzed three sets of data using a supervised machine learning method to detect prediction models that can accurately portray patterns between children's cognitive skills and progress in reading. Reusing the data from their other studies, they once again found no significant results. One limitation is that EF was measured only through neuropsychological tests, not behavioural rating scales, such as parent reports or teacher reports. This means that participants who struggle with EF in ways that are not captured by these tests, such as emotional control or planning, are not taken into account despite being related to the EF construct in question. Interestingly, in the remediation program used for these studies, parents were involved with the intervention sessions as well as completing homework assigned by the interventionist. Due to less supervision of the homework tasks, internal validity is more questionable. In other words, certain participants may have spent more time completing the homework compared to others, yet this information was not captured.

In contrast to the conclusions of Walda et al. (2014; 2022; 2023), a study by Wilson et al. (2021) found that executive functions uniquely predicted reading gains using a computer game intervention.[4] Taking place in the United Kingdom, the participants were a cohort of 391 children (aged 6-7 years) who completed in a randomised controlled trial. Half of participants were randomly assigned to the intervention game and the other half were assigned to a business-

---

[4] Interestingly, the computer game intervention, GraphoGame, has recently won a UNESCO Literacy Prize (https://www.unesco.org/en/articles/finland-based-programme-receives-2023-unesco-king-sejong-literacy-prize)

as-usual control. The intervention consisted of a game where players matched sounds to letter patterns, practicing letter-sound correspondence.[5] EF measures consisted of working memory via the digit span backwards, and inhibition/shifting via the heart and flower task (when shown a heart you press a button on the same side as the heart, when shown a flower you press a button on the opposite side of the flower). Reading was measured in the form of fluency via the TOWRE PDE and SWE, and spelling via the Single Word Spelling Test. Higher working memory and inhibition/shifting performance at pretest predicted higher progression made in the game and higher posttest reading scores. This held true after controlling for other cognitive factors such as phonological awareness. Gender did not significantly vary across playing time, rhythmic synchronisation, phonological skills and EF skills. The intervention format and age were distinct from the present study, which means their results may not translate to the present study sample.

Additionally, a study by Miciak et al. (2019) found that EF could predict intervention responsiveness to a small degree in an English sample comparing both children with and without RD. Based on the intervention results, children with RD were further distinguished as adequate responders or poor responders based on whether they continued to meet criteria for RD. Their study divided EF pretest measures into a common EF factor model, included a common EF composite score, and used five latent factors: (a) working memory as recall span with planning; (b) working memory as updating in active memory; (c) generative fluency; (d) self-regulated learning (SRL); and (e) metacognition. They found that non-RD children who scored

---

[5] The game focused specifically on rime identification, which is a component of phonological awareness. Rime is the part of a syllable that includes the vowel sound and subsequent consonants. For example, in the single-syllable word *clock*, the rime is “ock” because it includes the vowel sound from “o” as well as the subsequent consonant sound.

significantly higher on the general EF factor as well as on the teacher-reported Behavioural Rating Scale of Executive Functioning (BRIEF) made greater reading gains compared to RD poor responders. However, they failed to find significant results in EF when comparing the adequate responders with the poor responders. Therefore, they did not find that EF is related to intervention responsiveness specifically in students with RD. Notably, their method for determining which students did not adequately respond to intervention was by comparison of posttest reading scores only, specifically whether participants continued to meet criteria for RD.

The Present Study

The present study builds from previous research by examining EF as a predictor of responsiveness to a reading intervention with similarities and differences in methodology. Understanding the extent that EF relates to gains made from an intensive reading intervention can inform the way intervention systems, such as the RTI framework, address students with both poor reading and poor EF. For example, if students' EF performance shows a strong positive correlation with reading gains, then in future implementations of these programs, students with poor EF could be provided greater educational support to mitigate the relationship with the aim that they make similar reading gains as their peers. To accomplish this, the present study examined a sample of students with RD in Grade 3 and 4 who completed the PHAST intervention, seeking to examine the relationship between the core EF (inhibition, shifting, working memory) and reading skill at pretest and gains after 70 hours of intervention. Reading skill was measured in terms of relatively lower level (decoding and word reading accuracy) and higher level (fluency and passage comprehension). By using data from participants in this intervention, the present study will provide useful insights to researchers and school systems alike.

Although previous studies have attempted to determine the relationship between EF and intervention responsiveness, such studies were limited by either the measures of EF and reading, or the calculation of gains. Additionally, few studies have been conducted in English. Consequentially, there are many unique angles in the present study. As discussed earlier, one complication in studying EF is that there are numerous ways to conceptualize EF, whether as a single construct or as multiple related ones (Fisk & Sharp, 2004; Miyake et al., 2000), corresponding to many psychometric approaches to measurement. Here, EF involving the core of inhibition, shifting, and working memory, as well as emotional control and higher-order EF constructs representing metacognition and reasoning. In all but emotional control, these forms of EF were derived from neuropsychological and parent report methods. Furthermore, no studies have examined this research question using the PHAST Reading Program, a highly intensive evidence-based intervention that has become more popular in North American schools in the commercial form of Empower™ Reading. Another unique approach of the present study is the analysis of reading gains. While the Walda (2014, 2022, 2023) studies used gains between adjacent follow up tests, the Wilson (2021) study used pre-post differences, and the Miciak study used post-test only. Yet, the standard error of the mean on the specific reading test should be taken into account, as is done with the Jacobson-Truax reliable change index (RCI; Jacobson & Truax, 1984). Therefore, by using an intervention study with the PHAST program, and by measuring reliable change in reading, the present study seeks to address various gaps around EF predicting reading intervention responsiveness.

Research Questions

1a) To what extent do participants make reliable change across different reading measures?

1b) How might these outcomes vary across different demographics of the sample?

2a) To what extent do participants' executive functions at pretest predict reading skill at pretest?

2b) Which executive functions predict which reading outcomes?

3a) To what extent do children's executive functions at pretest predict reading gains made during intervention?

3b) Which executive functions predict which reading outcomes?

4) To what extent do multiple deficits in executive function scores (i.e., below average standard scores across multiple EF constructs) relate to reading gains?

Hypotheses

Because the majority of the intervention was dedicated to decoding strategies, whereas only a minor component of the intervention was dedicated to reading comprehension, my overall hypothesis was that more gains would be made in in decoding compared to comprehension, as well as the fact that differences in gains would be most prominent in comprehension measures. My specific hypotheses are:

1) The majority of the sample will make reliable change in reading skill following the intervention. The highest rate of reliable change will be made in decoding skill, with less gains in comprehension.

2) EF at pretest will have strong positive correlations with reading skills at pretest.

3) EF will positively predict reading gains, especially comprehension. Specifically, shifting will make the strongest prediction of intervention responsiveness, because it has been one of the more prevalent predictors of learning and reading performance (Horowitz-Krauss, 2023), and the card sorting task in Walda et al. (2014) had the greatest number of significant correlations with reading gains.

4) Multiple deficits in EF will relate to a lower proportion of reading gains, especially in comprehension.

Methods

Participants

One hundred and thirty-nine children (M_{age} = 9.28 years; *SD* = 0.67) participated. Thirteen participants were removed due to withdrawal from the intervention program, and eleven were removed via listwise deletion for not completing all EF measures. The remaining sample size was 115, comprised of 67 male (58.3%) and 48 female (41.7%) participants (M_{age} = 8.74; *SD* = .76; 70 in Grade 3[59.5%], 45 in Grade 4[40.5%]). Forty-three participants (37.4%) had ADHD. Seventy-two percent of participants were Black, 21.7% were White, 1.6% were Latino, 0.9% were Asian, and 3.2% were biracial. Household income levels were distributed on scale from 1-5, with 8.7% in the first bracket, 27.8% in the second, 26.1% in the third, 18.3% in the fourth, 14.8% in the fifth, and 4.3% unreported. A referral and purposive sampling approach was undertaken (McNeill & Chapman, 2005) where potential participants were referred by their schools as struggling readers, with preference to children who had previously received at least one year of extra reading assistance.[6]

The study was approved by the Institutional Research Board for Georgia State University as well as the Brock University Social Science Research Ethics Board. In accordance with ethical standards, all participants provided informed consent, and their parents provided informed consent prior to their child's participation. Participants who met the following criteria at the screening stage were excluded: (a) below average intellectual functioning (standard score

---

[6] This was due to the goal of identifying children whose reading difficulties result from underlying neurobiological deficits rather than from poor educational instruction. That being said, this was a preference, rather than mandatory, especially considering the fact that causation mechanisms (genetic versus environment) may interact.

<80) on each subscale of the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 2011); (b) chronic absenteeism (>15 absences per year); (c) hearing impairment (>25dB at 500+ Hz with both ears); (d) uncorrected visual impairment (>20/40); (e) serious emotional or psychiatric issues; (f) chronic medical/neurological condition (e.g., seizure disorder); or, (g) certain metals in their bodies (e.g., braces, cardiac pacemaker, hearing aid, plates, or other metal objects). Exclusion criteria (a), (b), (c), (d), (e) and (f) were in order to control substantial confounds. For example, the intervention was not designed to accommodate intellectual or sensory impairments. Criteria (c), (d), (e), (f) and (g) were also mandatory because the study was part of a larger neuroimaging study using functional magnetic resonance imaging (fMRI) requiring certain safety restrictions. All children were native speakers of English for the purpose of avoiding confounds related to language acquisition.

Materials

PHAST Intervention

A modified version of the Phonological and Strategy Training (PHAST; Lovett et al., 2000a) intervention was used. Details about the intervention components and the overall model have been delineated in existing literature (Lovett et al., 2008; Lovett et al., 2017; Morris, et al., 2012). This program focuses on teaching students five reading strategies for word reading, in addition to working on other reading skills such as vocabulary and comprehension. The five strategies are: sounding out, rhyming, peeling off, vowel alert, and spy. Delivered by trained special education teachers, the intervention took place in elementary schools in Atlanta, Georgia, USA. The intervention lasted 45 to 60 minutes each day across 14 weeks for a total of 70 hours throughout the entire program. Efforts were made to minimize disruption to participants' other school activities. Students were divided into groups of five to eight based on similar reading ability to ensure a homogenous group, allowing the instructor to progress at a pace that worked well for all students. Each session started with a review of specific content, then featured three to six smaller themed mini-lessons, and ended with a consolidation activity. The themed mini-lessons alternated between the five reading strategies, as well as spelling, vocabulary, and comprehension. PHAST sessions are structured in a thoughtful manner such that prerequisite reading skills required for successfully implementing strategies are taught and practiced before the strategies themselves are introduced. Strategies are not introduced all at once, but rather gradually introduced. Children are instructed in a metacognitive approach that enables them to choose, use, monitor, and assess their application of strategies while reading. Likewise, the comprehension aspect of the program was not introduced until later into the 2nd half of the intervention.

Reading Measures

There is a multitude of techniques for measuring reading skills. The present study used the Woodcock-Johnson III Tests of Academic Achievement (WJ III), Test of Word Reading Efficiency II (TOWRE), the Standardized Reading Inventory 2nd Edition (SRI-2), and the Challenge Word Test (CWT).

Woodcock-Johnson Tests of Academic Achievement, Third Edition (WJ III). This reading battery included four subtests: word identification, word attack, reading fluency, and passage comprehension (Woodcock et al., 2001), of which raw scores were collected, then used to calculate scaled scores and W scores. Reliability statistics are reported as being 0.80 and above. Each subtest evaluates distinct aspects of reading, offering triangulation of participants' reading ability, and at posttest, their responsiveness to the intervention.

WJ III Letter-Word Identification (LWI). This subtest evaluates participants' capacity to accurately identify and pronounce letters and familiar words as indicators of word recognition skills. The test starts with identifying letters, then words in increasing difficulty (e.g., starting with *car*, and *dot*, then *about*, which continues until the participant no longer recognizes words accurately.

WJ III Word Attack (WA). This subtest examines the capability to decode non-words. An example non-word is "zoop." This subtest measures the application of knowledge in letter-sound relationships and syllable patterns to decode new words.

WJ III Reading Fluency (RF). This subtest assesses fluent reading speed and semantic processing speed, as indicators of how automatically participants read a single sentence and make semantic judgements by indicating if the sentence is true or false. For example, a question may say, "A clock tells time" for which the correct answer would be *true*. This is a timed assessment. Unlike TOWRE, the WJ III construct of fluency targets fluent reading comprehension, thereby drawing on the automaticity of verbal reasoning skills.

WJ III Passage Comprehension (PC). This subtest evaluates the ability to understand and interpret a single written sentence by way of filling in a missing word (known as cloze assessment). Participants would read a short sentence with a missing word which would be guessed based on the clues provided by an adjacent picture. These questions provide insight regarding the cognitive processes involved in reading text beyond individual word reading, such as verbal reasoning skills.

Test of Word Reading Efficiency, 2nd Edition (TOWRE). This instrument has two subtests, *Sight Word Efficiency* and *Phonemic Decoding Efficiency*, which assess fluency of reading of real words and pseudowords respectively (Torgesen et al., 2012). Unlike WJ III RF, the TOWRE focuses on fluency at an individual word level. Participants read as many words or pseudowords as they can in the span of 45 seconds. Raw and standardized scores were calculated for each subtest.

TOWRE Sight Word Efficiency (SWE). This subtest measures how quickly and accurately a person can identify and pronounce real words. It primarily evaluates the automaticity of sight-word reading, reflecting the ability to recognize words without having to decode them letter by letter. For example, one word might be *room*.

TOWRE Phonemic Decoding Efficiency (PDE). Unlike SWE, which overlaps with one's vocabulary ability, the PDE subtest assesses the ability to decode non-words by applying one's decoding ability (i.e., familiarity of phonetic principles) in a fluent manner. It measures how well an individual can use letter-sound correspondence and phonological awareness to sound out and pronounce words efficiently. An example of a non-word is "ib" which could be decoded using the rhyme strategy to compare the known word "rib" to "ib."

Standardized Reading Inventory, 2nd Edition (SRI-2). Two subtests of this instrument were used: Word Recognition Accuracy (WRA) and Passage Comprehension (PC). Participants read a fictional text excerpt once aloud and once silently (Newcomer, 1999). Raw and standardized scores were calculated. Test-retest reliability exceeds .85, with inter-rater reliability measured between 90% to 95% agreement. Of note, the passages in this instrument are longer than those in the WJ III PC instrument.

SRI-2 Word Recognition Accuracy (WRA). Similar to the WJ III Letter-Word Identification, this subtest measures participants' ability to recognize words and pronounce them accurately. Unlike the WJ III LWI test, however, the scoring for SRI-2 WRA is done whilst the participant is reading the aforementioned fictional passage.

SRI-2 Passage Comprehension (PC). This subtest involves the same fictional passage, but scoring is instead derived from comprehension questions asked of the participant after reading once aloud and once silently. Comprehension questions include lexical, inferential, and factual open-ended prompts regarding the text.

Challenge Words Test (CWT). This test requires participants to put together the decoding strategies learned from the PHAST reading intervention, measuring how well participants can read long multisyllabic words not instructed during the intervention by applying each strategy (Lovett et al., 2000b). Participants' flexible application of learned strategies to new words taps into metacognition by applying their knowledge in an ambiguous decision-making context.

Executive Function Measures

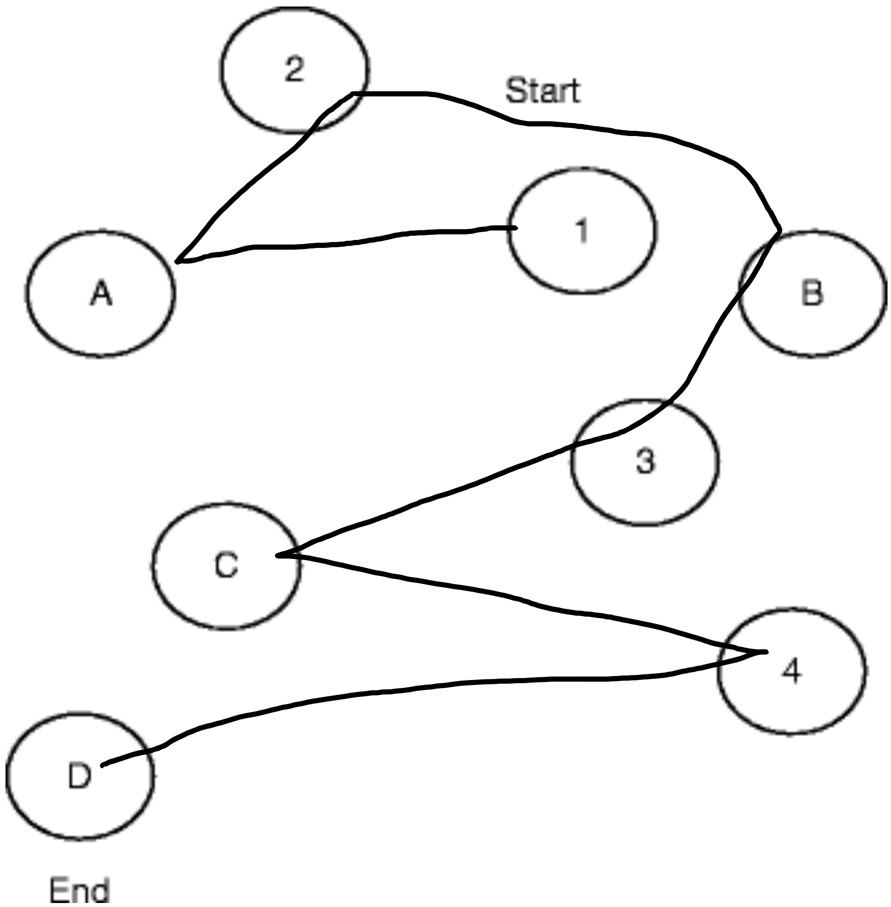
As with reading skill, EF can be measured in a multitude of ways (Anderson, 2002; Diamond, 2013; Packwood et al., 2011). In the present study, EF scores were measured at pre-intervention only and were derived primarily from the administration of neuropsychological tests. Because the validity of neuropsychological tests are specific to performance in the laboratory or clinical setting, the other method employed was parent-report behavioural rating scales, intended to capture behavioural patterns outside the lab setting that are associated with executive functions (e.g., the ability to keep one's belongings organized).

Delis-Kaplan Executive Function System (D-KEFS). This battery (Delis et al., 2001) has been extensively used in clinical and research settings (Delis et al., 2004). The present study used the Trail-Making, Colour-Word Interference, and Free Sorting subtests. Subtests had multiple conditions each. In other words, each subtest had different variations of the task to measure different aspects of task performance. Furthermore, scaled scores and contrast scores were derived from the raw scores, which is a common practice across the battery's wide usage, particularly in developmental contexts as is the case of the present study (Miyake & Friedman, 2012; Lee et al., 2013; Serpell & Esposito, 2016).

D-KEFS Trail-Making (TM). This subtest measures shifting and is broken down into five conditions and involves connecting lines between labeled circles (called "nodes") in specific sequences. The first condition was visual scanning, in which participants are asked to point to letters and numbers to ensure they have the visual-spatial ability to complete the task. Condition two was number sequencing, in which participants draw lines from numbers in order to verify their number sequencing ability. Similarly, condition three was letter sequencing, in which participants draw lines from A to B to C, and onward in order to verify their letter sequencing ability. Condition four was number letter switching, in which participants alternate between numbers and letters in their respective sequences. This means starting on the node labelled "1," then drawing a line to the node labelled "A," then drawing a line to "2," then "B," and so forth (See Figure 2). The fifth condition is motor speed, measuring how fast one can trace dotted lines between unlabelled nodes. The principal scores used from this measure were the raw and scaled scores for condition four (number-letter switching), as well as the contrast scaled score which takes the condition four score and controls for both condition two and three. Trail-Making internal consistency reliability is .60-.81 (Stephens et al., 2014).

Figure 2

Number-Letter Switching Condition of the Trail-Making Subtest

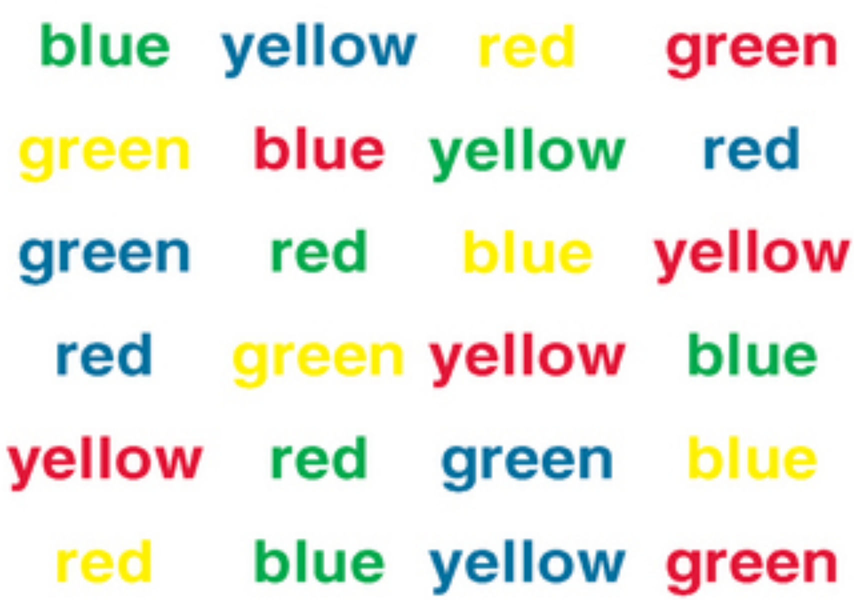


Note. Source: https://link.springer.com/referenceworkentry/10.1007/978-0-387-79061-9_2934

D-KEFS Colour-Word Interference (CWI). This task derives two useful measures: inhibition (CWI-I) and shifting (CWI-S). The premise is twofold: (a) participants read the names of colours, but instead of saying the name of the colour written in text, they say the name of the ink colour; and (b) switching rules when the word is outlined by a black rectangle. To account for basic perceptual processes, the first two conditions involve reading the names of colours written as text and reading the names of colours based on the ink colour. A D-KEFS variation of the widely known Stroop test, this assessment measures inhibition as participants must inhibit their automatic response to read the word on the page to complete the task (See Figure 3). The second part of CWI involves participants switching between two sets of rules: initially, they use the same set of rules from the previous task; when the word is contained in a black rectangular outline, they are expected to switch rules and read the word on the page as opposed to the colour of the ink. Therefore, this second task measures the shifting component of EF because participants need to be flexible in switching from one set of rules to another. Condition three (inhibition) test-retest reliability is .75, whereas condition four (shifting) test-retest reliability is .65 (Stephens et al., 2014).

Figure 3

Depiction of Text from Colour-Word Interference Task



Note. Source: Kinsella et al. (n.d.) https://neupsykey.com/executive-function-and-its-assessment/.

D-KEFS Free Sorting, Card Set 1 (CS1) and 2 (CS2). This card sorting assessment measures higher-order EF in the form of problem-solving and abstract reasoning. There are two sets of cards and thus two rounds of the test. Each set includes six cards of various shapes, sizes colours, and word labels. Participants sort the cards into two groups of three based on specific verbal or nonverbal rules. Verbal rules might include the label indicating an animal or a vehicle. Nonverbal rules include the colour, shape, or size of the card. Test-retest reliability ranges from .55-.81 (Stephens et al., 2014).

Digit Span Backward (DSB). This task is from the Wechsler's Intelligence Scale for Children, Fourth Edition (WISC-IV; Wechsler, 2003). The psychometrist will read a series of digits, and participants repeat the digits in backward order. For example, the psychometrist will say, "2, 1, 9" and the participant is supposed to say, "9, 1, 2." The psychometrist begins with a trial using only two digits to ensure the participant understands the concept of the task. Then, the number of digits increases to three, then four, and so forth, until the participant multiple mistakes in a row. The score represents working memory due to requirement of manipulating the order of the digits within one's short-term memory system. The backward subtest was used in the present study.

Behaviour Rating Inventory of Executive Function (BRIEF). This instrument is a reporting scale that assesses day to day behaviours associated with executive functioning in children and adolescents aged 5 to 18 years old (Gioia et al., 2000). There are three main forms of the BRIEF: parent-report, teacher-report, and self-report. The items in the BRIEF include higher-level executive functions such as planning, organizing, problem-solving, in addition to including "hot" EF in the form of emotional regulation. The present study uses the parent-report form, which is completed by parents or caregivers who report based on their experience of the child's behaviour at home. The questionnaire consists of 86 items that assess various aspects of executive function. These items are grouped into eight EF traits enumerated here with a short description of what they signify in the context of this instrument: (a) *inhibit*, as in the ability to control impulses and stop one's own behaviour; (b) *emotional control*, as in the ability to regulate emotional responses; (c) *shift*, as in the ability to move freely from one situation or activity to another as the circumstances demand; (d) *working memory*, referring here to the ability to hold information in mind for the purpose of completing a task; (e) *initiate*, as in the ability to independently begin a task or activity and generate ideas, responses, or problem-solving strategies; (f) *monitor*, as in the ability to sustain awareness of one's own actions and evaluate their effects on others or on a task; (g) *plan/organize*, as in the ability to manage current and future-oriented task demands effectively; (h) *organization of materials*, as in the ability to impose order on work, play, and storage spaces. Responses are scored to indicate the frequency of observed behaviours related to EF deficit such that higher scores on any given scale indicate greater difficulties in that area of EF. In addition to the raw scores, the BRIEF provides standard scores (T scores) based on norms for the child's age and gender. For example, although two nine-year-olds received the same raw score on BRIEF Inhibition, they had different T scores because one participant was female and the other was male. Test-retest reliability is reported as .59-.96 (Roth et al., 2014).

Procedure

Data Collection

Screening instruments were used to determine whether students met criteria for RD by assessing if they were below 85 standard score on a composite of WJ III or TOWRE-2. Concurrently, participants were screened for ADHD with the Strengths and Weaknesses of

ADHD-Symptoms and Normal-Behaviour Scale (SWAN; Swanson et al. 2001). Data were collected across 5 academic years from 2014 to 2019.

Data Processing

Standard Score Conversion. Standardization formulas were used to convert raw scores into standard scores relative to developmental norms. These are referred to as scaled scores in the case of D-KEFS and Digit Span and T scores in the case of BRIEF. The premise behind the scaled score calculation was to convert the raw score into a new score that accounts for age where a mean score of 10 is defined as the population norm for the respective age, setting the standard deviation to three (mean of 100 and SD of 15 for WJ III and TOWRE measures). The EF and reading measures all had scaled score variations. This conversion formula first turns the raw score into a *z*-score using the mean and standard deviation of the normative sample (rather than the sample of the present study). BRIEF T scores used the same process, with a norm mean of 50 and SD of 10.

Here is the breakdown with an example below:

$$\text{Scaled Score} = 10 + (Z \times 3)$$

$$Z = \frac{X - \mu}{\sigma}$$

X = Raw score

μ = Mean of the raw scores in the normative sample

σ = Standard deviation of the raw scores in the normative sample.

Say the normative sample mean (μ) and standard deviation (σ) of the raw scores are 50 and 10, respectively. A participant who obtained a raw score of 65 would have the following scaled score conversion:

1. Calculating the *z*-score:

$$Z = \frac{65-50}{10} = \underline{1.5}$$

2. Converting the *z*-score to a scaled score:

$$\text{Scaled Score} = 10 + (\underline{1.5} \times 3) = \mathbf{14.5}$$

So, the scaled score would be 14.5. Note that the normative sample mean was differentiated between subsamples based on age and gender.

Contrast Score Conversion. The D-KEFS TM and CWI measures also had contrast scaled scores to control for non-EF-related performance measures that may confound with measuring EF. TM had a contrast score that represented the condition 4 task performance, controlling for performance on the number switching and letter switching (conditions 3 and 4). CWI had contrast scores for both inhibition and shifting. CWI-I controlled for colour reading, and CWI-S controlled for CWI-I.

Reliable Change in Reading. The Jacobson-Truax Reliable Change Index (RCI; Jacobson & Truax, 1984, as modified in Christensen & Mendoza, 1986) was applied to transform pretest and posttest differences to a categorical (i.e., yes or no) score of whether the participant made reliable gains in each of the reading measures included above in addition to an overall reliable change in reading gains score.

$$\text{RCI} = \frac{(x_2 - x_1)}{\text{SEM} \times \sqrt{1-r}}$$

Where:

- x_1 = Initial score (pre-intervention score)
- x_2 = Follow-up score (post-intervention score)
- SEM = Standard Error of Measurement for the test
- r = Reliability coefficient of the test

This formula produces a standard error and reliability-adjusted gains score, from which a critical value of 1.96 was assigned for the sample's distribution to establish threshold by which a given participants gains score was significant or not, aligning with a *z*-score of 2 standard deviations increase in gains that is meaningful and not attributed to chance. A new distribution of *z* scores was generated centred around this critical value as the mean and the SD of the reliable gains score. This *z*-score distribution was then used to establish a continuous score meant to capture the strength of reliable gains relative to others in the sample, considering nuances beyond the dichotomy of making significant gains or not, as both categories feature notable variance in gains. These *z*-scores were then combined to form the following RCI composites: (a) reliable change in single-word decoding, consisting of WJ III LWI & WA, CWT, and SRI-2 WRA; (b) reliable change in fluency, consisting of TOWRE SWE and PDE; and (c) reliable change in text comprehension, consisting of WJ III RF, WJ III Comprehension, and SRI-2 Comprehension. An overall reliable change in reading variable was created as the average of the composites. The categorical variable for overall reliable change was derived from latent class analysis of the decoding measures.

Analysis

Within the dataset, there was diversity across EF and reading skill in terms of both the instruments used (e.g., WJ III and TOWRE) and the underlying constructs represented (decoding and comprehension), and the overlap thereof, the analyses grouped results based on underlying construct as opposed to instrument. For EF, the four main categories were:

- *Inhibition*, entailing CWI-I (including its contrast score), and BRIEF Inhibition;
- *Shifting*, entailing CWI-S (including its contrast score), TM (including its contrast score), and BRIEF Shifting;

- *Working Memory*, entailing DSB and BRIEF Working Memory;
- *Emotional Control*, entailing BRIEF Emotional Control;
- *Reasoning/Metacognition*, entailing CS1, CS2, BRIEF Initiate, Monitor, Plan/Organize, and Organize/Materials.

First, reliable change outcomes were observed across the sample, which involves a continuous variable representing the *z*-score of reliable gains and a corresponding nominal variable representing whether gains were reliable (yes or no). Chi-square analyses were run to determine if the proportion of reliable change varied across demographic factors. Then, correlations were run between pretest reading and gains measures to determine whether reading scores at pretest was a significant determinant of reading gains. Next, correlations between EF and pretest reading measures were conducted to determine the extent that EF is associated with reading ability prior to the intervention. Then, correlations were run to determine if any EF measures related to composite reliable change in reading gains measures. Specifically, the following correlations were conducted using their standard score variations: TM, TM Contrast, CWI-I, CWI-I Contrast, CWI-S, CWI-S Contrast, CS1, CS2, DSB, BRIEF with RCI Overall, RCI Decoding, RCI Fluency, and RCI Comprehension. The high number of correlations were conducive to the application of Bonferroni and Benjamini-Hochberg correction procedures to control for family-wise error rate (Benjamini & Hochberg, 1995). To complement the analysis of individual EF measures vis-à-vis reliable gains in reading, clusters of EF deficits was also created for examination, similar to previous studies (e.g., Iverson et al., 2020). After reviewing how individual EF correlated with one another, a multi-deficit variable was created using CWI- and CWI-S, CS1 and CS2, and DSB. In the form of independent samples *t*-tests, analyses were conducted to see if the lower multivariate EF group differed in their reading gains compared to the high group; (b) a chi-square analysis was run to determine what parts of the reading composite measures differed in the rate of meeting the threshold of reliable gains or not.

Results

Missing Data

Thirteen participants were excluded due to intervention withdrawal. Chi-square tests revealed that those who withdrew generally did not differ from those that completed the intervention in terms of most key demographic variables (grade, race, gender, and ADHD status; *p*s>.05). Nonetheless, considering the factor of household income scale of 1-5, the difference between the observed and expected ratios of participants who withdrew was near significant at the .05 level (χ^2 = 9.41; *p*=.052; Cramér's *V* = .27). Specifically, on the household income factor, no participant from the two levels of higher income (four and five) withdrew, whereas there was more than twice the number of withdrawals in the lowest income level than would be expected if withdrawal was random and proportionately distributed across income groups.[7] Therefore, it appears that lower income families were more likely to withdraw from the study, indicating a potential source of bias by inadequately representing this portion of the population. Additionally, independent samples *t*-tests determined that participants who withdrew had significantly poorer performance than those who did not on three of the eight pretest reading measures (37.5%) and on 51.6% of pretest EF measures (*p*s<.05). However, the considerable sample size difference (withdrawn, *n*=13; not withdrawn, *n*=126) caused the assumptions of normality to be violated, meaning this difference is not definitive. Eleven of the 126 participants who completed the

[7] Due to 8 missing data points on the income variable (3 from the withdrawn group), there was an adjusted *N* of 131: 10 in the withdrawn group, 121 in the non-withdrawn group.

intervention did not complete every single executive function measure. Therefore, analyses were conducted using listwise deletion, leaving $N = 115$.

Assumptions of Normality

Scores with significant skewness or kurtosis underwent two tests of normality – the Kolmogorov-Smirnov test and the Shapiro-Wilk test (Appendix A). While many of the EF and reading scores exhibited skewness, investigation of the Q-Q plots and histograms showed the distribution shapes were within the $z_{dev}= 0.5$ range and thus were assumed to be fit for parametric tests due to the relatively robust sample size. Variables with the most concerning skewness and kurtosis (i.e., above three) demonstrated some degree of multimodality or heavy tails, and their results from parametric tests were subsequently interpreted with greater caution, which is discussed later in this section.

Reading Descriptives

My first research question inquired about the reliable change outcomes across different reading measures and across different demographics of the sample. Table 2 presents pretest and posttest raw and standard score results on all reading measures as well as rates of reliable change.

Pre and Post Test Scores

For WJ III and SRI-2, the developmental norm is for standard scores to fall between 85 to 115, with 100 being the average. Average standard scores for all WJ III and TOWRE measures consistently fell either below this range or at the lower tail (i.e., just above 85) at pretest. For SRI-2, the norm is for standard scores to fall within 8.5 to 11.5, meaning the present sample was even further below norms on these measures. This pretest pattern is expected given the recruitment strategy was based on identifying struggling readers in this range. At posttest, the WJ

III scores no longer fell below the normal range, although posttest scores fell below the range on TOWRE and SRI-2 measures.

Table 2

Descriptive Statistics for Raw and Standard Reading at Pretest and Posttest

	Pretest		Posttest		% Made Reliable Gains
	Raw Score (SD)	Standard Score (SD)	Raw Score (SD)	Standard Score (SD)	
Decoding					
WJ III LWI	459.3 (17.9)	88.3 (8.1)	478.3 (17.3)	91.1 (8.5)	29.6%
WJ III Word Attack	472.4 (16.0)	88.3 (8.9)	485.9 (12.9)	92.1 (7.6)	22.6%
SRI-2 WRA	11.8 (6.2)	5.5 (2.3)	16.0 (5.7)	6.2 (2.3)	22.6%
Fluency					
TOWRE SWE	41.0 (12.2)	76.1 (10.8)	54.4 (11.7)	83.8 (12.0)	38.3%
TOWRE PDE	11.5 (5.9)	73.3 (8.4)	18.7 (8.3)	79.6 (10.8)	47.0%
Comprehension					
WJ III RF	482.1 (6.8)	87.0 (10.1)	490.8 (7.4)	93.8 (10.3)	27.8%
WJ III PC	468.4 (10.8)	80.7 (8.4)	483.8 (12.8)	87.6 (10.5)	41.7%
SRI-2 PC	23.6 (9.3)	6.8 (2.0)	29.8 (10.6)	7.2 (2.2)	35.7%
Challenge Word Test	4.1 (3.3)	--	8.9 (4.7)	--	67.8%

Note. Sample size = 115. Standard score means from normed sample are as follows: (i) WJ III and TOWRE = 100; (ii) SRI-2 = 10. Mean and standard deviation for all measures are rounded to one decimal place due to being scored on an integer scale.

Reliable Change

According to RCI calculations, 84 participants (73%) reached the threshold for the composite reliable change category, whereas 31 (27%) did not. Using this composite category, chi-square tests were conducted across income, race, gender, grade, and ADHD categories to determine if rates of reliable change were uniform across groups. Grade was the only factor that emerged significant such that Grade 3 participants had significantly higher rates of reliable change compared to their Grade 4 peers (χ^2=4.4; p=.04). Reliable change frequencies became more dispersed when examining reading constructs separately, each having less than 50% rate of reliable gains (See Table 2). The exception to this trend is performance on the intervention-specific CWT measure, on which 67.8% of participants made reliable gains.

Relationship Between Pretest Scores and Gains

As shown in Table 3, the following seven of the nine reading measures had significant negative correlations between pretest and gains: WJ III LWI, WA, and RF, TOWRE SWE, SRI-2 WRA and PC, and CWT. Applying the Bonferroni correction adjusted the significance value to .006 (.05 divided by 9 comparisons), leaving five as significant – WJ III WA and RF as well as the two SRI-2 measures and CWT. The remaining four, the WJ III LWI and PC, TOWRE SWE and PDE, were not significant (*p*s>.3). These results indicate that students who did poorly on pretest reading measures made greater gains than those who did better, although this depends on the specific reading skill. The exception is the CWT where a strong opposite relationship was found.

Table 3

Correlations Between Standard Pretest Reading and Gains

	r	***p***
Decoding		
WJ III LWI	–.19*	.05
WJ III Word Attack	–.56**	<.001
SRI-2 Word Recognition Accuracy	–.43**	<.001
Fluency		
TOWRE SWE	–.23*	.01
TOWRE PDE	–.07	.47
Comprehension		
WJ III Reading Fluency	–.38**	<.001

WJ III Passage Comprehension	–.09	.35
SRI-2 Passage Comprehension	–.31**	<.001
Challenge Word Test	.71**	<.001

**p<0.01 (2-tailed). *p<0.05 level (2-tailed).

Note. Degrees of freedom for all correlations = 114.

Executive Function Descriptives

Completion time for the D-KEFS and DSB EF measures as well as scores for the BRIEF parent report measures are listed in Table 4 in raw and standard format. Notably, all D-KEFS timed measures had multiple participants who reached the maximum amount of allotted time. Considering scaled scores have a norm average from 7 to 13, scores on TM were importantly below the average, and most were in the low average range. Although, unlike reading skill, having low EF scores was not a part of the present study's recruitment strategy. All BRIEF T-score means fell within the developmental average range of 40 to 60. Therefore, D-KEFS tasks, but not BRIEF, suffered from floor and ceiling effects. Specifically, ceiling effects were found for completion time scores such that many participants ran out of time, while floor effects were seen on CS2 where numerous participants clustered at the lowest possible score.

Table 4

Descriptive Statistics for Raw and Standard Score Means of Executive Functions

	Raw Score (SD)	Standard Score (SD)
Inhibition		
D-KEFS CWI Inhibit	112.7 (28.6)	8.4 (3.5)
D-KEFS CWI Inhibit Contrast	--	9.7 (3.1)
BRIEF Inhibit	15.8 (5.3)	51.8 (11.8)
Shifting		

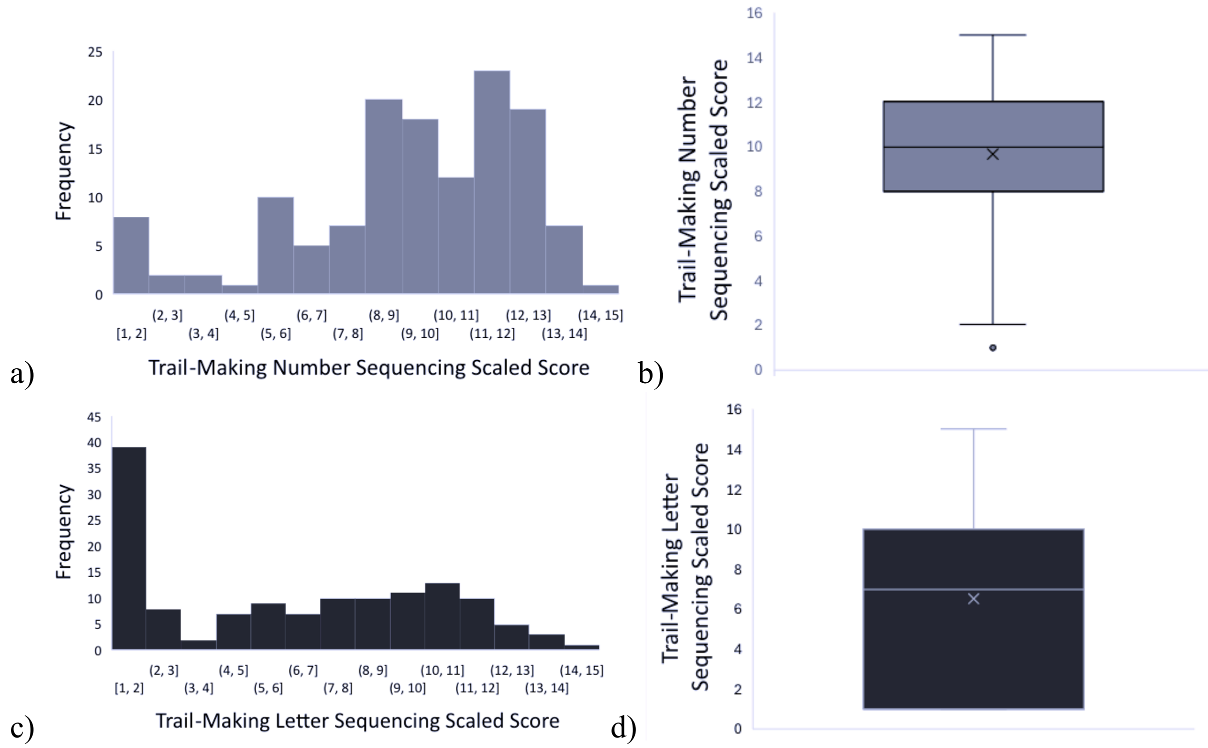
D-KEFS Trail-Making	179.6 (50.5)	6.8 (3.7)
D-KEFS Trail-Making Contrast	--	8.5 (3.3)
D-KEFS CWI Shifting	113.0 (30.8)	8.5 (3.7)
D-KEFS CWI Shifting Contrast	--	10.1 (3.2)
BRIEF Shift	12.6 (3.8)	52.0 (12.6)
Working Memory		
Digit Span Backward	5.5 (1.1)	8.2 (2.2)
BRIEF Working Memory	19.3 (5.3)	57.8 (11.5)
Emotional Control		
BRIEF Emotional Control	16.9 (5.5)	51.2 (12.2)
Reasoning and Metacognition		
D-KEFS Card Sort 1	2.8 (1.3)	7.8 (3.0)
D-KEFS Card Sort 2	2.6 (1.4)	7.5 (2.9)
BRIEF Initiate	13.8 (3.4)	53.4 (10.5)
BRIEF Monitor	15.0 (4.0)	52.3 (11.8)
BRIEF Plan/Organize	21.4 (6.2)	54.3 (12.8)
BRIEF Organization/Materials	12.2 (3.6)	52.2 (11.0)

Note. Only pretest scores were collected. Standard score means from normed sample are as follows: D-KEFS and Digit Span = 10; BRIEF = 50.

To further investigate the performance on timed D-KEFS assessments, the subtests, the TM patterns was interesting because of the number and letter processing tasks underpinning that task in addition to cognitive flexibility. There was a severe floor effect on the distribution of scores in the letter sequencing task of the D-KEFS Trail-Making task (See Figure 6). This pattern was interesting because there was a small negative skew on the number sequencing task. Therefore, this may reflect underlying issues with the letter/alphabet skill because the participants have RD. Because of the severe platykurtic behaviour, namely from scores four to twelve, the high number of scores at the low end (i.e., 1 or 2) is not technically skew behaviour although a large amount clusters at the bottom.

Figure 4

Histograms and Boxplots of the Distribution of Number and Letter Sequencing



Note. Histograms and boxplots depicting the spread of scores on the number (a, b) and letter (c, d) sequencing tasks in D-KEFS Trail-Making.

Relationships Between EF and Reading

EF and Pretest Reading

My second research question asked, “To what extent do participants’ executive functions at pretest predict reading skill at pretest?” were identified between EF measures and certain reading measures at pretest (See Table 5). Without adjustment, 67 out of 248 correlation coefficients (27%) of the correlations between pretest reading and EF were statistically significant. Bonferroni corrected alpha level, .05/248 = .0002 was deemed too conservative given the number of variables involved, so the Benjamini-Hochberg procedure was utilized, ranking each pre-corrected significant *p*-value from 1-67. Following this procedure, 16 correlations (6.5%) remained significant (shaded darker grey in the table; See Appendix B).

Table 5

Correlations Between Pretest Reading and EF

	Decoding			Fluency		Comprehension		
	WJ III LWI	WJ III WA	SRI-2 WRA	TOWRE SWE	TOWRE PDE	WJ III RF	WJ III PC	SRI-2 PC
Inhibition								
D-KEFS CWI-I Raw	.05	.05	–.14	–.03	–.03	–.05	.01	–.05
D-KEFS CWI-I Scaled	.01	.05	.17	.05	.05	.12	.09	.13
D-KEFS CWI-I Contrast	–.15	–.12	–.17	–.31**	–.07	–.21*	–.12	–.03
BRIEF Inhibit Raw	–.11	–.15	–.13	–.16	–.08	–.10	–.09	–.17
BRIEF Inhibit T	–.19*	–.22*	–.19*	–.18*	–.15	–.18	–.18	–.23*
Shifting								
D-KEFS TM Raw	–.19*	–.12	–.12	–.19*	–.18	–.18	–.32**	–.30*
D-KEFS TM Scaled	.08	.10	.05	.10	.10	.10	.23*	.26*
D-KEFS TM Contrast	.03	.11	.01	.01	.09	–.07	.14	.11
D-KEFS CWI-S Raw	–.07	–.03	–.17	–.19*	–.11	–.17	–.08	–.09
D-KEFS CWI-S Scaled	.03	.06	.15	.17	.08	.18*	.09	.09
D-KEFS CWI-S Contrast	.02	.01	.00	.16	.04	.10	.00	–.04
BRIEF Shift Raw	–.08	–.15	–.05	–.13	–.02	–.09	–.10	–.10
BRIEF Shift T	–.25*	–.23*	–.15	–.21*	–.10	–.23*	–.23*	–.20*
Working Memory								
DSB Raw	.11	.21*	.11	.09	.10	.04	.20*	.16
DSB Scaled	.08	.20*	.11	.06	.06	.03	.17	.10
BRIEF Working Memory Raw	–.18	–.08	–.11	–.31*	–.16	–.24*	–.15	–.21*
BRIEF Working Memory T	–.23*	–.13	–.15	–.37**	–.22*	–.30*	–.21*	–.25*
Emotional Control								
BRIEF Emotional Control Raw	–.03	–.07	–.11	–.12	.02	–.06	–.06	–.12
BRIEF Emotional Control T	–.15	–.14	–.18	–.17	–.05	–.15	–.16	–.17
Reasoning and Metacognition								
D-KEFS CS1 Raw	.25*	.09	–.10	.10	.12	.21*	.28*	.35**
D-KEFS CS1 Scaled	.25*	.12	–.06	.03	.09	.23*	.29*	.38**
D-KEFS CS2 Raw	.20*	.08	–.01	.10	.16	.08	.27*	.35**

D-KEFS CS2 Scaled	.25*	.15	.03	.08	.17	.15	.32**	.43**
BRIEF Initiate Raw	–.06	–.06	–.15	–.17	–.09	–.08	–.10	–.13
BRIEF Initiate T	–.17	–.14	–.22*	–.26*	–.17	–.17	–.21*	–.19*
BRIEF Monitor Raw	–.17	–.14	–.20*	–.25*	–.08	–.20*	–.20*	–.17
BRIEF Monitor T	–.19*	–.18*	–.22*	–.25*	–.12	–.23*	–.22*	–.17
BRIEF Plan/Organize Raw	–.17	.02	–.14	–.27*	–.06	–.21*	–.09	–.12
BRIEF Plan/Organize T	–.25*	–.04	–.18	–.39**	–.12	–.28*	–.16	–.17
BRIEF Organize/Materials Raw	–.08	–.06	–.15	–.16	–.11	–.12	–.07	–.08
BRIEF Organize/Materials T	–.12	–.09	–.16	–.20*	–.12	–.15	–.12	–.07

Note. Raw reading scores are used in rows with EF raw scores, and standard reading scores are used in rows with EF standard or contrast scores.
***p*<0.0002-0.0131 (2-tailed) based on Benjamini-Hochberg rankings. **p*<0.05 level (2-tailed).

EF and Reliable Change Composite Z-scores

Despite there being several significant correlations between EF and pretest reading, when calculating reliable gains after the intervention (See Table 6), most relationships were non-significant. BRIEF Initiate was the only EF measure that significantly correlated with any reliable change composite *z*-score. BRIEF Initiate T scores negatively correlated with composite fluency gains (r=.21; p<.05), although the significance would not persist following Bonferroni correction, .05/68 = .007. Scatterplots for this correlation is found in Figure 5.

Table 6

Correlations Between Reliable Change in Reading Composites and Standard Score EF

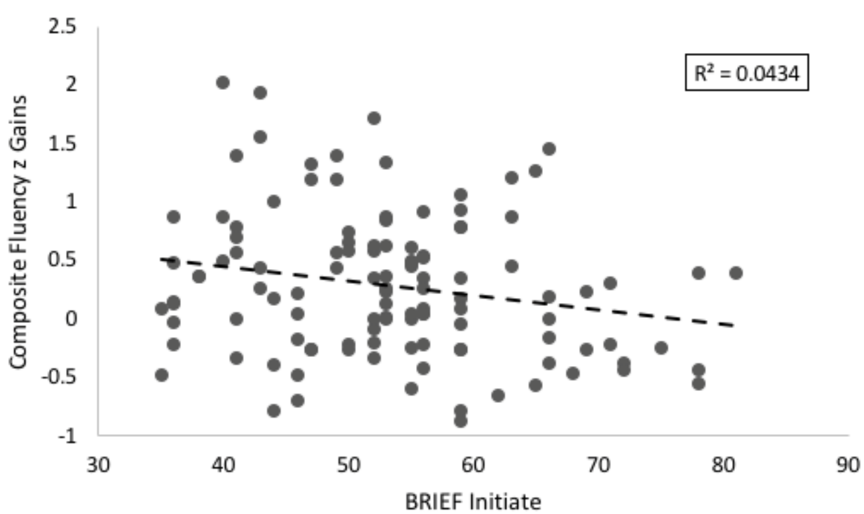
	Overall	Single-Word Decoding	Fluency	Text Comprehension
Inhibition				
CWI Inhibit	.12	.13	.05	.17
CWI Inhibit Contrast	.05	.07	–.14	.15
BRIEF Inhibit	–.03	.00	–.17	.09
Shifting				

Trail Making	.00	–.04	–.05	.11
Trail Making Contrast	–.04	–.02	.05	.01
CWI Shift	.17	.03	.14	.17
CWI Shift Contrast	.07	–.10	.11	.01
BRIEF Shift	–.05	–.01	–.11	.11
Working Memory				
Digit Span Backward	.05	–.09	.16	–.03
Working Memory	–.18	–.10	–.16	–.03
Emotional Control				
BRIEF Emotional Control	.03	.05	.00	.09
Reasoning and Metacognition				
Card Sort 1	–.05	–.01	–.02	.02
Card Sort 2	–.07	.06	–.08	.13
BRIEF Initiate	–.13	–.03	–.21*	–.04
BRIEF Monitor	–.04	.04	–.09	.08
BRIEF Plan/Organize	–.13	–.06	–.11	–.11
BRIEF Organize/Materials	–.16	–.06	–.16	–.08

*p<0.05 level (2-tailed).

Figure 5

Scatterplot of BRIEF Initiate T Scores and Composite RCI Fluency

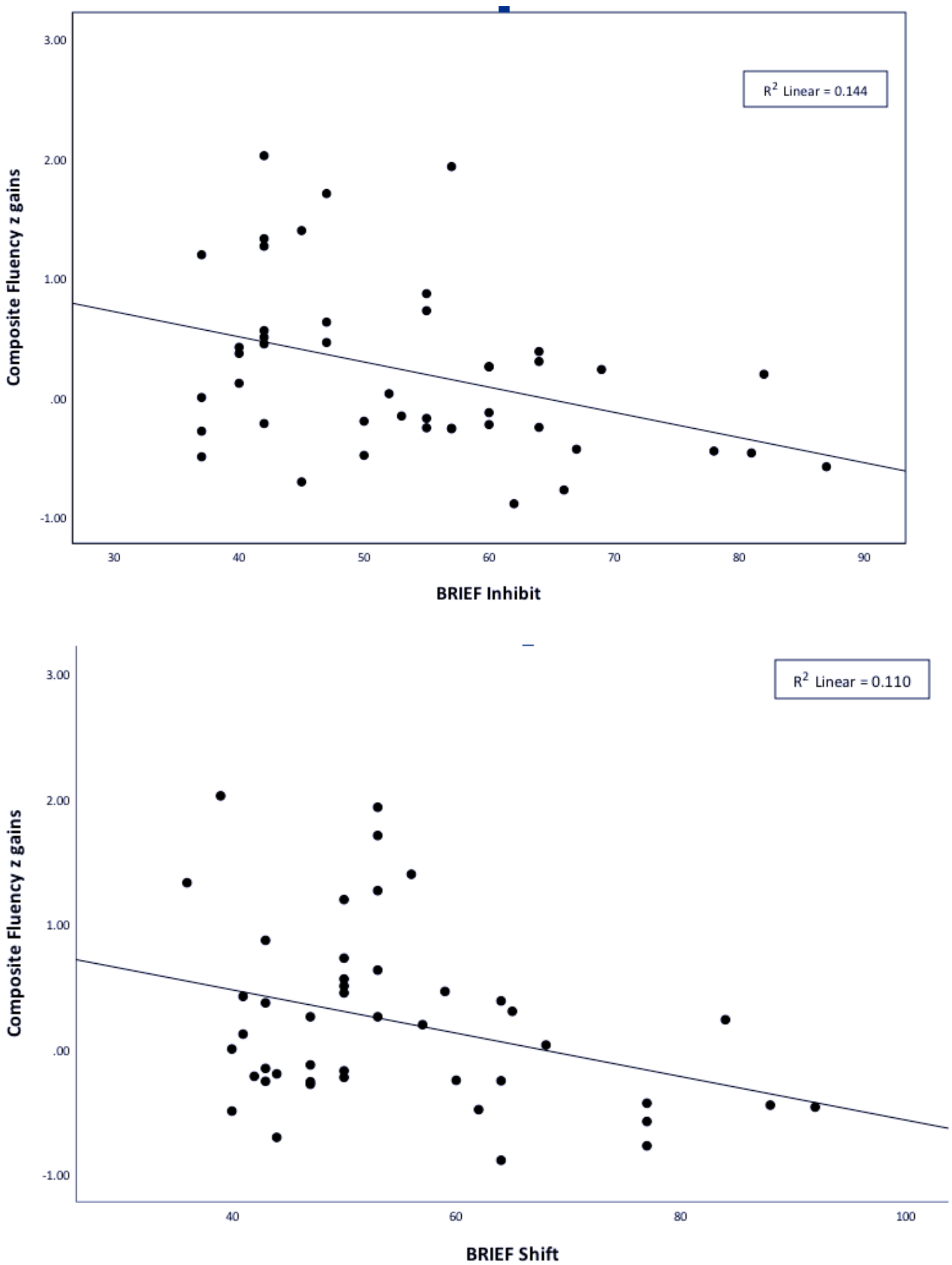


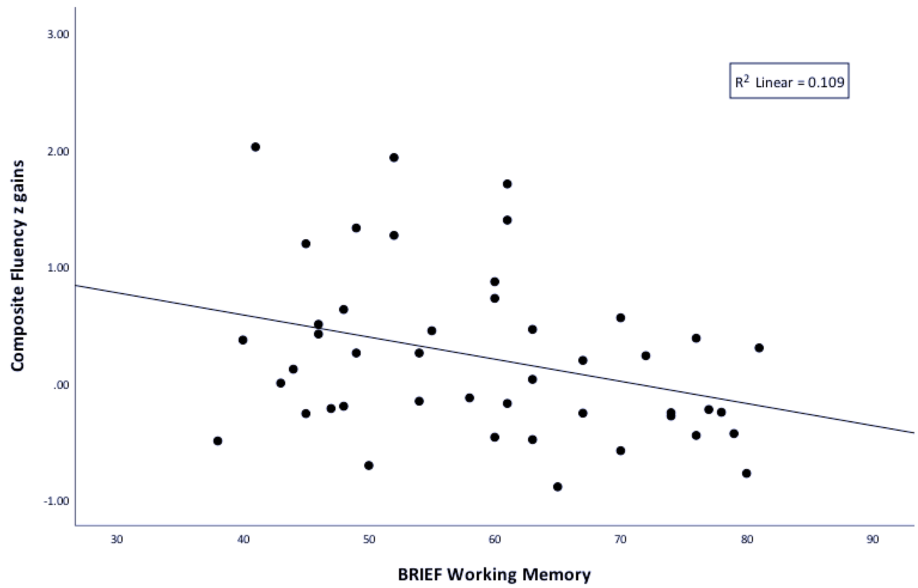
Note. T scores were used for the BRIEF measures. A higher T score indicates more EF difficulty.

When the correlations between EF and reliable gains were broken down by grade, the results mostly remained non-significant, but the variables that were significantly correlated differed. First, the correlation between BRIEF Initiate and composite fluency gains was not significant when examining Grade 3 (r=–.14; p=.26) and Grade 4 (r=–.17; p=.07) separately. In Grade 3 participants, the only significant correlation was between composite fluency gains and DSB (r=.24; p<.05), although this significance did not persist following the same Bonferroni correction applied earlier (2-tailed alpha = .007). For Grade 4 participants, the three significant correlations that emerged were between composite fluency gains and BRIEF Inhibit (r=–.38; p<.05), Shifting (r=–.33; p<.05), and Working Memory (r=–.33; p<.05). However, once again, the significance did not persist following Bonferroni correction. Scatterplots for BRIEF Inhibition, Shifting, and Working Memory are found in Figure 6.

Figure 6

Scatterplots of Grade 4 Fluency Gains and BRIEF Inhibition/Shifting/Working Memory





Note. T scores were used for the BRIEF measures. A higher T score indicates more EF difficulty.

Composite EF Deficit Variable

EF Intra-Correlations

Correlations between D-KEFS and DSB raw scores are seen in Table 7.

Table 7

D-KEFS/DSB Raw Score Correlations

	TM	CWI-I	CWI-S	CS1	CS2
CWI Inhibit	.38*				
CWI Shift	.44**	–.65**			
Card Sort 1	–.12	–.03	–.01		
Card Sort 2	–.33**	<.01	–.05	.59**	
DSB	–.07	–.04	–.16	–.08	–.06

***p*<0.01 (2-tailed). **p*<0.05 level (2-tailed).

All BRIEF Raw scores were highly correlated with one another, with the lowest correlation being Emotional Control and Plan/Organize (r = .43; p<.001), and the highest being Initiate and Working Memory (r=.76; p<.001). Nearly identical results were found with BRIEF T-Scores. Further, as shown in Table 8, there were significant correlations between D-KEFS/DSB and BRIEF measures, although few in proportion to the total amount of measures (10 out of 72), yet interesting nonetheless given how they were concentrated in a certain of measures (i.e., primarily with CWI scores). Aside from the CWI-I Contrast score, all correlations were in the direction where lower D-KEFS scores was linked to higher BRIEF scores. Because low D-KEFS scores and high BRIEF scores both indicate relative dysfunction on their respective constructs, higher dysfunction on D-KEFS was correlated with higher dysfunction on BRIEF.

Table 8

Standard Score Correlations Between BRIEF and D-KEFS/DSB

	CWI-I	CWI-I Contrast	TM	TM Contrast	CWI-S	CWI-S Contrast	CS1	CS2	DSB
Inhibit	–.02	.21*	–.05	–.02	–.05	–.04	–.09	–.18*	–.08
Emotional Control	–.08	.08	–.05	–.06	–.03	.05	–.02	–.15	–.04
Shift	–.09	.16	–.05	–.04	–.06	.04	–.03	–.13	–.06
Working Memory	–.14	.18*	–.10	.05	–.18*	–.07	–.06	–.16	–.09
Initiate	–.17	.16	–.16	–.03	–.21*	–.07	–.02	–.14	–.24**
Monitor	–.06	.16	–.09	–.04	–.10	–.06	–.05	–.11	–.14
Plan/Organize	–.19*	.15	–.11	.09	–.23*	–.06	–.12	–.15	–.16
Organize/Materials	–.18*	.17	–.11	.01	–.19*	–.02	.06	–.07	–.15

**p<0.01 (2-tailed). *p<0.05 level (2-tailed).

Thus, the sample was divided into a subsample that performed consistently poor on EF measures to compare their reading gains with the remainder of the sample that did not perform consistently poorly on the key EF measures. Given the drastic intercorrelations among BRIEF and their greater deviance from normality, only the neuropsychological EF measures were incorporated. Specifically, the low-EF subsample was derived through the calculation of scoring one standard deviation below (i.e., scaled scores at or below 7) at least two of three key EF measures: (a) Digit Span Backward; (b) CWI Inhibition or Shifting; and (c) Card Sort 1 or 2. Using this formula, there were 29 participants (25.2%) who met the criteria, and 86 (74.8%) who did not.

T-tests

Independent samples *t*-tests were conducted to determine whether children with consistently poor EF, a defined by the composite EF deficit variable, made significantly less reading gains than participants. This comparison was done using the composite *z*-score gains measures (overall, decoding, fluency, and text comprehension). Assumptions of the homogeneity of variance were met each time, as determined by Levene's test for equality of variances ($ps > .05$). As seen in Table 9, results show that children with consistent deficits across multiple EF made significantly less gains in fluency, $t(113) = 2.19$, $p = .03$. This effect size was medium, Hedge's $g = 0.47$, 95% CI [0.04, 0.89].[8] Even so, this result did not remain significant following Bonferroni correction (alpha level = .05/4 = .0125). No group differences were found for the other measures ($ps > .05$).

Table 9

T-Test on Reliable Change in Reading Using EF Multi-Deficit Category

Composite *z* Gains	***t***	***df***	***p***
Overall	1.92	113	.06
Decoding	.81	113	.42

[8] Hedge's *g* was chosen instead of Cohen's *d* due to the unevenness of groups.

Fluency	2.19	113	.03
Text Comprehension	1.45	113	.15

Chi Square Analyses

To further examine the relationship between consistent EF deficits and reading gains, chi-square analyses were run to compare groups in terms of a categorical reliable change result both overall and for each reading measure. As seen in Table 10, the chi-square analysis returned mostly non-significant results, although WJ III RF was marginally significant (p =.051), which would not hold after Bonferroni correction (0.05/8 = .0063). Importantly, the multi-deficit EF composite measure was non-significant in distinguishing reading outcomes categorically.

Table 10

Chi-Square Results for Reliable Change Across Multi-Deficit Grouping

	χ^2	p
Decoding		
WJ III Letter-Word ID	.07	.79
WJ III Word Attack	.64	.42
SRI-2 WRA	.64	.42
Fluency		
TOWRE SWE	.16	.69
TOWRE PDE	2.4	.12
Comprehension		
WJ III Fluency	3.8	.051
WJ III Comprehension	.002	.96
SRI-2 Comprehension	2.2	.13

Discussion

Broadly, the present thesis study sought to understand the relationship between EF and reading intervention responsiveness in children with RD to extend understanding of the dynamics of EF and reading intervention. The main findings were that EF did not significantly correlate with responsiveness whether EF was calculated individually or as a composite multi-deficit variable. In this section, I will first discuss how the results address the original research questions and the extent to which each hypothesis was supported. Next, I will consider the strengths and limitations of the methods, including measures, design, analyses, and situate findings within theoretical and conceptual frameworks. Finally, I will reflect on the implications of the study findings in relation to research, policy, and practice.

RQ1: Reliable Change Outcomes Across Reading Measures and Demographics

My first RQ asked "To what extent do participants make reliable change across different reading measures? How might these outcomes vary across different demographics of the sample?" The fact that most participants (73%) made reliable reading gains confirms my hypothesis and complements previous studies using the PHAST intervention (Lovett et al., 2008; Lovett et al., 2017; Morris, et al., 2012). These overall gains in reading were spread out across the various measures, as reliable change rates for individual reading measures were lower. This pattern suggests that the gains reflect an accumulated improvement to several underlying reading skills rather than concentrated in a specific facet of reading skill. Contrary to my hypothesis, comprehension measures generally had higher rates of reliable change than decoding. Specifically, the WJ III RF (27.8%), WJ III PC (41.7%) and SRI-2 PC (35.7%) measures were higher than the non-timed decoding measures of WJ III LWI and Word Attack as well as SRI-2 WRA. This demonstrates that, despite the majority of the intervention being dedicated to decoding strategies, greater gains were made in comprehension than decoding accuracy. That being said, there were high reliable change rates on the timed assessments of fluent single-word decoding measures of TOWRE SWE (38.3%) and PDE (47.0%). This suggests that participants improved on their fluent single-word decoding skills more than they improved on their general decoding accuracy. This finding may be justified considering the intensive nature of the intervention which provided constant opportunities to practice reading, making the process more automatic, and automaticity of reading (whether decoding or comprehension) is the phenomenon underlying fluency.

The reason why more gains were made in comprehension compared to decoding accuracy may be because participants were further behind typically developing peers on comprehension compared to decoding, with the exception of SRI-2 WRA. However, this may explained by the fact that the SRI-2 decoding task was derived from oral word recognition amidst reading a longer set of passages, whereas the WJ III LWI and WA involved recognizing/decoding isolated words. In a similar vein, it is worth considering that, at both pretest and posttest, the sample was furthest behind their typically developing peers (based on the instruments' respective standard score median) in performance on the SRI-2 instrument, followed by performance on the TOWRE battery; performed was closest between participants and the normative sample on the WJ III battery. The reason for poorer relative performance on the SRI-2 may be because the way the two tasks were embedded in a longer text. Furthermore, the SRI-2 is flexible in the complexity of the passage, ranging from kindergarten to Grade 8 in reading level. In a study investigating a sample relatively older than the present sample (students at an average age rounded to 12 years old), Jacobson et al. (2017) found performance on SRI-2 WRA and PC standard scores fell in the range of 7, as opposed to 5 in the present study. This lower performance relative to developmental norms suggests that interpreting standards scores on SRI-2 is more difficult for younger samples because the wider range of reading skill.

Interestingly, it may be a laudable result that greater gains were made in fluency and comprehension, as research has discussed the problematic pattern of intervention focusing on decoding rather than fluency and comprehension (Compton et al., 2014) because the latter are skills more relevant for everyday functioning. There are other ways in which the interventions results connect with existing literature. Lovett et al. (2008) conducted a study comparing students whose first language was English versus another language in multiple interventions, including PHAST. They found that in both groups, participants in the PHAST intervention outperformed those in a business-as-usual special education remediation program. Furthermore, the fact that some reading measures were predicted by their pretest scores follows suit with many studies that have found positive correlations between higher pretest reading skill and higher gains (e.g., Lovett et al., 2017; Tran et al., 2011).

RQ2: EF at Pretest Predicting Reading Skill at Pretest

My second RQ was "To what extent do participants' executive functions at pretest predict reading skill at pretest? Which executive functions predict which reading outcomes?" Confirming my hypothesis, EF at pretest had strong positive correlations with reading skills at pretest. Here it was found that 65 out of 248 correlation coefficients (26.2%) between EF and pretest reading were statistically significant. I expressly hypothesized that higher EF was associated with higher reading skill. The result met expectations, as all correlations were in the direction such that higher pretest reading scores were associated with higher EF performance.[9]

[9] For BRIEF, this meant high pretest reading scores were linked to lower BRIEF T Scores because low both low raw and T Scores reflect higher executive functioning.

Further, unlike what would be seen if the results were due to chance, the significant correlations clustered in certain configurations, offering moderately interpretable patterns.

The most prominent associations were between the card sorting tasks and WJ III Letter-Word Recognition and Passage Comprehension as well as SRI-2 Passage Comprehension. The fact that fewer significant correlations were found between card sorting and WJ III Word Attack compared to WJ III LWI may be because the WA uses pseudowords, which may be more daunting because the task is less representative of how words are encountered in regular life. One salient result was that WJ III and SRI-2 Passage Comprehension pretest measures all had highly significant correlations with both card sort tasks (reaching the $p<.001$ significance level). Circling back to the Scarborough Reading Rope theory (Scarborough, 2001), the way reading comprehension involves a larger combination of language skills compared to single-word decoding may explain how these measures of abstract reasoning and problem-solving contributed the most. Further, card sorting may better represent the role of EF in reading because the task involves using abstract reasoning to solving problems amidst ambiguity and switching between multiples rules (sorting according to shape, size, colour, etc.) as opposed to two rules in Trail-Making (number sequencing versus letter sequencing) and CWI Shifting task (word was inside a box versus not inside a box).

Not all reading and EF measures showed notable patterns. For examples, TOWRE Phonemic Decoding Efficiency did not correlate with any EF. The EF measures that did not correlate with any reading scores were CWI-I, CWI-S, BRIEF Emotional Control, DSB, and Organization of Materials. The lack of results for EF and pretest TOWRE PDE may be explained by the fact that general classroom instruction may not utilize pseudowords to teach decoding compared to interventions. In addition to the fact that encountering pseudowords is a less natural task, there are two other possible reasons for this pattern: first, the pretest averages were lower in PDE than SWE; second, greater gains were found in PDE compared to SWE. The lack of association between EF and Emotional Control is not entirely surprising, but does contradict the results of Al Otaiba et al. (2002; Al Otaiba & Fuchs, 2006) who reviewed numerous studies and found that behavioural issues was a predictor of poor reading. However, their review was conducted in 2002, using primarily studies from the pre-2000s; special education may have improved mitigation of behavioural issues since then (e.g., professional development). The fact that CWI Shifting had far less results than Trail-Making aligns with the result that they did not significantly correlate with one another, suggesting that they tap into distinct components of shifting. Therefore, there appears to be various EF constructs that have far less association with reading skill levels prior to intervention participation and therefore would not likely be relevant in future studies on reading intervention responsiveness.

Overall, the presence of associations at pretest suggests that EF predicts how well reading skills are acquired before participating in an intensive reading intervention (i.e., how well one learns from the general education level), aligning with the well-supported research on how EF predicts academic outcomes in the general education level (e.g., Serpell & Esposito, 2016; Yeniad et al., 2013). These results support previous studies that found relationships between EF and reading (Carretti et al., 2009). Using an amalgamated sample of children 9 years old on average, Ober et al. (2020) conducted a meta-analysis with 65 studies, finding small to moderate associations between EF and decoding ability. Walda et al. (2014) found that 80% of correlations between EF and reading scores at pretest or posttest were significant. Further, their methods employed highly similar constructs such as Berg's Card Sorting Task, which is related to the Wisconsin Card Sorting Task that inspired the D-KEFS CS1 and CS2. They also found the most significant correlations with the card sorting task. Overall, these findings support existing understanding that students with poorer EF tend to be further behind their peers on reading skill before participating in an intervention, especially for passage comprehension and when poor EF is derived using complex shifting tasks such as card sorting.

RQ3: EF at Pretest Predicting Reading Gains Made During Intervention

My third RQ was "To what extent do children's executive functions at pretest predict reading gains made during intervention? Which executive functions predict which reading outcomes?" I predicted that EF scores would positively correlate with reading gains such that those with higher EF would show greater reading gains using the reliable change composite *z*-score gains. I further predicted that such a relationship would be most pronounced with measures of comprehension, and that shifting would make the strongest prediction based on previous studies (e.g., Horowitz-Krauss, 2023; Walda et al., 2014). Contrary to expectations, the only significant outcome was between BRIEF Initiate and the composite reading fluency gains variable. This result denotes that participants who demonstrate greater initiative in daily activities (as reported by parents) showed noticeably larger improvements in word and pseudoword reading fluency. When broken down by grade, results for Grade 4 participants were marginally more noteworthy in that there were significant correlations between composite fluency gains and BRIEF Inhibition, Shifting, and Working Memory. However, this should be interpreted cautiously given the high number of correlations computed (Simmons et al., 2011) and the fact that the significance diminished after *p*-value correction.

Regarding possible underlying mechanisms, there may be an explanation that considers a theoretical interaction between the BRIEF Initiate construct, referring to the ability to begin a task or solve problems independently, and the two TOWRE variables that comprise the fluency composite which involve automaticity of reading words and non-words. In particular, the scatterplot revealed that scores at the lower end of BRIEF Initiate (representing higher task initiation) had more variance in composite fluency gains, where as participants at the higher end (representing greater difficulty with task initiation) was clustered near the bottom on the composite fluency gains axis. Independent task initiation is part of the metacognition construct, and one of the goals of PHAST was to prepare students to metacognitively employ the five types decoding strategies learned by deciding which to use in certain circumstances. Therefore, there are two possible advantages in benefitting from the intervention for participants with greater independent task initiation. First, given EF was measured at pretest, participants with higher independent task initiation may be more prepared for the metacognitive nature of the intervention and thus engage with the activities more effectively. Second, participants with higher independent task initiation may exhibit independence in terms of practicing reading on their own, such as reading at home, where they could practice strategies learned in the intervention. Much research has demonstrated that consistent practice with reading, even re-reading books one already knows, is an excellent way to improve reading fluency (Ardoin et al., 2016; Ardoin et al., 2018). The latter explanation especially makes sense considering that the BRIEF scores were parent-report measures and thus most valid in inferring participants' EF ability as manifested in the home context.

It is also noteworthy that BRIEF Initiate T scores had a significant correlation with pretest TOWRE Sight Word Efficiency (r=–.26; p<.001), and SWE was one of the two measures contributing to the RCI fluency composite variable. The negative correlation makes sense because it means that higher reading gains were associated with higher EF (higher scores on BRIEF measures represent higher dysfunction). Furthermore, TOWRE SWE pretest scores significantly correlated with SWE gains, although this significance did not withstand Bonferroni correction. Therefore, participants with higher pretest SWE performance may have had an advantage in making SWE gains, and higher SWE pretest performance appears to be significantly more probable for those with higher BRIEF Initiate. The explanation of students' independent reading at home remains probable even when the relationship between BRIEF Initiate and fluency gains appears to be through an indirect advantage of making gains based on pretest Sight Word Efficiency. This is because students with higher independent task and strategy may have exhibited beneficial independent reading habits, such as more frequent and consistent reading during which reading strategies may be practiced. In turn, this practice may have benefitted their single-word decoding fluency prior to the intervention and positioned them to make greater gains on average than their peers specifically on single-word decoding fluency.

Despite significance failing after Bonferroni correction, it is also noteworthy how all three main EF did achieve a significant correlation with gains in at least one instance, all of which were with the composite fluency gains variable. The correlation coefficients were all weak, but all in the correct direction such that higher EF correlated with higher single-word decoding gains. For third graders, the significant link was with Digit Span Backward, indicating that participants with auditory working memory may play a role in intervention responsiveness. For fourth graders, the significant links were with BRIEF Inhibition, Shifting, and Working Memory, indicating that these EF abilities, as reported by parents, may play a role in intervention responsiveness. However, in all four of these cases, just like with BRIEF Initiate, the pattern is likely explained by the fact that there was also significant correlations between these measures and TOWRE SWE at pretest. Coupled with the fact that SWE at pretest predicted higher gains in

SWE suggests that there may be a subtle, indirect role of basic EF through association with pretest reading ability.

While these possibilities should be further explored, another explanation is that the correlation was significant as a result of Type I error. Aside from a minority of situations, the capacity of EF to predict reliable change in reading gains was absent from a vast array of EF and reading measures, suggesting that the PHAST intervention effectiveness was independent from participants' executive functioning. Explaining this pattern is the likelihood that difficulties with inhibiting automatic responses, switching from different rulesets, and retain instructions in memory appears to be mitigated, even though such a pattern is usually found in the regular classroom. In many ways, this inability to predict reading gains with EF matches with existing literature. Walda et al. (2014) found few significant correlations when examining reading gains to the extent that the significant correlations were not meaningful after adjusting *p*-values for the number of analyses run. The types of EF measurements largely align with the present study, such as focusing on basic-level EF of inhibition, shifting, and working memory.[10] However, the key differences was their lack of exploration of reading comprehension measures.

On the contrary, Wilson et al. (2021) found that EF was positively correlated with reading gains, although their study sample was younger (Grade 1 instead of Grade 3 and 4) and the intervention format was a computer game. Furthermore, the intervention duration was considerably shorter. One explanation for why significant results were not found is that the small heterogenous group setting in which the PHAST intervention was delivered may have mitigated the learning challenges that arise from executive dysfunction. Specifically, smaller groups mean

[10] Although higher-order EF measures were also used, such as planning and problem-solving, they were the minority.

less distractions for students and more time for teachers to support each individual student. Consequentially, it may be easier for the teacher to help students focus, remember instructions, and transition from one task to the next because of this small group structure, reducing demand on students' inhibition, working memory, and shifting respectively, enabling an equal playing field with regard to participation and benefitting from sessions. In sum, the present findings largely align with existing studies EF that may be elaborated in future studies to address different samples and intervention formats.

RQ4: Multiple EF Deficits in Relation to Reading Gains

My final RQ was "To what extent do multiple deficits in EF scores (i.e., below average standard scores across multiple EF constructs) relate to reading gains?" I expected that multiple deficits in EF in the form of a composite measures would relate to a lower proportion of reading gains, especially in comprehension. The only significant *t*-test result was with composite fluency gains. This result indicates that children who had consistently substantial EF deficits fared worse on average in developing fluent single-word/pseudoword reading capacity. Interestingly, the main noteworthy chi-square result was with the WJ III Reading Fluency approaching significance (p=.051), revealing that children with multiple EF deficits may less frequently surpass the threshold of making reliable gains in fluent reading comprehension.[11] This combination of results is intriguing because the TOWRE and WJ III fluency measures reflect different constructs – one detecting fluency of word and pseudoword decoding, and the other reflecting a fluency of comprehension. A multiple deficit in EF may reflect assorted implications that connect with intervention responsiveness, such as poorer engagement with the intervention

[11] For clarity, WJ III Reading Fluency measures fluency (i.e., automaticity) of reading comprehension rather than fluency of single-word decoding which is captured by the TOWRE measures.

or difficulties completing daily tasks, leading to less opportunities to practice reading strategies learned in the intervention.

Similar to the previous research question, these results generally supporting existing knowledge. The overall lack of results matches that found by Miciak et al. (2019), where a composite EF measure did not predict intervention responsiveness status, although it did differentiate the RD nonresponding group to the typically developing group. This connection demonstrates that, although EF predicts pre-intervention reading skill, as it does with many other academic skills (e.g., Yeniad et al., 2013), EF does not play a substantial role in intervention responsiveness. As mentioned earlier, the reasons why EF would play a role in intervention responsiveness would primarily be in relation to their ability to benefit from the lessons and practice within the intervention context.

Emergent Patterns

The previous section discussed possible explanations for how the results connect to my research questions and hypotheses. Building from these themes, there are additional layers of the present study results as well as further connections to the scholarly literature. For example, while reading gains typically were not predicted by EF, in a typically developing (i.e., non-RD) populations, the nature of EF and reading gains looks different. DeBruin-Parecki and Cartwright (2023) conducted a narrative comprehension intervention and found a statistically significant improvement compared to business-as-usual controls, and that greater performance in cognitive flexibility, as measured by a card sorting task, correlated with higher gains. Similarly, Gao et al. (2018) conducted a reading intervention for all students and found that cognitive flexibility, measured using card sorting task, influenced the results. In these cases, the goal was not to remediate RD, but to determine if EF influenced the effectiveness of a particular strategy meant to improve typically developing students' reading skill beyond the business-as-usual approach. It is noteworthy that the PHAST intervention was delivered in small groups, as opposed to whole classes, as was done with DeBruin-Parecki and Cartwright (2023) and Gao et al. (2018). Another sample difference is that both these studies with a smaller sample than the present study. Nonetheless, a key nexus to these studies in my results is the fact that cart sort scores (D-KEFS CS1 and CS2) presented greater amounts of correlations between EF and pretest reading skill and higher effect sizes, although in the present study the task emphasized abstract reasoning and problem solving by having participants determine sorting rules on their own rather than receiving them via direct instruction as is done with the Wisconsin Card Sorting task that inspired the other studies. While difference in sample composition and intervention group size may explain why EF was mitigated in the present study despite predicting reading gains in these other studies, the card patterns warrant further investigation into the role of EF in reading skill and intervention gains which appears to be mediated or moderated by pretest performance.

Test Performance Relative to Developmental Norms

Through conversion of EF and reading variables into standard score formats, there is an opportunity to consider how the sample performed relative to expected results in the wider population. Besides one measure, all standard scores for reading and EF measures were below their respective developmental average. (i.e., a standard score of 10 on D-KEFS or SRI-2, 50 on the BRIEF, and 100 on the WJ III or TOWRE; Delis et al., 2001; Woodcock et al., 2001; Torgesen et al., 2012). On the D-KEFS measures, scaled scores of 8.5-11.5 represent 'average' for one's age. Six and seven represents 'low average,' and 5 represents 'borderline.' In particular, all means for D-KEFS measures were below 10 besides the CWI Shifting mean, which rounded to 10, and BRIEF T-scores were in the average range of 50-57. Performing in the "low average" range may not seem problematic, as it still falls under the umbrella of average and does not yet meet the threshold of clinically concerning. However, the consistent pattern across this whole sample where these scores being in several units below the population average should be investigated further. This may suggest that there is a pattern of dyslexic children exhibiting poorer EF performance, which has been found in other studies (e.g., Walda et al., 2022) Similarly, means for the WJ III and TOWRE consistently fell below 100 at pretest and posttest. This is to be expected given how the sample consists of students with RD.

Neuropsychological EF Performance

The pattern that all D-KEFS timed measures had multiple participants who reached the maximum amount of time, skewing scores, speaks to two possibilities. First, the link between RD and poor processing speed (for a review, see Horowitz-Krauss, 2023) appears to be supported by how the trail-making number and letter sequencing tasks yielded remarkably poor results in relation to developmental norms. In particular, there was much greater floor effects on the letter sequencing task, which may reflect underlying issues with the letter/alphabet knowledge since the participants have RD. The severe platykurtic behaviour suggests the RD population struggles with these kinds of neuropsychological tasks and the measures are not ideal for analyzing with parametric tests.

Strengths, Limitations, and Significance

Measures

One limitation across all EF measures was that they were measured at pretest only. Thus, future studies may explore the possibility of measuring EF at posttest as well to determine whether certain children make more EF gains than others over a 14-week span, and if such a pattern has a relationship with reading gains. This approach was taken by Swanson & Jerman (2007) with working memory specifically.

D-KEFS. The D-KEFS measures were mostly correlated with one another, meaning they provided a relatively wide depiction of participants EF capacity. A particular advantage was the way that some tasks involved accuracy, whereas some involved speed. Additionally, the TM and CWI tasks involve letters and words respectively, meaning that while they do not specifically involve inhibition, shifting, or working memory in the context of decoding words, they do involve processing of letters and words in a manner that may advantageously draw from emergent literacy skills such as letter knowledge.

Processing time is another confound, as it has been studied as an EF construct in other projects (e.g., Horowitz-Krauss, 2023) but not the present study. If a student does not have enough time to articulate their comprehension, they may end up with an answer that does not fully reflect what they know. The confound is especially relevant for students with executive dysfunction, which was an area of interest for the present study. The results from the D-KEFS timed tests (Trail-Making, Colour-Word Interference) had a high number of participants who maxed out the completion time for the EF tasks. While completion time is directly related to the measurement of EF constructs, completion time is typically only incorporated into reading when evaluating reading fluency. Therefore, it is reasonable to assume that completion time for the reading comprehension tasks involve processing speed, whether it be processing of the question/prompt, or whether it be processing and thinking of one’s answer.

DSB. The DSB task did not predict reading gains and One limitation is that DSB only captured auditory-based WM, known as the phonological loop. As RD involves phonological deficits, the auditory component of WM likely is an important one. Nonetheless, as reading first requires one to first visually process letters, there may be relevance of other visuospatial working memory (VWM) tasks in a reading gains prediction model. A VWM task may provide a more fulsome picture of participants' domain-general WM capacity, which may play a role in reading, such as the "peel off" decoding strategy .

BRIEF. This parent-report measure had the advantage of capturing daily behaviours known by the parents that may be signs of executive dysfunction that may not be detected through neuropsychological screening. All BRIEF Raw scores were highly correlated with one another, the lowest correlation being Emotional Control and Plan/Organize (r = .43; p<.001), and the highest being Initiate and Working Memory (r=.76; p<.001). Nearly identical results were found with BRIEF T-Scores. This suggests that the BRIEF instrument measured EF mechanisms in a way that overlapped far more compared to the D-KEFS/DSB.

Design and Analysis

Sampling. One particular note of the racial composition of the sample is how the majority of participants were Black (72%). While the ability for these findings to represent that demographic are strong compared to studies with predominantly White samples, this interpretation is complicated by the fact that the USA, Black students often speak a dialect known as African American English (AAE; see Craig et al., 2009). Because the study was part of a larger neuroimaging project that included fMRI, the following exclusion criteria were mandatory: speaking English as a secondary language, low intellectual functioning, chronic absenteeism, hearing impairment, uncorrected visual impairment, severe emotional or psychiatric issues, chronic medical/neurological condition (e.g., seizure disorder), or certain metals in their bodies (e.g., braces, cardiac pacemaker, etc.). Many of these exclusions were required for safety purposes due to the neuroimaging equipment. However, future studies that do not involve such equipment would likely benefit from including these participants for a heterogenous sample to maximize ecological validity. While these exclusion criteria may help reduce confounds and simplify interpretations of the results, they leave a gap where the current evidence does not reflect those excluded. For example, a future study could investigate the role of EF with English language learners by including them in the sample, as was done in previous studies (e.g., Lovett et al., 2008). Overall, follow-up studies sampling from those excluded would be a laudable next step considering the inability to generalize results of the present study to those populations.

Children with chronic absenteeism were excluded from the study based on the logistics required to collect reliable and valid data. Nevertheless, future research should incorporate a design that is able to include these children while retaining reliability and validity. For example, studies could consider a design where absenteeism is accounted for via advanced statistical procedures, such as machine learning statistics. Alternatively, the study could be designed around a sample exclusively focused on these children with high absenteeism. Interestingly, one way that this exclusion factors speaks to the limitations of these results is that the phenomenon of chronic absenteeism may be related to executive functions. Yet, part of the purpose of this research is to better support students who experience barriers to learning how to read. Students with RD may face comorbidity with other conditions, whether it be traumatic brain injury, epilepsy, or any other condition. In the spirit of inclusion of marginalized students, it would be ideal for evidence on reading interventions to reflect all children with reading disability, not just those from specific demographic backgrounds. Therefore, future studies that do not use neuroimaging techniques should ensure they use more representative samples.

Standard and Raw Scores. Both raw and standard scores were used throughout analyses because they both offer different perspectives on the same construct. One interesting dilemma for the calculations of scores and subsequent analysis was the way that age and grade factor in using scaled scores. Specifically, standard scores make adjustments based on treating age categorically, rather than as a continuous variable. For instance, children in Grade 3 will have variance in their age, despite them being in the same grade. Likewise, someone in grade 4 could be extremely similar in age to a Grade 3 student by virtue of December and January birth dates, as theoretically they could be closer in age than to their other peers. In these early ages, the difference in age when examined continuously may have important developmental implications, especially how EF is substantially driven by natural developmental factors (Anderson, 2002) as opposed to reading which is a learned skill (Church et al., 2019). A key limitation in using standard scores is how the factors of age and gender are controlled for, whereas other factors are not. To compensate for this, future studies employing multiple regression or another multivariate statistical model may consider accounting for continuous age within said model rather than using standard scores which were created primarily for clinical purposes to assist the diagnostic process without the need for regression-based statistical controls.

Analyses. The analyses undertaken for the present study were primarily Pearson correlations, accompanied by *t*-tests and chi-square tests. However, considering reading comprehension under the Scarborough (2001) rope model entails multiple simultaneous cognitive skills (i.e., verbal reasoning, decoding, fluency, etc.), future studies should utilize multiple regression or other multivariate statistical procedures to explore shared versus unique variance across a range of relevant measures. Another limitation to the analysis is that ADHD grouping was inferred based on a questionnaire (SWAN; Swanson et al. 2001) rather than a formal diagnosis from a psychologist. Therefore, the grouping reflects behavioural patterns that may be caused by other factors (e.g., another mental disorder) that would otherwise be more prudently distinguished in the formal diagnostic process. Another point revealed through the scatterplots is that the significant correlations found between composite fluency gains and EF seem to be driven by clustering near the higher end of BRIEF scale (reflecting lower EF). Quantile regressions may help for accounting for a multitude of cognitive abilities simultaneously while extracting understanding about how the role of EF in intervention responsiveness might vary depending on participants' EF capacity relative to the rest of the sample.

Theoretical and Conceptual Frameworks

Executive function. Conceptualizations of EF differ across the literature, and so one natural limitation to this study is how EF was approached. One conceptual idea assumed in the present study is the notion of EF being domain-general. While EF was originally conceptualized as a domain-general faculty, research has evolved to frame EF in diverse ways, many of which are tied to more specific aspects of functioning, such as language-specific shifting (Gosselin & Sabourin, 2024). The present study approached EF on low and high levels, focusing on low levels and conceptualizing them as three main constructs of inhibition, shifting, and working memory, which has been similarly used in other reading intervention studies (e.g., DeBruin-Parecki & Cartwright, 2023). The focus on low-level EF through use of the D-KEFS and DSB was because low-level EF is more relevant for young children with less responsibility than adolescents or adults, when high level EF becomes more relevant, such as planning and organizing, constructs inferred through components of the BRIEF instrument used in the present study (Gioia et al., 2000). Nonetheless, future studies would benefit from further investigating the role of metacognition considering the substantial connection between BRIEF Initiate and fluency gains.

Another question debated in the scholarship of EF is the validity of neuropsychological measurements versus behaviour ratings (Barkley & Murphy, 2010; Wasserman & Wasserman, 2013). As noted by Koziol and Lutz (2013), the task of switching from letter to number (i.e., Trail-Making test) or sorting cards based on arbitrary shapes and colours are not in and of themselves natural tasks, meriting caution in interpreting the extent to which performance thereon may validly infer variance in the cognitive processes responsible for daily tasks. An area of EF minimally explored in the present study was the distinction between high and low level EF (Diamond et al., 2013). Under this model, low level EF include simple cognitive tasks, such as inhibition and working memory, with high level EF comprising more complex tasks, such as planning and problem solving. Both the D-KEFS sorting test and BRIEF metacognition scores from the present study captured some behavioural-rating derived measure of high-level EF, such as planning, organization, and problem solving. The correlations among the neuropsychological measures in particular demonstrated the differentiation between low and high levels of EF, supporting the diversity model, as has been done in previous studies (Crawford et al., 2008). Finally, EF supposedly has distinct domain-general and domain-specific manifestations. While initially conceptualized as domain-general, researchers have increasingly explored reading-specific executive functioning (e.g., Cartwright et al., 2010; Gosselin & Sabourin, 2024). Cartwright et al. (2017) found that the reading-specific shifting measure significantly predicted comprehension, but the domain-shifting general task failed to do so, suggesting that literacy research may find reading-specific EF as a more useful construct than domain-general EF.

Reading. The reading measures were triangulated across multiple instruments, addressing decoding, fluency, and comprehension. Additionally, the metacognitive dynamic of decoding skills emphasized in the intervention were captured using a specialized Challenge

Word Test that has shown high reliability in previous studies (Lovett et al., 2000b). One conceptual limitation is reading comprehension. While WJ III Reading Fluency and Passage Comprehension involved single sentence comprehension at a lower level, the SRI-2 undertook a more complex angle to comprehension by involving comprehension of increasingly complex passages and having open-ended answers. In order to investigate reading comprehension on a deeper level, a more grade-specific text could be explored for greater precision of the sample. In doing so, reconceptualizing comprehension beyond multiple choice or cloze assessment may provide various means for students to express their understanding (Adams et al., 2023). An additional factor affording strengths and limitations is the use of the reliable change index, which is useful in accounting for measurement instruments' reliability when determining if change is reliable (Frijters et al., 2013). Compared to other studies examining EF as a predictor of reading gains, the RCI method is a unique aspect of this study. Specifically, reliable change was helpful in accounting for reliability of each of the tests.

Intervention context. An area needed for further research is clarity around the circumstances of the intervention. While the MTSS approach, such as RTI, has been widely adopted in North American schools (Ontario Ministry of Education, 2013), but few papers in the literature formally situate their research in this framework. It is important to understand the interpretation of the intervention context because some studies require students to have already completed certain amount of remediation (e.g., Walda et al., 2014) whereas some interventions did not consider whether students had a certain amount of remediation before participating. These factors of interpretation are important because the more remediation a student had completed prior to the intervention, the more the participants' lack of response is indicative of an underlying problem (e.g., learning disability) manifesting as resistance to treatment as opposed to poor instruction. The present study screened participants to only include students who struggling with reading for at least one year and did not make significant gains as the result of the convention supports offered by the school. Therefore, the present results reflect students who persistently resistant to treatment.

Language. Another limitation concerns language systems. The intervention and the majority of literature on which this study was built was for English-speaking communities. Therefore, these findings will likely not effectively translate to other languages, especially different types of languages that do not use the Latin alphabet (Share, 2021). Therefore, studies investigating how EF relates to intervention responsiveness should be completed in these regions for more ecologically valid evidence.

Contributions to Education

From the standpoint of preventing RD, the fact that more Grade 3 students showed reliable gains in reading reinforces the fact that reading interventions should start as early as possible (e.g., Lovett et al., 2017). Furthermore, from the lens of inclusion in education, the present study demonstrated various points pertaining to accommodating executive dysfunction, such as in universal design for learning (UDL) and multi-tiered systems of support (MTSS). First, the fact that many EF positively correlated with pretest reading skills demonstrates how regular classroom learning has room to improve from a UDL perspective, because participants with poorer EF exhibited lower average pretest reading skill compared to their peers with higher EF, and these pretest scores largely reflects classroom reading instruction, or Tier 1 of MTSS. Aligning with UDL guidelines has been the goal of numerous reading interventions (e.g., Coyne et al., 2012; Daley et al., 2020). In the most recent UDL guidelines by CAST (2024), it is suggested that classroom learning should scaffold executive functioning, reducing the demand of lower-level EF. Second, the fact that reading gains made in an intensive reading intervention for students in Grade 3 and 4 did not significantly vary based on EF demonstrates how a highly structured intervention process may mitigate the role of EF in learning. In particular, the structure may establish consistent routines that differentially benefit students who struggle with EF by providing external scaffolding that may be absent or at least inconsistent in the general classroom context.

Certain ideals of UDL appear to be exemplified in the intervention by reducing the demand on EF, thereby addressing barriers to learning for students with poorer EF. Therefore, small group interventions wherein participants are of similar skill level of the intervention topic such that this delivery format may mitigate the EF demands that are often present typical classroom learning. This effect is supported by the lack of consistent association between participants' EF scores and reliable reading gains made in intervention despite a strong association between EF and pretest reading skill. Therefore, literacy interventionists should value the benefit of small group intervention, especially for students with poorer EF. While the present study provided support to the ideas of increasing learning through small groups, this does not necessarily mean that in the regular classroom setting small groups will always be helpful. In this particular context, there was a trained interventionist working with each group the entire time, minimizing distraction from peer interaction that may arise in unsupervised small groups. Similarly, in this context, groups members were assigned based on a similar skill level, meaning the interventionists' instructions were more targeted than if the groups were to have been assigned at random, in which case there would likely be greater variability in students' individual reading levels.

Implications and Future Directions

Persisting Questions

Despite triangulating participants' EF capacity as well as triangulating reading skills with multiple assessments, the present study did not produce any convincing evidence that EF is a prominent factor to be addressed in the implementation of intensive reading interventions. Although EF problems are typically more prevalent among students with RD (Cartwright et al., 2017; Daucourt et al., 2018; Doyle et al., 2018), the structured and systematic approach of the intervention appears to have mitigated the relationship between EF and learning that is frequently seen in the normal classroom environment (Center on the Developing Child at Harvard University, 2011; Serpell & Esposito, 2016). Nonetheless, based on evaluation of the present study's limitations and possible mechanisms for its results, the following future research questions would be of benefit to the field of reading intervention:

- Can the intervention effectiveness and the role of EF be replicated in a sample that includes students with medical barriers such as epilepsy or congenital heart disease that are excluded from neuroimaging studies? How might results differ using a sample of students whose first language is not English?
- To what extent might small and large intervention groups vary in the manner by which EF predicts intervention gains? Here, a quasi-experimental design would be most ethical by using groups that, due to their educational context, are already going to be divided into distinct group sizes (e.g., 3-5 versus 10-13) due to the fact that previous research has already indicated that smaller group size yields higher reading gains than larger ones (Vaughn et al., 2003).

- How might an EF intervention benefit students who struggle with EF, and how might it be connected to broader improvements to academic performance such as in reading skill? How might its effectiveness be inspected by applying the reliable change index, autoregression, and growth curve analysis to EF?
- How might home and independent reading practices connect with EF and reading skill? To what extent does socioeconomic status play a role in this connection?

Teaching Independent Reading Habits

Because of the significant correlation between participants' task and strategy initiation (BRIEF Initiate) and their gains in fluent word reading, it may be of benefit to further promote in schools the transferable skill of independent task initiation. Schools could teach explicit habits that support this skill, such as guidance for how to complete independent reading at home, which may be helpful to complement learning taking place at school, especially for student who may not have otherwise have strong independent reading habits. However, further investigation is warranted, especially regarding students' motivation to read due to the reciprocity of reading skill and motivation that may exacerbate gaps between adequate and struggling readers in the absence of appropriate intervention (McNamara et al., 2011).

Considering the notable amount of correlations between EF and pretest reading, the present study replicated findings that EF performance is associated with reading skill levels before providing intervention (Arrington et al., 2014; Jacobson et al., 2017). Given reading performance highly predicts academic performance (Yeniad et al., 2013) and other non-academic outcomes important for quality of life (Eloranta et al., 2019; Gibson & Kendall, 2010; Hulme & Snowling, 2016; Swanson et al., 2017; Willcutt et al., 2007), this study reinforces the already well-established understanding that educational policy should actively address the role of EF in learning in order to benefit students with poorer EF (Center on the Developing Child at Harvard University, 2011; Serpell & Esposito, 2016). Such factors to address should be the manner that EF manifests as a moderator of reading habits in tandem with factors such as reading motivation (Kambara & Lin, 2023) and beliefs about reading (Frijters et al., 2018).

Home Literacy and Mixed-Methods Research

To recognize factors impacting reading skill development that are external to the intervention, such as home literacy practices, it is worthwhile to contemplate how innovative research methods can build upon the present findings in subsequent studies.[12] First, on the side of quantitative research, we have access to high quality methods such as machine learning that can help analyze large data sets. This has already been done with studies on EF and RD intervention in a study by Walda et al. (2023). While their study did not find EF as a predictor of reading gains, there were many methodological distinctions from the present study context, such as a different language and different intervention style. Alongside the addition of capturing EF longitudinally (i.e., measuring at follow-up, not just pretest), which the present study did not do, machine learning may help address statistical conundrums such as the shaky reliability of EF measures such as contrast scores while still appreciating the various contributions of parallel cognitive processes such as motor speed (Crawford et al., 2008).

On the side of qualitative research, methods that "peer into" participants' lives outside of school may help to understand factors such as home literacy practices on a richer level.[13] As mentioned earlier, the reason why BRIEF Initiate was found to correlate with fluency gains may

[12] Pondering back on the scenario of the student unresponsive to intervention mentioned in the 'Introduction' section, I believe it is not merely a matter of scientific strategy, but a matter of ethical responsibility, to contemplate future possibilities of research and practice in recognition of the fulsomeness of contemporary research method options that can unlock new insights for how to support students facing learning difficulties.
[13] Relatedly, WJ IV has added a qualitative component to LWI.

be explained by independent reading practices. In a meta-analysis, Sénéchal and Young (2008) identified how home literacy factors have played a role in family literacy interventions. Parental involvement and access to literacy resources are all factors that were not captured in the present study, but should be further explored. For example, Swain and Cara (2017) divide home literacy practices into four categories: (a) family resources, (b) parental literacy behaviours and attitudes, (c) parental beliefs and understandings, and (d) family literacy activities and practices. Investigating these factors may help identify moderators of reading intervention responsiveness when an intervention transpires over a long period of time, such as with PHAST comprising fourteen weeks.

A related issue that supports the notion of investigating home literacy practices is the fact that many literacy scholars theorize that a purely cognitive lens of reading skill is literacy as a social practice, because the real-world act of reading is always done in social context just like all forms of language and communication (e.g., Purcell-Gates et al., 2004). Along these lines, digital literacy practices may also be considered to understand how student may be practicing their reading skill. Reading comprehension has been shown in various studies to vary between print and digital text formats, although many studies have found no difference (for a review, see Hare et al., 2024).[14] Depending on the type of digital text, there may be a stronger or weaker effect on the development of reading skill, which, if linked to other independent variables included in a study, may need to be factored into analyses. While the evidence that reliable gains in single-word decoding fluency (i.e., TOWRE composite) can be predicted by the ability to self-initiate tasks and problem-solving strategies (i.e., BRIEF Initiate) is not groundbreakingly salient, further exploration of students' independent reading habits with greater precision may prove

[14] For a review of reading interventions in a digital game format, see Ostiz-Blanco et al. (2021)

valuable by elucidating the broader social circumstances influencing their reading skill development. For example, a tailored parent-report questionnaire that captures specific independent reading and literacy habits may be a useful variable to measure to control for in future intervention responsiveness studies.

Therefore, there are at least two possibilities missing from the present study offer potential in subsequent studies. First is to measure EF and in a more comprehensive targeted way, such as measuring at follow-up, perhaps as a growth curve, and to employ a more precise battery of both domain-general and reading-specific measures, especially focusing more complex shifting tasks similar to the card sorts from D-KEFS. Second is to capture contextual moderators such as literacy practices peripheral to the intervention, including qualitative measures, and applying them using integrative mixed-methods (for existing qualitative research on dyslexia using students' voices, see Gibson & Kendall, 2010). These methodological approaches may enable deeper examination of how contextual practices of literacy may moderate or mediate the strength of certain factors, EF or otherwise, to predict intervention responsiveness.

Remediation of EF

Considering the lower standard scores on EF across the entire sample in the present study, and given the multiplicity of ways that MTSS is used in soft skills (e.g., behavioural skills), not just curriculum material (Fletcher et al., 2018), one crucial new direction encouraged by these findings could be to re-examine how EF is taught in schools. Preparing students for lifelong success resonates with the reasons why much research and resources are supposed to be invested into improving students' literacy outcomes (Ontario Human Rights Commission, 2022). Similarly, reading and EF are linked in the sense that they are both foundational skills that unlock other skills and knowledge, both academic and non-academic, throughout one's entire developmental course. Therefore, a vital question to ask is, should EF development be more explicitly and comprehensively addressed by schools through an RTI approach? While the PHAST intervention appears to have mitigated the effect of participants' EF on their reading gains, these results do not speak to the many other ways the EF influences students' lives, both academically and non-academically. Therefore, future studies should look into an EF intervention with frequent monitoring and a tiered approach as it is with reading because MTSS frameworks are beneficial in balancing the needs of all students with those who struggle the most (Fletcher et al., 2018). Similarly, EF interventions could be designed, as advocated by many scholars including those at the Centre on the Developing Child at Harvard University (2011) who suggest that the importance of EF warrants attention to EF intervention. While excessive executive dysfunction in children is clinically remediated when part of a set of pathological symptoms, such as part of a treatment plan for ADHD (American Psychiatric Association, 2022), schools typically do not provide EF remediation as an academic remediation. The Centre on the Developing Child at Harvard University (2011) claims:

> Although additional replications and evaluations will add important information, the current evidence base is strong enough to warrant systematic, scaled-up initiatives to teach executive function skills in early care and education programs that focus on vulnerable populations in conjunction with evidence-based curricula that promote early literacy and numeracy skills. (p. 12)

This quotation advocates for education systems to teach EF in an explicit manner, just as reading is successfully taught when undertaking a structured approach.

Other studies have looked at the effects of an EF intervention (e.g., Rueda et al., 2012). Yang et al. (2017) report two experiments in which working memory training, verbal and visuospatial respectively, were provided to an experimental group compared to a control group from a sample of Chinese children with dyslexia. In both experiments, the experimental group outperformed the control group on rapid naming as well as their specialty cognitive trait test, a visual rhyming task and an orthographic awareness test respectively. The generalization of these results in other languages needs to be explored in future studies. Although limited research exists on successful EF interventions, school boards could partner with researchers to ensure collaborative research projects that further pursue this possibilities to better support these students in the future. EF is a cognitive process more of an inherent ability where knowledge is not what is needed to increase the ability, but explicit practice. The present study investigated the way students with poorer executive functioning face certain learning barriers. Since the relationship between EF and learning seems to disappear when provided with an intensive intervention, potentially due to the highly structured nature of lessons and the closer attention by the adult educator (i.e., serving as an external support for EF), there is justification to explore ways to promote EF among students systematically, such as through MTSS. It would be particularly beneficial for EF intervention to employ a longitudinal research design to track not only short-term effects but also long term effects (Dunn & Banati, 2015), such as impact on high school grades or standardized testing.[15] Finally, researching EF remediation in tandem with a reading intervention may precipitate not only further understanding of mechanisms influencing intervention effectiveness, but may help illuminate psychometric explanations of reading skill measures and EF measures. The present study found the diverse outcomes across reading skill

---

[15] In the context of Ontario, Canada, the key standardized test of literacy in high school is the Ontario Secondary School Literacy Test (OSSLT), of which successful completion is a requirement for an Ontario Secondary School Diploma (https://www.eqao.com/the-assessments/osslt). A longitudinal study could link elementary school reading intervention responsiveness to results on these tests, providing richer understanding of long-term effects.

measures as well EF measures at pretest and posttest, such as the fact that pretest WJ III performance was far closer to the developmental average than pretest SRI-2 performance despite both involving similar reading constructs. By controlling for testing effects, understanding students' EF performance in a more dynamic fashion (i.e., via responsiveness to an EF intervention) may explain whether the link between pretest reading skill and EF is explained by the role of EF in performance on a reading assessment versus the role of EF in learning how to read, whether in the classroom or beyond.

Pause and Ponder

Try the following task:

When a word is in bold, say the word 'bold' instead of the word written. When a word is in italics, say the word 'italics' instead of the word written. When a word is underlined, say the word 'underline' instead of the word written.

underline *bold* italics *underline* bold
italics *bold* italics bold *underline*
italics **underline italics** *underline*

Could you envision proactive intervention of EF through tasks similar to this at the general classroom level? Might the advantages outweigh the disadvantages?

Conclusion

In sum, the present thesis explored the relationship between EF and gains in reading skill following an intensive intervention. The existing literature on EF comprises a diverse set of ways to define, categorize, and measure goal-driven aspects of cognition, and the methods of the present study triangulated some of the most common approaches. Similarly, participants' reading skill was captured as the core components of decoding, fluency, and comprehension, including both timed and untimed assessments. Reliability of the reading measures was incorporated by way of calculating scores with the reliable change index. The main pattern observed was that EF positively correlated with a notable amount of reading skill at pretest, but not in the form of gains. Future studies can address the gaps in the present study, such as sampling exclusions and the particular combinations of EF measures. Additionally, in recognition of home literacy practices, future studies may undertake a mixed-methods approach, consider evaluating EF from a growth curve perspective, and even consider EF interventions. The ways that the human brain stores, manipulates, and responds to information remains relatively elusive, and sometimes it seems that with more data come not more answers, but rather more questions. Nonetheless, I still believe learning is more of a science than an art, with this study revealing yet another piece of the puzzle, bringing us closer to our goals - making the acquisition of reading skill a truly inclusive pillar of our education system.

References

Al Otaiba, S., & Fuchs, D. (2002). Characteristics of children who are unresponsive to early literacy intervention: A review of the literature. *Remedial and Special Education, 23*(5), 300-316. https://doi.org/10.1177/07419325020230050501

Al Otaiba, S., & Fuchs, D. (2006). Who are the young children for whom best practices in reading are ineffective? An experimental and longitudinal study. *Journal of Learning Disabilities, 39*, 414–431. https://doi.org/10.1177/00222194060390050401

Altemeier, L. E., Abbott, R. D. & Berninger, V. W. (2008). Executive functions for reading and writing in typical literacy development and dyslexia, *Journal of Clinical and Experimental Neuropsychology, 30*(5), 588-606, https://doi.org/10.1080/13803390701562818

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). https://doi.org/10.1176/appi.books.9780890425787

Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology, 8*(2), 71-82. https://doi.org/10.1076/chin.8.2.71.8724

Ardoin, S. P., Binder, K. S., Foster, T. E., & Zawoyski, A. M. (2016). Repeated versus wide reading: A randomized control design study examining the impact of fluency interventions on underlying reading behavior. *Journal of School Psychology, 59*, 13–38. https://doi.org/10.1016/j.jsp.2016.09.002

Ardoin, S. P., Binder, K. S., Zawoyski, A. M., & Foster, T. E. (2018). Examining the maintenance and generalization effects of repeated practice: A comparison of three interventions. *Journal of School Psychology, 68*, 1–18. https://doi.org/10.1016/j.jsp.2017.12.002

Arrington, C. N., Kulesz, P. A., Francis, D. J., Fletcher, J. M., & Barnes, M. A. (2014). The contribution of attentional control and working memory to reading comprehension and decoding. *Scientific Studies of Reading, 18*, 325-346. https://doi.org/10.1080/10888438.2014.902461

Arrington, C. N., Malins, J. G., Winter, R., Mencl, W. E., Pugh, K. R., & Morris, R. (2019). Examining individual differences in reading and attentional control networks utilizing an oddball fMRI task. *Developmental Cognitive Neuroscience, 38*. https://doi.org/10.1016/j.dcn.2019.100674

Barkley, R. A., & Murphy, K. R. (2010). Impairment in occupational functioning and adult ADHD: The predictive utility of executive function (EF) ratings versus EF tests. *Archives of Clinical Neuropsychology, 25*(3), 157–173. https://doi.org/10.1093/arclin/acq014

Barnes, M. (2023). Executive functions: What are they good for? A perspective from intervention research. *Mind, Brain, and Education*. https://doi.org/10.1111/mbe.12371

Baron, I. S. (2000) Test review: Behavior rating inventory of executive function. *Child Neuropsychology, 6*(3), 235–238. https://doi.org/10.1076/chin.6.3.235.3152

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: Series B (Methodological), 57*(1), 289-300.

Brydges, C. R., Reid, C. L., Fox, A. M., & Anderson, M. (2012). A unitary executive function predicts intelligence in children. *Intelligence, 40*(5), 458-469. https://doi.org/10.1016/j.intell.2012.05.006

Burns, M., Appleton, J. J., & Stehouwer, J. D. (2005). Meta-analytic review of responsiveness-to-intervention research: Examining field-based and research-implemented models. *Journal of Psychoeducational Assessment, 23*, 381-394.

CAST (2024). *Universal Design for Learning Guidelines* (version 3.0). https://udlguidelines.cast.org

Carretti, B., Borella, E., Cornoldi, C., & De Beni, R. (2009). Role of working memory in explaining the performance of individuals with specific reading comprehension difficulties: A meta-analysis. *Learning and Individual Differences, 19*(2), 246-251. https://doi.org/10.1016/j.lindif.2008.10.002

Cartwright, K. B., Coppage, E. A., Lane, A. B., Singleton, T., Marshall, T. R., & Bentivegna, C. (2017). Cognitive flexibility deficits in children with specific reading comprehension difficulties. *Contemporary Educational Psychology, 50*, 33–44. https://doi.org/10.1016/j.cedpsych.2016.01.003

Cartwright, K. B., Marshall, T. R., Dandy, K. L., & Isaac, M. C. (2010). The development of graphophonological-semantic cognitive flexibility and its contribution to reading comprehension in beginning readers. *Journal of Cognition and Development, 11*(1), 61–85. https://doi.org/10.1080/15248370903453584

Centre on the Developing Child at Harvard University (2011). *Building the Brain's "Air Traffic Control" System: How Early Experiences Shape the Development of Executive Function:* Working Paper No. 11. http://www.developingchild.harvard.edu

Chapman, M. L. (2003). Phonemic awareness: clarifying what we know. *Literacy Teaching and Learning, 7*, 91-114.

Christensen, L. & Mendoza, J. L. (1986). A method of assessing change in a single subject: An alteration of the RC index. *Behavior Therapy, 17*, 305–308.

Christopher, M. E., Hulslander, J., Byrne, B., Samuelsson, S., Keenan, J. M., Pennington, B., DeFries, J. C., Wadsworth, S. J., Willcutt, E., & Olson, R. K. (2015). Genetic and environmental etiologies of the longitudinal relations between prereading skills and reading. *Child Development, 86*(2), 342-361. https://doi.org/10.1111/cdev.12295

Church, J. A., Cirino, P. T., Miciak, J., Juranek, J., Vaughn, S., & Fletcher, J. M. (2019). Cognitive, intervention, and neuroimaging perspectives on executive function in children with reading disabilities. *New Directions for Child and Adolescent Development*, 165, 25-54. https://doi.org/10.1002/cad.20292

Compton, D. L., Miller, A. C., Elleman, A. M., & Steacy, L. M. (2014). Have we forsaken reading theory in the name of "quick fix" interventions for children with reading disability? *Scientific Studies of Reading, 18*, 55–73. https://doi.org/10.1080/10888438.2013.836200

Coyne, M. D., Oldham, A., Leonard, K., Burns, D., & Gage, N. (2016). Delving into the details: Implementing multitiered K-3 reading supports in high-priority schools. *New Directions for Child and Adolescent Development, 154*, 67-85. https://doi.org/10.1002/cad.20175

Coyne, P., Pisha, B., Dalton, B., Zeph, L. A., & Smith, N. C. (2012). Literacy by design: A universal design for learning approach for students with significant intellectual disabilities. *Remedial and Special Education, 33*(3), 162-172. https://doi.org/10.1177/0741932510381651

Craig, H. K., Zhang, L., Hensel, S. L., & Quinn, E. J. (2009). African American English-Speaking Students: An Examination of the Relationship Between Dialect Shifting and

Reading Outcomes. *Journal of Speech, Language, and Hearing Research, 52*(4), 839–855. https://doi.org/10.1044/1092-4388(2009/08-0056)

Crawford, J. R., Sutherland, D., & Garthwaite, P. H. (2008). On the reliability and standard errors of measurement of contrast measures from the D-KEFS. *Journal of the International Neuropsychological Society, 14*(6), 1069-1073. https://doi.org/10.1017/S1355617708081228

Daley, S. G., Xu, Y., Proctor, C. P., Rappolt-Schlichtmann, G., & Goldowsky, B. (2020). Behavioral engagement among adolescents with reading difficulties: The role of active involvement in a universally designed digital literacy platform. *Reading & Writing Quarterly, 36*(3), 278–295. https://doi.org/10.1080/10573569.2019.1635545

Daucourt, M. C., Schatschneider, C., Connor, C. M., Al Otaiba, S. & Hart, S. A. (2018). Inhibition, updating, working memory, and shifting predict reading disability symptoms in a hybrid model: Project KIDS. *Frontiers Psychology, 9*(238). https://doi.org/10.3389/fpsyg.2018.00238

DeBruin-Parecki, A., & Cartwright, K. B. (2023). Supporting inferential comprehension in the preschool classroom: The roles of theory of mind and executive skills. *The Reading Teacher, 77*(2), 146–155. https://doi.org/10.1002/trtr.2223

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan Executive Function System (D-KEFS): Examiner's manual*. San Antonio, TX: The Psychological Corporation.

Delis, D. C., Kramer, J. H., Kaplan, E., & Holdnack, J. (2004). Reliability and validity of the Delis-Kaplan Executive Function System: An update. *Journal of the International Neuropsychological Society, 10*(2), 301–303. https://doi.org/10.1017/S1355617704102191

Denton, C. A., Hall, C., Cho, E., Cannon, G., Scammacca, N.,& Wanzek, J. (2022). A meta-analysis of the effects of foundational skills and multicomponent reading interventions on reading comprehension for primary-grade students. *Learning and Individual Differences, 93*, 102062. https://doi.org/10.1016/j.lindif.2021.102062

Diamond A. (2013). Executive functions. *Annual Review of Psychology, 64*. 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Dolean, D. D., Lervåg, A., Visu-Petra, L., & Melby-Lervåg, M. (2021). Language skills, and not executive functions, predict the development of reading comprehension of early readers: evidence from an orthographically transparent language. *Reading & Writing, 34*(6), 1491–1512. https://doi.org/10.1007/s11145-020-10107-4

Doyle, C., Smeaton, A. F., Roche, R. A. P., & Boran, L. (2018). Inhibition and updating, but not switching, predict developmental dyslexia and individual variation in reading ability. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.00795

Dunn, K. & Banati, P. (2015). Strength in numbers: How longitudinal research can support child development, UNICEF Office of Research - Innocenti, Florence. http://www.unicef-irc.org/files/upload/documents/glori-report.pdf

Eloranta, A., Närhi, V. M., Eklund, K. M., Ahonen, T. P. S., & Aro, T. I. (2019). Resolving reading disability—Childhood predictors and adult-age outcomes. *Dyslexia, 25*(1), 20–37. https://doi.org/10.1002/dys.1605

Fisk, J. E., & Sharp, C. A. (2004). Age-related impairment in executive functioning: Updating, inhibition, shifting, and access. *Journal of Clinical and Experimental Neuropsychology, 26*, 874–890.

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2018). *Learning disabilities: From identification to intervention*. Guilford Publications.

Fletcher, J. M., Shaywitz, S. E., & Shaywitz, B.A. (1999). Comorbidity of learning and attention disorders: Separate but equal. *Pediatric Clinics of North America, 46*(5), 885-897.

Fletcher, J. M., Stuebing, K. K., Barth, A. E., Denton, C. A., Cirino, P. T., Francis, D. J., & Vaughn, S. (2011). Cognitive correlates of inadequate response to reading intervention. *School Psychology Review, 40*(1), 3-22. https://doi.org/10.1080/02796015.2011.12087725

Fletcher, J. M, & Vaughn, S. (2009). Response to intervention: Preventing and remediating academic difficulties. *Child Development Perspectives, 3*(1), 30-37. https://doi.org/10.1111/j.1750-8606.2008.00072.x

Follmer, D. J. (2018). Executive function and reading comprehension: A meta-analytic review. *Educational Psychologist, 53*(1), 42-60. https://doi.org/10.1080/00461520.2017.1309295

Foorman, B. R., Francis, D. J., Fletcher, J. M., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology, 90*(1), 37-55. https://doi.org/10.1037/0022-0663.90.1.37

Friedman, N. P., & Gustavson, D. E. (2022). Do rating and task measures of control abilities assess the same thing? *Current Directions in Psychological Science, 31*(3), 262-271. https://doi.org/10.1177/09637214221091824

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General, 133*, 101-135.

Frijters, J. C., Lovett, M. W., Sevcik, R. A., & Morris, R. D. (2013). Four methods of identifying change in the context of a multiple component reading intervention for struggling middle school readers. *Reading and Writing: An Interdisciplinary Journal, 26*(4), 539–563. https://doi.org/10.1007/s11145-012-9418-z

Frijters, J. C., Lovett, M. W., Steinbach, K. A., Wolf, M., Sevcik, R. A., & Morris, R. (2011). Neurocognitive predictors of reading outcomes for children with reading disabilities. *Journal of Learning Disabilities, 44*(2), 150-166. https://doi.org/10.1177/0022219410391185

Frijters, J. C., Tsujimoto, K. C., Boada, R., Gottwald, S., Hill, D., Jacobson, L. A., Lovett, M. W., Mahone, E. M., Willcutt, E. G., Wolf, M., Bosson-Heenan, J., & Gruen, J. R. (2018). Reading-related causal attributions for success and failure: Dynamic links with reading skill. *Reading Research Quarterly, 53*(1), 127–148. http://doi.org/10.1002/rrq.189

Gao, T., Zhao, J., Dou, K., Wang, Y., Li, X., & Harrison, S. E. (2018). Impact of cognitive flexibility on rapid reading skills training outcomes for primary school students in China. *School Psychology International, 39*(3), 273–290. https://doi.org/10.1177/0143034318773787

García, J. R., & Cain, K. (2014). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of Educational Research, 84*(1), 74-111. https://doi.org/10.3102/0034654313499616

Gerst, E. H., Cirino, P. T., Fletcher, J. M., & Yoshida, H. (2017). Cognitive and behavioral rating measures of executive function as predictors of academic outcomes in children. *Child Neuropsychology, 23*, 381-407. https://doi.org/10.1080/09297049.2015.1120860

Gibson, S., & Kendall, L. (2010). Stories from school: Dyslexia and learners' voices on factors impacting on achievement. *Support for Learning, 25*(4), 187–193. https://doi.org/10.1111/j.1467-9604.2010.01465.x

Gioia, G., Isquith, P. K., Guy, S. C., and Kenworthy, L. (2000). *Behavior Rating Inventory of Executive Function*. Odessa, FL: Psychological Assessment Resources.

Gosselin, L., & Sabourin, L. (2024). Language-specific cognitive flexibility is related to code-switching habits and interactional context; domain-general cognitive flexibility is not. *Journal of Cognitive Psychology, 36*(6), 688-708. https://doi.org/10.1080/20445911.2024.2365463

Hogan, T. P., Catts, H. W., & Little, T. D. (2005). The relationship between phonological awareness and reading. *Language, Speech & Hearing Services in Schools, 36*(4), 285–293. https://doi.org/10.1044/0161-1461(2005/029)

Horowitz-Kraus, T. (2023). The role of executive functions in fluent reading: Lessons from reading acquisition and remediation. *Mind, Brain, and Education*. https://doi.org/10.1111/mbe.12357

Hulme, C., & Snowling, M. (2012). Learning to read: What we know and what we need to understand better. *Child Development Perspectives, 7*(1), 1–5. https://doi.org/10.1111/cdep.12005

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

Iverson, G. L., Karr, J. E., Terry, D. P., Garcia-Barrera, M. A., Holdnack, J. A., Ivins, B. J., & Silverberg, N. D. (2020). Developing an executive functioning composite score for research and clinical trials. *Archives of Clinical Neuropsychology, 35*(3), 312–325. https://doi.org/10.1093/arclin/acz070

Jacobson, L. A., Koriakin, T., Lipkin, P., Boada, R., Frijters, J. C., Lovett, M. W., Hill, D., Willcutt, E., Gottwald, S., Wolf, M., Bosson-Heenan, J., Gruen, J. R., & Mahone, E. M. (2017). Executive functions contribute uniquely to reading competence in minority youth. *Journal of Learning Disabilities, 50*(4), 422-433. https://doi.org/10.1177/0022219415618501

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Clinical and Consulting Psychology, 59*, 12–19.

Kambara, H., & Lin, Y.-C. (2023). Factors impacting reading motivation: Insights from the bioecological model of human development. *The Journal of Educational Research, 116*(5), 309–321. https://doi.org/10.1080/00220671.2023.2265881

Kinsella, G., Storey, E., & Crawford, J. R. (n.d.). *Executive function and its assessment*. Neupsy Key. https://neupsykey.com/executive-function-and-its-assessment/

Lam, E. A., & McMaster, K. L. (2014). Predictors of responsiveness to early literacy intervention: A 10-year update. *Learning Disability Quarterly, 37*(3). https://doi.org/10.1177/0731948714529772

Learning Disabilities Association of Canada (2015, March 2). *TO REVISE OR NOT TO REVISE: The Official LDAC Definition of Learning Disabilities Versus DSM-5 Criteria*. [Position paper]. https://www.ldac-acta.ca/official-definition-of-learning-disabilities/

Lee, K., Bull, R., & Ho, R. M. H. (2013). Developmental changes in executive functioning. *Child Development, 84*, 1933-1953. https://doi.org/10.1111/cdev.12096

Levesque, K. C., Kieffer, M. J., & Deacon, S. H. (2019). Inferring meaning from meaningful parts: The contributions of morphological skills to the development of children's reading comprehension. *Reading Research Quarterly, 54*(1), 63–80. https://doi.org/10.1002/rrq.219

Lovett, M. W., Barron, R. W., & Frijters, J. C. (2013). Word identification difficulties in children and adolescents with reading disabilities: Intervention research findings. In Swanson, H. L., Harris, K. & Graham, S. (Eds.), *Handbook of learning disabilities* (2nd ed., pp. 329–360). Guilford Press.

Lovett, M. W., De Palma, M., Frijters, J., Steinbach, K., Temple, M., Benson, N., & Lacerenza, L. (2008). Interventions for reading difficulties. *Journal of Learning Disabilities, 41*(4), 333–352. https://doi.org/10.1177/0022219408317859

Lovett, M. W., Frijters, J., Wolf, M., Steinbach, K. A., Sevcik, R. A., & Morris, R. D. (2017). Early intervention for children at risk for reading disabilities: The impact of grade at intervention and individual differences on intervention outcomes. *Journal of Educational Psychology, 109*(7), 889-914. https://doi.org/10.1037/edu0000181

Lovett, M. W., Lacerenza, L., & Borden, S. L. (2000a). Putting struggling readers on the PHAST track: A program to integrate phonological and strategy-based remedial reading instruction and maximize outcomes. *Journal of Learning Disabilities, 33*, 458-476. https://doi.org/10.1177/002221940003300507

Lovett, M. W., Lacerenza, L., Borden, S. L., Frijters, J. C., Steinbach, K. A., & De Palma, M. (2000b). Components of effective remediation for developmental reading disabilities: Combining phonological and strategy-based instruction to improve outcomes. *Journal of Educational Psychology, 92*, 263–283. https://doi.org/10.1037/0022-0663.92.2.263

Lovett, M. W., Lacerenza, L., Steinbach, K. A., & De Palma, M. (2014). Development and evaluation of a research-based intervention program for children and adolescents with reading disabilities. *Perspectives on Language and Literacy, 40*(3), 21.

Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A definition of dyslexia. *Annals of Dyslexia, 53*, 1-14. https://doi.org/10.1007/s11881-003-0001-9

McNamara, J. K., Scissons, M., & Gutknecth, N. (2011). A longitudinal study of kindergarten children at risk for reading disabilities: The poor really are getting poorer. *Journal of Learning Disabilities, 44*(5), 421-430.

McNeill, P., & Chapman, S. (2005). *Research methods (3rd ed.)*. Routledge.

Melby-Lervåg, M., Lyster, S. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin, 138*(2), 322–352.

Mercugliano, A., Bigozzi, L., De Cunto, A., Graziani, D., Pecini, C., Usai, M. C., Vecchi, S., & Ruffini, C. (2024). Which childhood predictive indices forecast reading and writing skills in school-age children: A systematic review. *Child Neuropsychology*, 1–36. https://doi.org/10.1080/09297049.2024.2347381

Miciak, J., Cirino, P. T., Ahmed, Y., Reid, E., & Vaughn, S. (2019). Executive functions and response to intervention: Identification of students struggling with reading comprehension. *Learning Disability Quarterly, 42*(1), 17-31. https://doi.org/10.1177/0731948714529772

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science, 21*(1), 8-14. https://doi.org/10.1177/0963721411429458

Monchi, O., Petrides, M., Strafella, A. P., Worsley, K. J., & Doyon, J. (2006). Functional role of the basal ganglia in the planning and execution of actions. *Annals of Neurology, 59*(2), 257-264.

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

Nelson, J. R., Benner, G. J., & Gonzalez, J. (2003). Learner characteristics that influence the treatment effectiveness of early literacy interventions: A meta-analytic review. *Learning Disabilities Research & Practice, 18*(4), 255–267. https://doi.org/10.1111/1540-5826.00080

Newcomer, P. L. (1999). *Standardized Reading Inventory* (2nd ed.). Austin, TX: Pro-Ed.

Niendam, T. A., Laird, A. R., Ray, K. L., Dean, Y. M. Glahn, D. C. Carter, C. S. (2012). Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cognitive, Affective and Behavioral Neuroscience, 12*, 241-268. https://doi.org/10.3758/s13415-011-0083-5

Nilvius, C., & Svensson, I. (2022). Efficacy evaluation of a full-scale response to intervention program for enhancing student reading abilities in a Swedish school context. *Reading and Writing, 35*, 1239–1264. https://doi.org/10.1007/s11145-021-10237-3

Norton, E. S., & Wolf, M. (2012). Rapid automatized naming (RAN) and reading fluency: Implications for understanding and treatment of reading disabilities. *Annual Review of Psychology, 63*, 427-452. https://doi.org/10.1146/annurev-psych-120710-100431

Ober, T. M., Brooks, P. J., Homer, B. D., & Rindskopf, D. (2020). Executive functions and decoding in children and adolescents: A meta-analytic investigation. *Educational Psychology Review, 32*(3), 735-763. https://doi.org/10.1007/s10648-020-09526-0

Ontario Human Rights Commission (2022). *Right to read report*. https://www.ohrc.on.ca/en/right-to-read-inquiry-report/introduction

Ontario Ministry of Education. (2013). *Learning for all: A guide to effective assessment and instruction for all students, Kindergarten to Grade 12*. Toronto, ON: Queen's Printer for Ontario. https://files.ontario.ca/edu-learning-for-all-2013-en-2022-01-28.pdf

Ostiz-Blanco, M., Bernacer, J., Garcia-Arbizu, I., Diaz-Sanchez, P., Rello, L., Lallier, M., & Arrondo, G. (2021). Improving reading through videogames and digital apps: A systematic review. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.652948

Packwood, S., Hodgetts, H. M., & Tremblay, S. (2011). A multiperspective approach to the conceptualization of executive functions. *Journal of Clinical and Experimental Neuropsychology, 33*(4), 456-470. https://doi.org/10.1080/13803395.2010.533157

Pelham, W. E. Jr, Gnagy, E. M., Greenslade, K. E., Milich, R. (1992). Teacher ratings of DSM-III-R symptoms for the disruptive behavior disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 31*(2), 210–218. https://doi.org/10.1097/00004583-199203000-00006

Peng, P. (2023). The role of executive function in reading development and reading intervention. *Mind, Brain and Education, 17*(4), 246–256. https://doi.org/10.1111/mbe.12375

Perdue, M. V., Mahaffy, K., Vlahcevic, K., Wolfman, E., Erbeli, F., Richlan, F., & Landi, N. (2022). Reading intervention and neuroplasticity: A systematic review and meta-analysis of brain changes associated with reading intervention. *Neuroscience and Biobehavioral Reviews, 132*, 465-494. https://doi.org/10.1016/j.neubiorev.2021.11.011

Poon, K. (2018). Hot and cool executive functions in adolescence: Development and contributions to important developmental outcomes. *Frontiers in Psychology*, 8. https://doi.org/10.3389/fpsyg.2017.02311

Purcell-Gates, V., Jacobson, E. & Degener, S. (2004). The seeming incommensurability of the social and the cognitive. In *Print Literacy Development: Uniting Cognitive and Social Practice Theories* (pp. 63-80). Cambridge, MA and London, England: Harvard University Press. https://doi.org/10.4159/9780674042377-007

Rabiner, D. L., Malone, P. S., & Conduct Problems Prevention Research Group. (2004). The impact of tutoring on early reading achievement for children with and without attention problems. *Journal of Abnormal Child Psychology, 32*, 273-284. https://doi.org/10.1023%2Fb%3Ajacp.0000026141.20174.17

Scammacca, N. K., Roberts, G., Vaughn, S., & Stuebing, K. K. (2015). A meta-analysis of interventions for struggling readers in Grades 4–12: 1980–2011. *Journal of Learning Disabilities, 48*(4), 369–390. https://doi.org/10.1177/0022219413504995

Scarborough, H. S. (2001). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory, and practice. In S. Neuman & D. Dickinson (Eds.), *Handbook for research in early literacy* (pp. 97 – 110). New York: Guilford Press.

Sénéchal, M., & Young, L. (2008). The effect of family literacy interventions on children’s acquisition of reading from kindergarten to Grade 3: A meta-analytic review. *Review of Educational Research, 78*(4), 880–907. https://doi.org/10.3102/0034654308320319

Serpell, Z. N., & Esposito, A. G. (2016). Development of executive functions: Implications for educational policy and practice. *Policy Insights from the Behavioral and Brain Sciences, 3*(2), 203-210. https://doi.org/10.1177/2372732216654718

Share, D. L. (2021). Is the science of reading just the science of reading English? *Reading Research Quarterly, 56*(S1), S391–S402. https://doi.org/10.1002/rrq.401

Silverman, R. D., Johnson, E., Keane, K., & Khanna, S. (2020). Beyond decoding: A meta-analysis of the effects of language comprehension interventions on K-5 students' language and literacy outcomes. *Reading Research Quarterly, 55*.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359-1366. https://doi.org/10.1177/0956797611417632

Soares, C. S., Guerra, A., Roy, A., Hazin, I., & Azoni, C. S. (2020). Developmental dyslexia and executive functions: Evidence on main evaluation methods. *Estudos de Psicologia (Natal), 25*(1), 1-9.

Stuebing, K. K., Barth, A. E., Trahan, L. H., Reddy, R. R., Miciak, J., & Fletcher, J. M. (2015). Are child cognitive characteristics strong predictors of responses to intervention? A meta-analysis. *Review of Educational Research, 85*(3), 395-429.

Swain, J. M., & Cara, O. (2017). Changing the home literacy environment through participation in family literacy programmes. *Journal of Early Childhood Literacy*, 1-28. https://doi.org/10.1177/1468798417745118

Swanson, J., Deutsch, C., Cantwell, D., Posner, M., Kennedy, J. L., Barr, C. L., Moyzis, R., Schuck, S., Flodman, P., Spence, M., Wasdell, M. (2001). Genes and attention-deficit hyperactivity disorder. *Clinical Neuroscience Research, 1*(3), 207–216. https://doi.org/10.1016/S1566-2772(01)00007-X

Swanson, H. L., & Jerman, O. (2007). The influence of working memory on reading growth in subgroups of children with reading disabilities. *Journal of Experimental Child Psychology, 96*, 249-283. https://doi.org/10.1016/j.jecp.2006.12.004

Swanson, E., Stevens, E.A., Scammacca, N.K. et al. (2017). The impact of tier 1 reading instruction on reading outcomes for students in Grades 4–12: A meta-analysis. *Reading and Writing, 30*, 1639–1665. https://doi.org/10.1007/s11145-017-9743-3

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2012). *Test of Word Reading Efficiency Second Edition*. Austin, TX: Pro-Ed.

Tran, L., Sanchez, T., Arellano, B., & Lee Swanson, H. (2011). A meta-analysis of the RTI literature for children at risk for reading disabilities. *Journal of Learning Disabilities, 44*(3), 283-295. https://doi.org/10.1177/0022219410378447

Vaughn, S., Linan-Thompson, S., Kouzekanani, K., Bryan, D. P., Sickson, S., & Blozis, S. A. (2003). Reading instruction grouping for students with reading difficulties. *Remedial and Special Education 24*, 301–315. https://doi.org/10.1177/07419325030240050501

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R., Denckla, M. B., & Levin, J. R. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology, 88*(4), 601–638. https://doi.org/10.1037/0022-0663.88.4.601

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101*(2), 192–212. https://doi.org/10.1037/0033-2909.101.2.192

Wagner, R., Torgesen, J.K., & Rashotte, C.A. (1999). *Comprehensive Tests of Phonological Processing (CTOPP)*. Austin, TX: Pro-Ed.

Wagner, R. K., Zirps, F. A., Edwards, A. A., Wood, S. G., Joyner, R. E., Becker, B. J., Liu, G., & Beal, B. (2020). The prevalence of dyslexia: A new approach to its estimation. *Journal of Learning Disabilities, 53*(5), 354-365. https://doi.org/10.1177/0022219420920377

Walda, S., Hasselman, F., & Bosman, A. (2023). Identifying determinants of dyslexia: An ultimate attempt using machine learning. *Frontiers in Psychology, 13*. https://doi.org/10.3389%2Ffpsyg.2022.869352

Walda, S. A., van Weerdenburg, M., Wijnants, M. L., & Bosman, A. M. (2014). Progress in reading and spelling of dyslexic children is not affected by executive functioning. *Research in Developmental Disabilities, 35*(12), 3431-3454. https://doi.org/10.1016/j.ridd.2014.08.013

Walda, S., van Weerdenburg, M., van der Ven, A., & Bosman, A. (2022). Literacy progress in children with dyslexia and the role of attention. *Reading & Writing Quarterly, 38*(1), 19-32. https://doi.org/10.1080/10573569.2021.1897910

Wanzek, J., Stevens, E. A., Williams, K. J., Scammacca, N., Vaughn, S., & Sargent, K. (2018). Current evidence on the effects of intensive early reading interventions. *Journal of Learning Disabilities, 51*(6), 612-624. https://doi.org/10.1177/0022219418775110

Wanzek, J., & Vaughn, S. (2007). Research-based implications from extensive early reading interventions. *School Psychology Review, 36*(4), 541-561. https://doi.org/10.1080/02796015.2007.12087917

Wanzek, J., & Vaughn, S. (2011). Is a three-tier reading intervention model associated with reduced placement in special education? *Remedial and Special Education, 32*(2), 167-175. https://doi.org/10.1177/0741932510361267

Wanzek, J., Vaughn, S., Scammacca, N. K., 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

Wasserman, T., & Wasserman, L. D. (2013). Toward an integrated model of executive functioning in children. *Applied Neuropsychology: Child*, 88-96. https://doi.org/10.1080/21622965.2013.748394

Wechsler D. (2003). *WISC-IV Technical and Interpretive Manual*. The Psychological Association, San Antonio, TX.

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence - Second Edition*. San Antonio, TX: Pearson.

Willcutt, E. G., Betjemann, R. S., Pennington, B. F., Olson, R. K., DeFries, J. C., & Wadsworth, S. J. (2007). Longitudinal study of reading disability and attention-deficit/hyperactivity disorder: Implications for education. *Mind, Brain, and Education, 1*(4), 181-192. https://doi.org/10.1111/j.1751-228X.2007.00019.x

Wilson, A., Ahmed, H., Mead, N., Noble, H., Richardson, U., Wolpert, M. A., & Goswami, U. (2021). Neurocognitive predictors of response to intervention with GraphoGame Rime. *Frontiers in Education (Lausanne), 6*. https://doi.org/10.3389/feduc.2021.639294

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III*. Itasca, IL: Riverside Publishing.

Yang, J., Peng, J., Zhang, D., Zheng, L., & Mo, L. (2017). Specific effects of working memory training on the reading skills of Chinese children with developmental dyslexia. *PLOS One*. https://doi.org/10.1371/journal.pone.0186114

Yeniad, N., Malda, M., Mesman, J., van IJzendoorn, M. H., & Pieper, S. (2013). Shifting ability predicts math and reading performance in children: A meta-analytical study. *Learning and Individual Differences, 23*, 1-9. https://doi.org/10.1016/j.lindif.2012.10.004

Appendix A

Significant Skewness and Kurtosis Results

		Raw Score		Standard Score	
		γ_1 (SE)	γ_2 (SE)	γ_1 (SE)	γ_2 (SE)
CWI Inhibition		.50 (.217)	--	–.47 (.217) [c]	--
CWI Shifting		.68 (.218)	--	–.50 (.218) [c]	--
CWI Shifting Contrast		--	--	--	1.15 (.433)
Card Sort 1		--	--	.79 (.216)	2.17 (.428)
Card Sort 2		--	--	--	–.851 (.428)
Digit Span Backward		--	1.25 (.440)	--	--
BRIEF Inhibition		5.15 (.217)	--	11.64 (.217)	--
BRIEF Monitor		3.95 (.217) [c]		11.55 (.217) [a, c]	
BRIEF Shifting		3.85 (.217)	--	12.60 (.217)	--
BRIEF Emotional Control		5.44 (.217)	--	12.11 (.217)	--
BRIEF Plan/Organize		--	--	--	1.43 (.430) [a]
BRIEF Organization/Materials		--	–1.01 (.430)	--	–1.01 (.430) [a]
WJ III Letter Word ID	**pretest**	--	--	–.73 (.216)	1.12 (.428)
	posttest	--	--	--	1.26 (.428) [a, b]
	gains	.471 (.216) [a, b]	1.22 (.428) [a, b]	.92 (.216)	3.50 (.428)
WJ III Word Attack	**pretest**	–1.04 (.216)	2.86 (.428)	–.79 (.216) [a]	1.47 (.428) [a]
	gains	2.51 (.216)	14.31 (.428)	2.23 (.216)	12.45 (.428)
WJ III RF	**pretest**	–.78 (.216)	1.96 (.428)	–1.20 (.216)	3.58 (.428)
	posttest	.463 (.216) [a, b]	--	--	--
	gains	1.62 (.216)	6.40 (.428)	1.41 (.216)	5.13 (.428)
WJ III PC	**pretest**	--	.93 (.428) [b, c]	--	1.86 (.428) [a]
TOWRE SWE	**pretest**	–.47 (.216) [a, b, c]	--	--	--
	posttest	–.72 (.216)	--	–.52 (.216)	--
TOWRE PDE	**pretest**	.64 (.216)	--	--	--
	posttest	.47 (.216)	--	--	--
SRI-2 WRA	**posttest**	--	--	–.49 (.216)	--
SRI-2 PC	**pretest**	.56 (.216) [b, c]	1.35 (.428) [b, c]	--	--

gains	--	--	--	1.28 (.428)

Note. All *p*s <.05 when deriving the z-score from the skewness or kurtosis statistic over the SE. Total number of variables with significant skewness or kurtosis = 67 (13 of these were conditions in the TM subtest that were not directly used in any analyses).
[a] Consistent with normality according to the Kolmogorov-Smirnov test. [b] Consistent with normality according to the Shapiro-Wilk test. [c] Significance of skewness or kurtosis disappeared in the *n*=115 adjusted sample.

Appendix B

Benjamini-Hochberg Rankings for Significant Correlations Between EF and Pretest Reading

Correlation	*p*	Rank	BT	Sig.
Standard CS2, SRI PC	2.24091E-06	1	.00020161	True
Standard BRIEF P/O, TOWRE SWE	1.98842E-05	2	.00040323	True
Standard BRIEF WM, SWE	3.49814E-05	3	.00060484	True
Standard CS1, SRI PC	3.71033E-05	4	.00080645	True
Raw CS2, SRI PC	0.000119232	5	.00100806	True
Raw CS1, SRI PC	0.000124282	6	.00120968	True
Raw TM, WJ III PC	0.000430028	7	.00141129	True
Standard CS2S, WJ III PC	0.000449808	8	.0016129	True
CWI-I Contrast, TOWRE SWE	0.000648181	9	.00181452	True
Raw BRIEF WM, TOWRE SWE	0.000710472	10	.00201613	True
Raw TM, SRI PC	.001	11	.00221774	True
Standard BRIEF WM, WJ III RF	.001	12	.00241935	True
Raw CS1, WJ III PC	.002	13	.00262097	True
Standard CS1, WJ III PC	.002	14	.00282258	True
Standard BRIEF P/O, WJ III RF	.002	15	.00302419	True
Raw BRIEF P/O, TOWRE SWE	.003	16	.00322581	True
Raw CS2, WJ III PC	.004	17	.00342742	False
Standard TM, SRI PC	.005	18	.00362903	False
Standard BRIEF Initiate, TOWRE SWE	.005	19	.00383065	False
Raw CS1, WJ III LWI	.006	20	.00403226	False
Raw BRIEF Monitor, TOWRE SWE	.006	21	.00423387	False
Standard CS1, WJ III LWI	.007	22	.00443548	False
Standard CS2, WJ III LWI	.007	23	.0046371	False
Standard BRIEF Monitor, TOWRE SWE	.007	24	.00483871	False
Standard BRIEF Shift, WJ III LWI	.007	25	.00504032	False
Standard BRIEF WM, SRI PC	.007	26	.00524194	False
Standard BRIEF P/O, WJ III LWI	.008	27	.00544355	False
Raw BRIEF WM, WJ III RF	.009	28	.00564516	False
Standard CS1, WJ III RF	.012	29	.00584677	False
Standard BRIEF Shift, WJ III WA	.012	30	.00604839	False
Standard BRIEF Shift, WJ III PC	.012	31	.00625	False
Standard BRIEF WM, WJ III LWI	.012	32	.00645161	False
Standard TM, WJ III PC	.013	33	.00665323	False
Standard BRIEF Inhibit, SRI PC	.013	34	.00685484	False
Standard BRIEF Monitor, WJ III RF	.014	35	.00705645	False

Standard BRIEF Monitor, WJ III PC	.016	36	.00725806	False
Standard BRIEF Shift, WJ III RF	.016	37	.00745968	False
Standard BRIEF Inhibit, WJ III WA	.017	38	.00766129	False
Standard BRIEF Initiate, SRI WRA	.019	39	.0078629	False
Standard BRIEF WM, TOWRE PDE	.019	40	.00806452	False
Standard BRIEF Monitor, SRI WRA	.021	41	.00826613	False
Raw BRIEF WM, SRI PC	.023	42	.00846774	False
Raw CS1, WJ III RF	.025	43	.00866935	False
Standard BRIEF WM, WJ III PC	.025	44	.00887097	False
Raw BRIEF P/O, WJ III RF	.026	45	.00907258	False
Standard BRIEF Shift, TOWRE SWE	.026	46	.00927419	False
Standard BRIEF Initiate, WJ III PC	.026	47	.00947581	False
Raw DSB, WJ III WA	.027	48	.00967742	False
CWI-I Contrast, WJ III RF	.028	49	.00987903	False
Raw BRIEF Monitor, SRI WRA	.029	50	.01008065	False
Raw CS2, WJ III LWI	.031	51	.01028226	False
Raw BRIEF Monitor, WJ III PC	.031	52	.01048387	False
Standard BRIEF Shift, SRI PC	.032	53	.01068548	False
Standard BRIEF O/M, TOWRE SWE	.033	54	.0108871	False
Raw BRIEF Monitor, WJ III RF	.034	55	.01108871	False
Raw DSB, WJ III PC	.037	56	.01129032	False
Standard DSB, WJ III WA	.037	57	.01149194	False
Raw TM, WJ III LWI	.038	58	.01169355	False
Standard BRIEF Inhibit, WJ III LWI	.038	59	.01189516	False
Standard BRIEF Initiate, SRI PC	.038	60	.01209677	False
Raw CWI-S, TOWRE SWE	.039	61	.01229839	False
Standard BRIEF Inhibit, SRI WRA	.04	62	.0125	False
Raw TM, TOWRE SWE	.046	63	.01270161	False
Standard BRIEF Monitor, WJ III LWI	.046	64	.01290323	False
Standard CWI-S, WJ III RF	.049	65	.01310484	False
Standard BRIEF Inhibit, TOWRE SWE	.049	66	.01330645	False
Standard BRIEF Monitor, WJ III WA	.05	67	.01350806	False