8-13-2019

The Relation Between Working Memory and Reading Intervention Outcomes in Children with Developmental Dyslexia

Rebecca Winter

doi: https://doi.org/10.57709/14864000

THE RELATION BETWEEN WORKING MEMORY AND READING INTERVENTION OUTCOMES IN CHILDREN WITH DEVELOPMENTAL DYSLEXIA

by

REBECCA WINTER

Under the Direction of Robin Morris, PhD

ABSTRACT

This study explored three theoretical models for understanding working memory (WM) in struggling readers. Research has shown relations between working memory and reading, especially on foundational phonologically based tasks like non-word repetition, however there is limited research exploring the predicted relations between non-language based WM tasks and basic reading abilities. Here three models of WM are used to explore the optimal theoretical framework for understanding both foundational reading abilities and response to intervention among a group of elementary school diagnosed with developmental dyslexia (DD). The three models were driven by theories from: 1. Baddeley and Hitch model; 2. Engle et al., model; and 3. Brown and Hulme model. 108 children (mean age: 9.01) in grades 3-4 were assessed as meeting

criteria for a reading disability and subsequently participated in a 70-hour intensive reading intervention. Children were administered nine working memory, intelligence, and language tasks prior to the intervention and then tested at four time points on four single word reading tasks at baseline, 23, 45 and 70 hours over the course of the intervention. Three theoretical models were fit to the resulting data using structural equation modeling. Model 1 (Baddeley and Hitch, 1974) revealed the best fit for the data (RMSEA = 0.04; Model Chi Squared = 43.65, p=.18 (df=33); CFI =0.96), with Language, Intelligence and Phonological Working Memory factors showing strong relations to initial reading scores, but no factor showed relations to single word reading intervention change scores. These results suggest that phonological WM, visuo-spatial WM, and central executive WM reflect separate but related constructs. These data also confirm the fundamental connection between single word reading and phonological working memory and supports previous data showing minimal support for cognitive assessment as a means of predicting outcomes on reading intervention.

INDEX WORDS: Developmental Dyslexia, Reading intervention, Working memory, Language, Phonological loop, Central executive

THE RELATION BETWEEN WORKING MEMORY AND READING INTERVENTION OUTCOMES IN CHILDREN WITH DEVELOPMENTAL DYSLEXIA

by

REBECCA WINTER

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2019

THE RELATION BETWEEN WORKING MEMORY AND READING INTERVENTION OUTCOMES IN CHILDREN WITH DEVELOPMENTAL DYSLEXIA

by

REBECCA WINTER

Committee Chair: Robin Morris

Committee: Lee Branum-Martin

Jan Frijters

Rose Sevcik

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2020

DEDICATION

This dissertation is dedicated to my husband, Josh Gottlieb. You have put up with me, supported me, and helped me keep everything in perspective every step of the way. Thank you for everything you do for all of us. I love you so much. Looking forward to the next great thing!

ACKNOWLEDGEMENTS

I am grateful to acknowledge a host of people without whom I could never have completed this paper or degree. First and foremost, my academic advisor, Robin Morris, has led me through the highs and lows of this rigorous program with honesty, steady guidance and tremendous empathy. He helped me through every unexpected and unconventional hiccup along the way, and I am certain that without his advocacy on my behalf, my family would not have maintained its sense of the relative calm that it did over these past 6 years. Similarly, my clinical supervisors at GSU and within the Emory and CHOA hospitals have provided me with excellent training and have encouraged me to attain a 'work-life-balance' that I never thought possible. Thank you all for reinforcing for me that the important parts of life often happen outside of work and school and that being a parent and a clinician can enhance the quality of care for patients, despite a slightly more hectic schedule.

My dissertation committee Drs. Sevcik, Branum-Martin and Frijters, has helped me understand nuances in literature, complex statistics (well, complex to me!), and delve deeper into my assumptions and theories espoused by previous research. I am grateful for your consistent availability, encouragement, critiques and for giving me the feeling of "you can do this!"

To my lab mates, current and past, Cortney, Cynthia, Emily, Phebe and Ally; I am so grateful for your friendship, collaboration and support over these years. I'm so going to miss our lab chats/complaining sessions ☺. Emily, I cannot believe that we are not continuing to this next stage together. You have provided me with support that goes above and beyond. You are a true friend, a stellar clinician and one of the best humans I have ever known. I am so excited to celebrate your future successes.

Finally, my family near and far have always been my biggest advocates. It has truly taken a village to get me through this degree and I could not have done it without your visits, babysitting, emotional support, phone calls, school pick ups and drop offs, moving to Atlanta (!!!!) and general awesomeness. Thank you.

Eitan, Alex and Ranan: You are the lights of my life. Sometimes I feel like I finished this degree in spite of you, but mostly I feel like you got me through it. Thank you for the happiness and pride that you bring me everyday.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

1.1 Working Memory and Reading

The Baddeley and Hitch (1974) model of working memory (WM) has, over time, been supported as the gold-standard conceptual model for describing working memory abilities and explaining its role in impairments across a wide range of clinical and developmental populations. As an example, injury to the phonological slave system, which is described as responsible for temporarily storing and manipulating sublexical and phonological units of language, has been used to explain deficits in neuropsychological patients displaying atypical fluent aphasias such as anomic and conduction aphasias (Baddeley, 2003). These findings have further influenced seminal studies focused on the developmental origins of language acquisition with researchers suggesting that the foundation of language acquisition lies in the strength of the phonological slave system. A limitation in the development or functioning of this system will inevitably lead to maladaptive vocabulary growth and language learning over the long term (Gathercole, 1995; Gathercole & Baddeley, 1989). Because developmental dyslexia (DD) has long been associated with deficits in the phonological slave system, it has been suggested that this deficit might drive, or underlie, difficulties in reading development (Brooks, Berninger, & Abbott, 2011; Ehri, 2014). Additional support for the role of WM in reading development has been found in recent reading intervention studies which have identified predictors of successfully remediated children, including WM (Al Otaiba & Fuchs, 2002; Frijters et al., 2011; Miciak et al., 2014).

Researchers interested in further specifying the underlying constructs of the original Baddley and Hitch (1974) theory have expanded or developed alternative models to this

classic WM model. For example, Engle, Tuholski, Laughlin, and Conway (1999) put forward a model conceptualizing WM as including executive control and emphasizing the need for increased attentional control. This ‘executive’ component of WM directly impacts short-term memory (STM) which would subsume both phonological and visual spatial components of the Baddeley and Hitch (1974) model, as well as relate to overall intellectual abilities. Thus, for Engle et al. (1999), phonological capabilities stem from this broader ability to manage attention.

Despite its widespread appeal in clarifying the basis for the relationship of language in reading development, the Baddeley and Hitch (1974) WM model has been opposed by language theorists who posit that the described phonological slave system is in fact synonymous with the language system, thus eliminating the role of WM in language acquisition (see Acheson & MacDonald, 2009; Baddeley, 2003; MacDonald, 2016). Although never directly studied in a DD population, language theorists would posit that reading difficulties stem from errors in overall language system development and functioning, including its phonological capabilities, but these language attributes are unrelated to WM (e.g., MacDonald, 2016). Based on these different conceptualizations of the relationship between language, phonological capabilities, WM, and reading, this project aims to (1) directly study the similarities and differences between three different WM and language constructs and models in a sample of formally diagnosed children with DD (see Figure 1), and to (2) evaluate the predictive power of these various WM models on reading scores following an intensive reading intervention, which will provide evidence for which model constructs are most critical in such learning outcomes.

Baddeley and Hitch (1974) Model of WM | Engle et al. (1999) model of WM | Brown and Hulme (1996) language driven model of WM

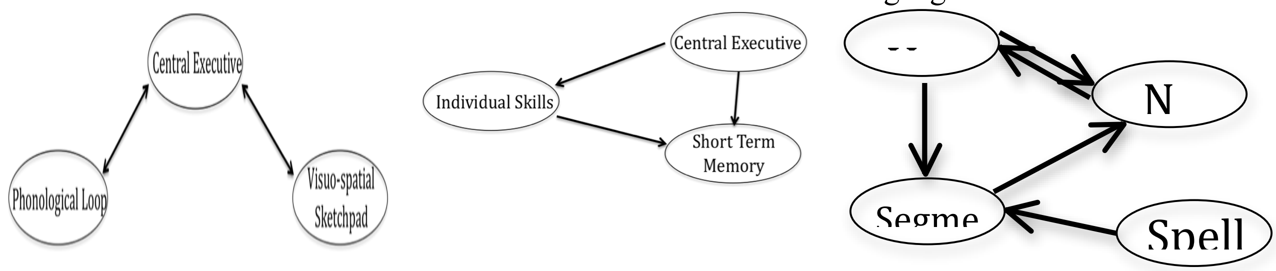


Figure 1: Three theoretical models for understanding working memory. Models are described in detail in subsequent sections.

1.2 Model 1: The Baddeley and Hitch Model of Working Memory

The Baddeley and Hitch (1974) model of WM proposes a tripartite WM system which includes two domain-specific slave systems, the phonological loop and the visuospatial sketch pad, both of which feed into the 'domain general component' of the model (Cowan et al., 2005), now known as the central executive (Baddeley, 2000; described also by: Coolidge & Wynn, 2005; Cowan et al., 2005). The phonological slave system is responsible for storing phonemic sounds and maintaining vocally or sub vocally rehearsed information. The visuospatial sketchpad has been shown to involve maintenance and integration of visual and spatial material. According to this model, both slave systems combine information to form a domain-general (not modality driven) component, the central executive, whose function is hypothesized to activate controlled attention, memory units, and perhaps divide and switch attention necessary for higher order cognitive processes (as described by Engle et al., 1999). The central executive has been shown to correlate with

tasks requiring the integration of information, regardless of modality, such as that required, for example, in reading and comprehension (Cowan et al., 2005).

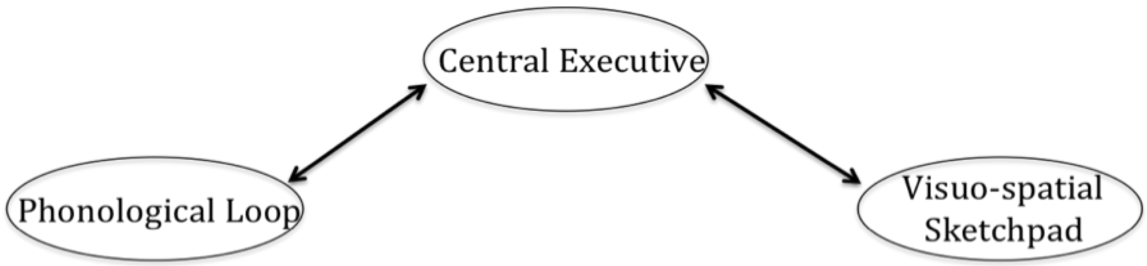


Figure 2: The Baddeley and Hitch (1974) model of working memory

Following 25 years of research largely supporting the tripartite model, Baddeley (2000) proposed a fourth component to the model: The episodic buffer (See Figure 3). This component attempted to account for findings of a temporary storage system capable of manipulating and holding complex information, beyond the assumed capacity of the two originally proposed slave systems. The episodic buffer has been approximated by tasks, which draw on previous exposure (with information stored in long-term memory) and repeated in real time during a WM task. Immediate story recall and sentence repetition have been frequently used as proxies for the episodic buffer (Dawes, 2015; Henry, 2010).

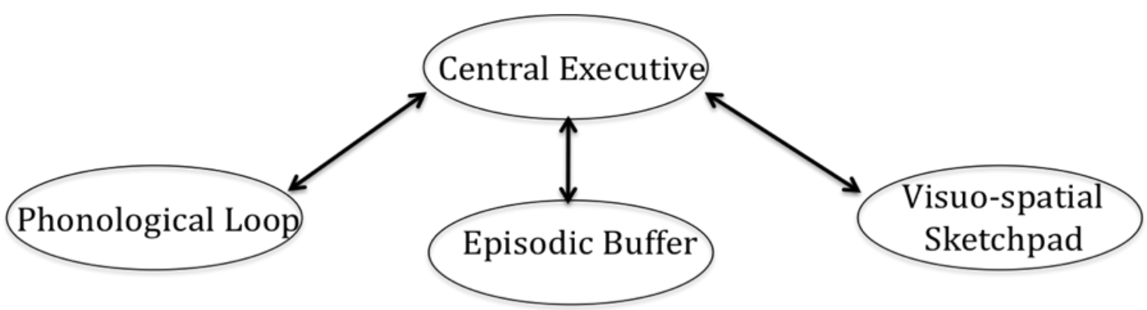


Figure 3: The four-component model as proposed by Baddeley (2000). The addition of the episodic buffer allows for the integration of information from a variety of sources for a short period of time.

1.3 Model 2: Intelligence, Short-term and Working memory Model

Research on the aforementioned WM model has introduced some debate around conceptually differentiating between WM and short-term memory (STM). Many studies frequently refer to the two constructs synonymously (Anderson, 1990; also described by Engle et al., 1999) while others consider WM as a subset of the larger construct of STM (Cowan, 2008). Engle et al. (1999) proposed and provided empirical evidence for conceptualizing STM and WM both as distinct but related constructs. These authors define STM as a temporary store or rehearsal mechanism, similar to Baddeley and Hitch's definition of the phonological loop and visuo-spatial sketchpad slave systems in their tripartite model (Baddeley & Hitch, 1974). Tasks that utilize these slave systems are only related to WM in as much attentional control is required to perform them well. Therefore, there are no 'pure' WM or STM tasks in this conceptualization; all memory related tasks inevitably fall under both domains, but central executive functions will be required more or less depending on the task demands (see Figure 4). This research therefore supports the idea that task categorization (as either WM or STM) will differ on an individual subject level depending on their developmental state, their ability to solve novel problems and their vocabulary and general knowledge capacity. Taken together this set of characteristics is commonly referred to as intelligence (g) (see Horn & Cattell, 1967). As a developmental example, repeating three letter words, while simple for most 20 year olds, is a complex and involved task for three year olds, and thus involves more attentional control or central executive involvement for the younger children due to their comparatively lower intelligence (or g). This concept is particularly important when relating these constructs to clinical populations for whom, due to deficits inherent to their condition, specific tasks

might involve greater demands on attentional control as compared to their typically developing same aged peers.

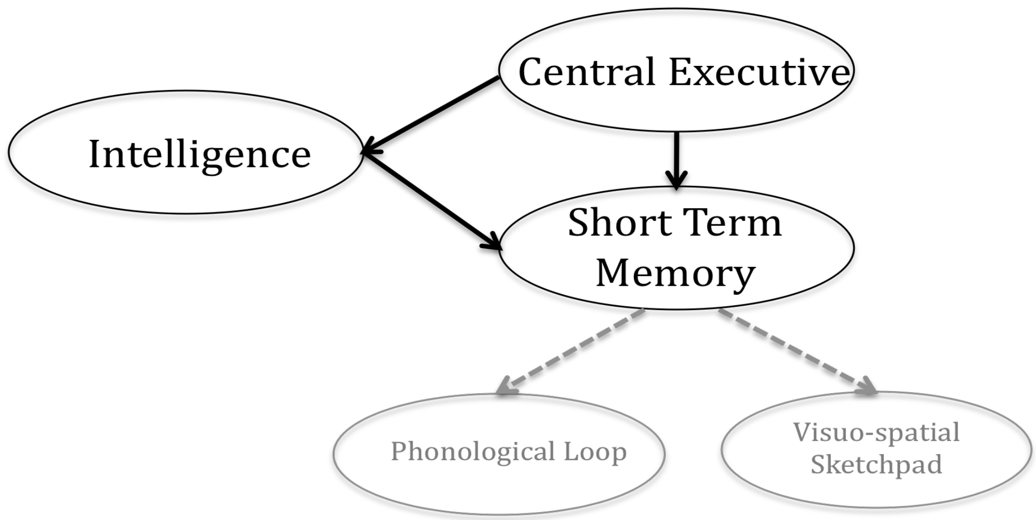


Figure 4:Theoretical relationships between components of the WM system: A representation of the WM model as drawn by Engle et al. (1999). The magnitude of the path between the Central Executive and Intelligence constructs differs depending on individual intellectual capacity as described above. The Short Term Memory (STM) component can be theorized to capture both the phonological loop and visuo-spatial sketchpad slave systems as well as the episodic buffer described by the Baddeley and Hitch (1974) theory (represented by the grayed in addition to the model above). The Baddeley and Hitch (1974) model does not account for intelligence and would likely theorize this construct to be external but related to their overall theory.

1.4 Model 3: Working Memory and Language Acquisition Model

In the last decade, the importance of the phonological loop and its role in language acquisition has been fiercely debated by both memory and language researchers (i.e., Acheson & MacDonald, 2009; Baddeley, 2003; Gathercole, 2006; Gupta & Tisdale, 2009). Early findings suggested that letters or words that are phonologically similar are more difficult to recall, as are longer and more complex words or sentences (Baddeley, 1966a; Baddeley, 1966b). Such results were consistent with findings from a patient with a pure phonological immediate memory deficit who had normal language production and

comprehension but faltered when material increased in complexity (Shallice & Vallar, 1990; Vallar & Baddeley, 1984b).

Memory researchers interpreted these findings as clearly supporting the causal role of the phonological loop in language acquisition, explaining that language acquisition is dependent on the capacity of the phonological loop, such that greater phonological capacity should be causally correlated with a richer vocabulary. Indeed, Gathercole and Baddeley (1989) found, in a cross-lagged correlational study, that non-word repetition (a measure for the phonological loop slave-system) measured at age 4 predicted vocabulary scores at age 5, but not the other way around. This relationship is developmentally driven such that as we age, vocabulary ability can in turn facilitate phonological abilities (like non-word repetition), thus the original tripartite model of WM becomes reliant on the crystallized system which it draws on to increase efficiency and capacity (Baddeley, 2003).

Researchers from the language-primary tradition have challenged this WM-driven basis of language acquisition for its ability to effectively account for phonological errors in both skilled and struggling readers. This language as primary focused school of thought posits that the similarities between errors in immediate recall of a list of words (a classic WM task) and common errors observed in language production, such as misordering of sublexical units, point to a common phonological encoding process which supersedes, or underlies, the WM system. This model, first proposed by Brown and Hulme (1996; see Figure 5), would suggest that existing language development and habits (which might be mis-learned or error ridden) facilitate multimodal learning (Acheson & Macdonald, 2009; Gupta & Tisdale, 2009). Indeed, Gathercole (1995)'s findings that nonwords which more closely approximate real English words were consistently easier to repeat than less familiar

phonemic sequences, suggest that existing language capabilities influence performance on a non-word repetition task (Baddeley, 2003; Gathercole, 1995). Acheson and MacDonald (2009) suggest that language structure underlying verbal WM tasks allows for ease of performance on tasks that are driven by syntax rules. For example, it is easier to recall a ten-word sentence than ten individual words because of the language structure that guides the former. Thus, according to these primary language theorists, there is an underlying feature of language ability driving performance on supposed verbal WM tasks, which are really just a proxy for acquired language abilities. In other words, the phonological loop is language production itself (see also MacDonald, 2017; Montag, 2016). Thus, in relation to the original Baddeley and Hitch (1974) model, the phonological loop would be supplanted by language acquisition depicted by Figure 5, while central executive and visual spatial sketchpad capabilities would be outgrowths of language as depicted in Figure 5.

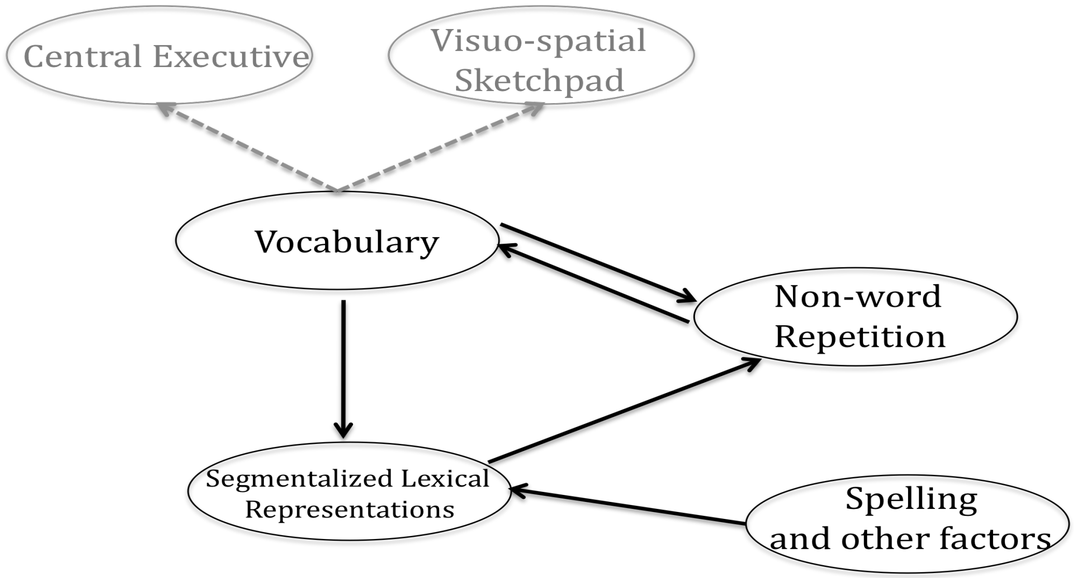


Figure 5: Brown and Hulme (1996) hypothesis of vocabulary growth described as the assumed relationship between vocabulary[1], nonword repetition and other related factors. Baddeley (2003) describes this model as eliminating the role of phonological WM in explaining language growth. This model might presume that visual-spatial sketchpad (described by Baddeley and Hitch, 1974) and central executive capabilities (described by both Baddeley and Hitch, 1974 and Engle et al., 1999) are driven by language acquisition (depicted here by dotted lines and grayed factors).

1.5 Working Memory and Reading

WM skills are highly related to successfully learning to read (Dawes et al., 2015; Nicholson, Fawcett and Baddeley, 1992). In developing reading, children must efficiently integrate phonemic units with their correct visually symbolic representations, connect with the semantic system, as well as utilize successful motoric articulation (as described by Dehaene, 2009). The foundation for the development of this highly complex skill is rooted in phonological awareness or phonemic decoding. Phonological awareness includes the ability to recognize rhyming, blend syllables, segment words, manipulate

[1] It is relevant to note that Brown and Hulme (1996) theorized 'vocabulary' as 'vocabulary growth', connoting that changes in vocabulary acquisition will impact and be impacted by non-word repetition and that growth will causally impact segmentalized lexical representations. Since growth can only be measured across time, thus demanding multi-year longitudinal data collection, for the purposes of this study we theorized this construct as only 'vocabulary' collected at a discrete time point.

word sounds and repeat unfamiliar words (Snowling, 2000; Torgesen et al., 1990). The ability to repeat unfamiliar words, measured by a test like the Non-Word Repetition (NWR) task, has been described as tapping into both phonological and related WM or STM (depending on the theoretical definition) memory abilities in DD children, as individuals are asked to listen to a pseudo-word, hold it in memory and reproduce it (Coady & Evans, 2008; Hoff, Core & Bridges, 2008; Wagner, Torgesen, Rashotte, & Pearson, 2013). Indeed, the NWR task correlates highly with the Digits Span Forward task, a verbal short-term memory task involving repeating strings of numbers immediately after hearing them (Torgesen, Wagner, Simmons, & Langhon, 1990). Due to the link between phonological skills and the phonological loop in the classic Baddeley and Hitch (1974) WM model, memory researchers have hypothesized that the core deficit of phonological awareness in children with DD might, in part, be driven by weakness in the phonological loop slave system in the WM model (White et al., 2006).

Despite the identified relation between phonological working memory in children with DD, no group has yet evaluated the similarities and differences among the Baddeley and Hitch (1974), Engle at al. (1999), or Brown and Hulme (1996) models in a formally diagnosed DD population. Based on the Baddeley and Hitch model of WM, it is easy to surmise that a faulty phonological slave system might at least partially underlie this developmental reading deficit, and restrict individual response to intervention. Researchers, like Engle at al (1999), who support a more generalist perspective, might view such phonological deficits as stemming from underlying executive difficulties coupled with overall cognitive capacity or intellectual acumen. The language theorists would disagree and posit that early language development is equivalent to phonological

development, and thus once vocabulary is accounted for, phonological abilities would not hold any significant predictive capacity of appropriate reading development.

Interestingly, recent research has shown that DD individuals also struggle on tests measuring abilities that tap into other components of the Baddeley and Hitch (1974) WM model. Some studies point to decreased performance even on visuospatial WM tasks (Menghini, Finzi, Carlesimo, & Vicari, 2011; Smith-Spark & Fisk, 2007), providing evidence of a possible deficit in the 'visual-spatial sketchpad' slave system in those with DD. Others have shown more general 'central executive' difficulties (Smith-Spark, Fisk, Fawcett, & Nicolson, 2003) in DD individuals. These studies have led to recent conclusions that the phonological loop deficit hypothesis (Snowling, 2000; White et al., 2006), as typically measured by the NWR task, may limit our understanding of the range of WM deficits associated with DD (Menghini et al., 2011). This conclusion highlights the need to better understand WM tasks across non-phonological modalities, to contribute to a more complete understanding of the deficits identified in DD.

1.6 Working memory's influence on reading intervention outcomes

Research has shown that reading interventions targeting phonological deficits by teaching foundational single word decoding skills produce strong long-term outcomes for elementary school children with DD (Torgesen, 2005). Younger children benefit most from phonology-based intervention strategies and the effects of these strategies can last for years following the intervention (Lovett et al., 2017; Torgesen, Wagner, Rashotte, Burgess, & Hecht, 1997). Despite these promising intervention outcomes, there remains a significant group of children whose reading scores hardly improve (poor-responders) despite undergoing explicit and intensive intervention.

Researchers have attempted to characterize this poor-responder or treatment resistant group by examining possible cognitive markers that might distinguish them from their responding peers. Unsurprisingly, a majority of intervention-focused studies have found that baseline phonological skill was predictive of treatment outcomes (as examples see: Al Otaiba & Fuchs, 2002; Frijters et al., 2011; Stuebing et al., 2015). These studies have also examined alternative predictors including attention, oral language, spelling and working memory, and all accounted for unique variance in understanding changes in reading over the course of an intervention, but no single predictor was better (Stuebing et al., 2015). Stuebing et al. (2015) and Frijters et al. (2011) also emphasized that it is difficult to even parse out the differences between the 'all star' predictor, phonological awareness, and other cognitive predictors due to the variance they share. Stuebing et al. (2015) in fact ascertains that "It would be difficult to support conclusions that phonological awareness is more related to outcomes than WM, especially because many phonological awareness assessments include a WM component." (p. 21). Thus, if phonological awareness (PA) and WM are so closely linked, it is possible that findings supporting the strong predictive abilities of PA might actually be masking the global influence of WM.

Attempting to parse out the differential contribution of WM from PA has proven difficult for both review and meta-analytic studies largely due to two factors. First, in an attempt to categorize tasks, all tasks related in any way to phonological awareness have been grouped together into an overarching phonological awareness category without true consideration for task components external to, or independent from, the 'phonological component'. As an example, Stuebing et al. (2015) group CTOPP tasks (which include

both phonological working memory tasks like non-word repetition, phonemic isolation tasks and rapid naming tasks) with sound discrimination tasks (Hatcher & Hulme, 1999; which contain no WM component) and articulatory awareness tasks (e.g. in Wise et al., 1997) under one umbrella she termed 'Phonological Awareness'. Since individual studies rely on disparate operationalization of the term 'phonological awareness', understanding such aggregate results across studies forces reviewers to broaden their categorization definitions, thus losing important accountability for components of task variance along the way. Second, no study to date has examined the predictive value of phonological awareness as a component of WM. In fact, studies have focused primarily on the cognitive predictors of reading by utilizing tasks directly related to reading and language while disregarding alternative components of contribution. To use Stuebing et al. (2015) as an example again: Tasks categorized as WM in their study are largely language based and include the aforementioned CTOPP (which was also categorized as a task of phonological awareness) and the Sentence Span Task (O-Shaughnessy & Swanson, 2000). These authors, as well as Frijters et al. (2011), also included the WISC-III working memory index, which, by many accounts, are tasks of short-term memory, and do not consider the central executive or divided attention component of some of these tasks which both Baddeley and Hitch (1974), and Engle et al. (1999), view as the instrumental component of the WM definition or system. The current study therefore attempts to understand the categorization of some of these tasks within the conceptual framework of WM in addition to exploring the differential contribution of these cognitive variables in predicting reading changes over the course of an intervention. Understanding the relations between these system components can allow us to better articulate and

categorize areas of difficulty for those with DD and in turn allow for greater intentionality and precision in intervention development and implementation.

1.7 Project Aims and General Procedures

This study examines the degree to which cognitive indicators collected from a group of children with DD best fit three different hypothesized models of WM: Baddeley and Hitch (1974) WM model; Engle et al. (1999) component model; and Brown and Hulme (1996) language model. In order to directly compare the models, ten tasks which were thought to tap into the theoretical latent factors of each model were carefully chosen. The tasks were: (1) Test of non-word repetition (TNWR), (2) Digit Span Forward, (3) Digit Span Backward, (4) Spatial Span Forward, (5) Spatial Span Backward, (6) Counting Span, (7) Comprehensive Test of Phonological Processing (CTOPP-2) Phonological Awareness composite, (8) Peabody Picture Vocabulary Test (PPVT), (9) Clinical Evaluation of Language Fundamentals (CELF) Sentence Repetition Task, (10) Wechsler Abbreviated Scale of Intelligence (WASI) Matrix Reasoning task. Confirmatory factor models (CFA) were used to evaluate the relative contribution of each indicator to the critical latent factors.

Indicators for all models were tasks administered to the children with DD prior to their participation in a 70 hour explicit reading intervention program. Additionally, each model contained two additional free indicators; 'Initial Reading Score' and 'Reading Intervention Gains', both of which are z-scores computed by utilizing composites of initial reading scores across several reading measures and reliable change scores (RCS). Description of the calculation and brief theoretical basis of change scores are discussed in the Methods.

1.7.1 Model 1: The Baddeley and Hitch theory

As previously described, the Baddeley and Hitch model of WM is a three-factor model with two slave systems (Figure 8). Given our chosen indicators, the Baddeley and Hitch model would predict that NWR and Digit Span forward and backward would load onto the Phonological Loop latent factor, while Spatial Span forward and backward would load onto the Visual Spatial latent factor. A core component of this model, the Central Executive, is indicated by the Counting Span task, and is hypothesized to be strongly related to both previously described latent factors. Under this model, nonverbal IQ, Vocabulary and language variables would be external to the model but possibly loosely related as evidenced by previously described research indicating the relation between phonological WM, vocabulary and intellectual abilities (Baddeley, 2001; Gathercole, 2012).

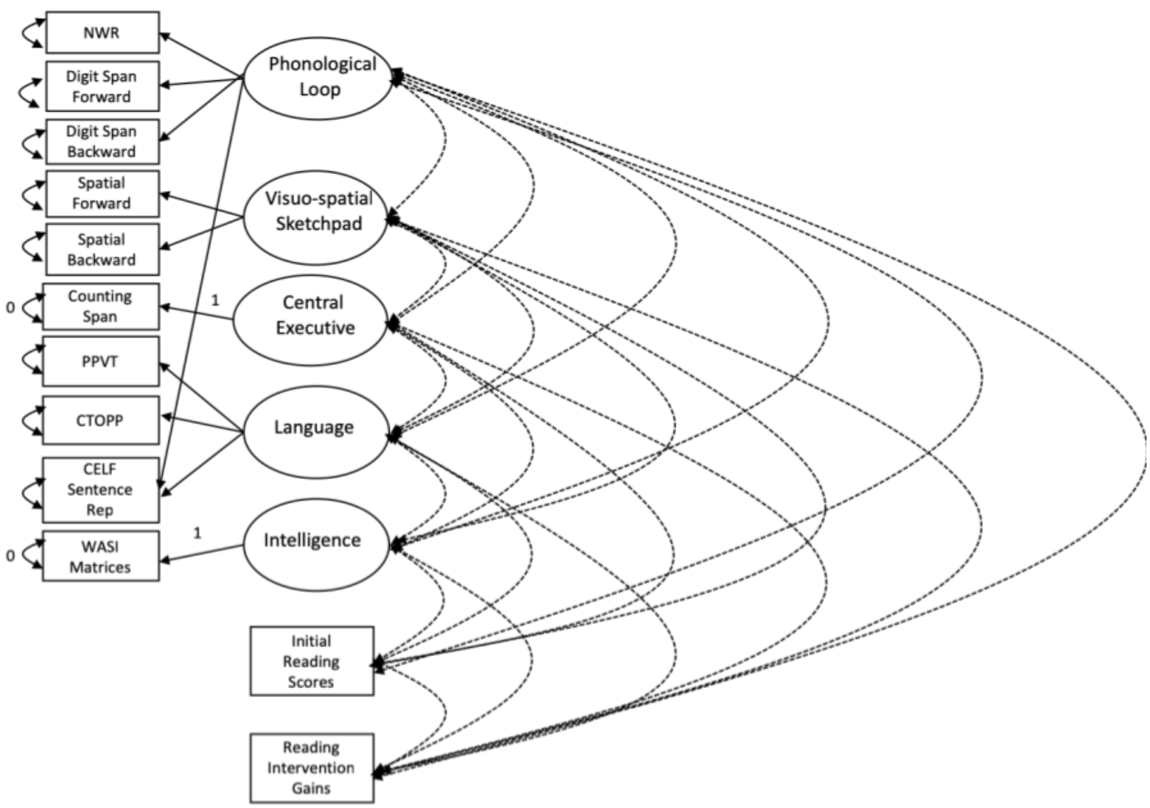


Figure 5: Model 1; Driven by the Baddeley and Hitch (1974) WM theory - The Counting Span Task which is the sole indicator of the Central Executive, is measured as phantom factor, related to the phonological and visuo-spatial latent factors. Intelligence

(WASI) and Language (PPVT, CTOPP and CELF) are theorized as external factors only loosely related to the other latent factors. Tasks and their theoretical construct classifications are described in the Methods section of this paper.

1.7.2 Model 2: The Engle et al. (1999) theory:

The Engle et al. (1999) component model of WM proposes that all tasks falling under the tripartite two component slave systems are re-categorized as a single 'Short-Term Memory' (STM) latent factor (Figure 9). This model would also propose that both the PPVT (Vocabulary) and WASI Matrix Reasoning indicators would fall under a second latent factor 'Intelligence' (Engle et al., 1999). The final indicator, counting span, would be a proxy for Engle's 'Central Executive' and is theorized as a phantom factor in the model related to both latent factors.

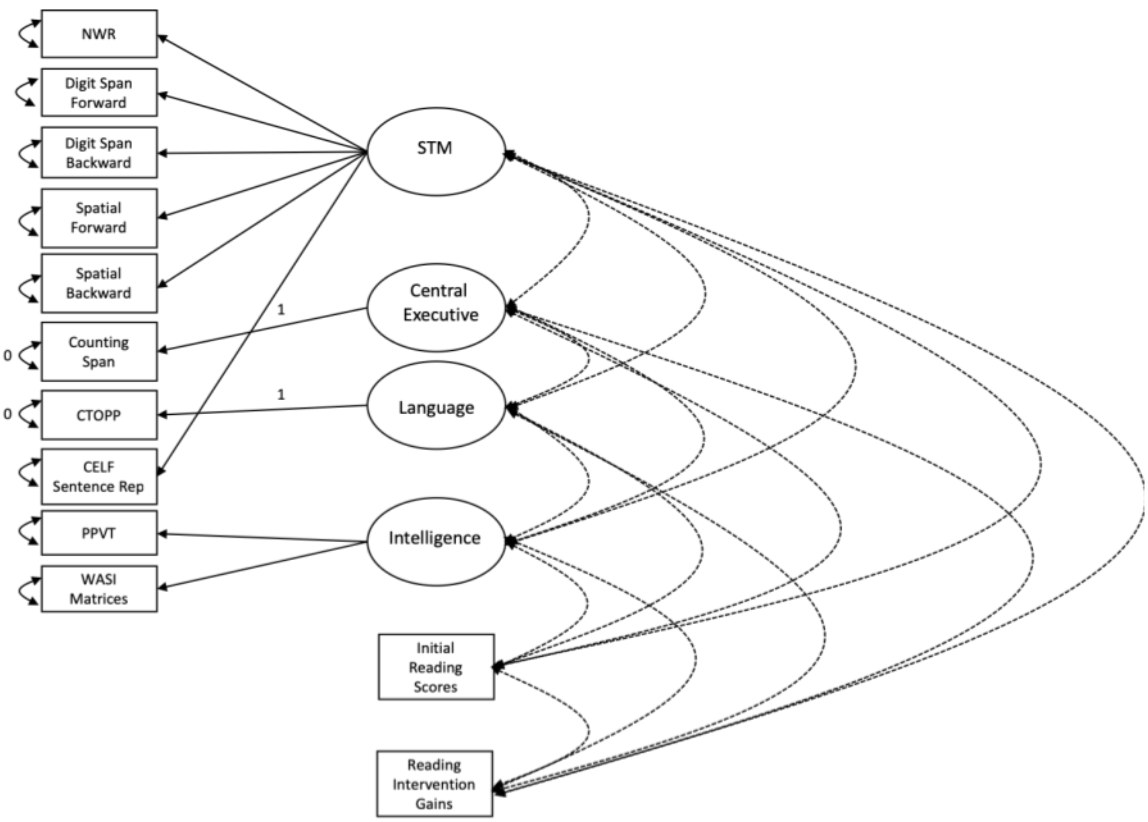


Figure 6: Model 2; The Engle et al. (1999) model:
The Counting Span task is the sole indicator of the latent construct Central Executive and is therefore theorized as a phantom factor related to the other two latent factors. The CTOPP task (falling under the phantom factor Language) is theorized as external but loosely related to the overall model. Tests and their associated theoretical construct classification are described in the Methods section.

1.7.3 Model 3: Hulme and Brown (1996) Language Theory:

Hulme and Brown's theory, expanding on the above, posits that language knowledge and phonological abilities are closely linked, whereas memory abilities are separate and perhaps predicted by underlying language abilities (Figure 10). Therefore, for this language driven model all language related indicators are hypothesized to be subsumed by the 'Language Abilities' latent factor, whereas other memory capabilities (regardless of modality) are collapsed under a 'Memory' factor theorized to be only loosely related to but separate from Language. As the Digit Span Tasks can be theorized as either WM and phonological, the theoretical decision to place them under Memory is

discussed in the Methods section below. The WASI Matrix Reasoning indicator is placed under the phantom factor 'Intelligence' and would be hypothesized to be external but loosely related to the two latent factors.

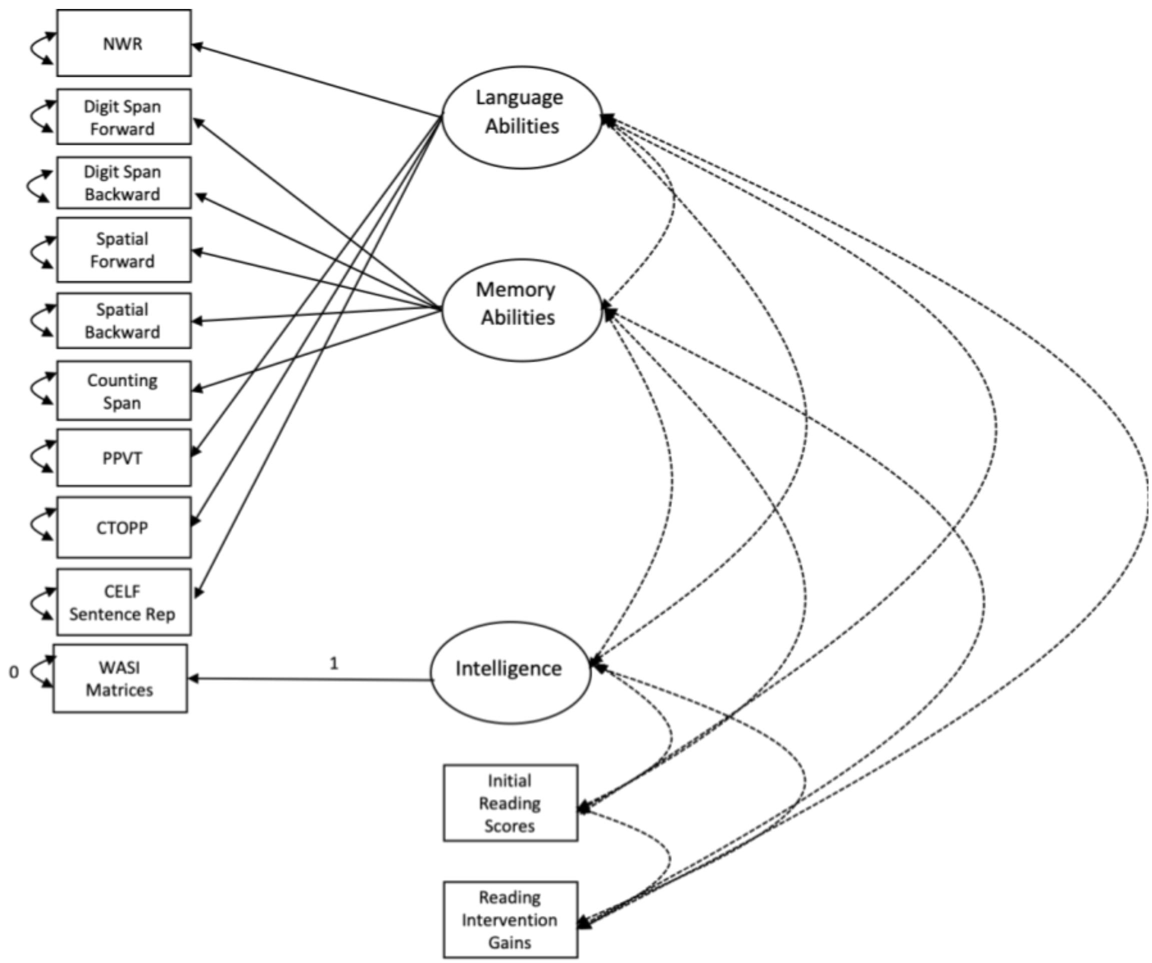


Figure 7: Model 3; Brown and Hulme (1996) Language Model:

Four language related tasks (NWR, PPVT, CTOPP and CELF Sentence Rep) are subsumed by the Language Abilities latent factor, whereas the remaining memory related indicators are subsumed by the Memory factor. WASI Matrices is placed under the latent factor Intelligence and is theorized to be related to the other two latent factors. Test and their associated theoretical construct classification are described in the Methods section.

1.8 General Hypotheses:

Based on the above literature review and conceptual proposals, three general hypotheses are made:

1. We hypothesized that the Baddeley and Hitch theory-based model of WM would be the best-fit model for the current data. Additionally, even though this model does not boast the parsimony offered by the Engle et al. (1999) or Brown and Hulme (1996) models, given its large number of free indicators (or statistically restricted phantom factors) compared to the other proposed models, it would likely have the best statistical fit (Bollen & Bauldry, 2011).
2. We predicted that the Phonological Loop WM factor from the Baddeley and Hitch model would be the strongest predictor of reading reliable change score indices over the course of an intensive reading intervention.
3. Based on predictions of #2, we also expected that the Phonological WM factor from the Baddelely and Hitch model would share the most variance with intervention gains.

METHODS

1.9 **Participants and Inclusion/Exclusion Criteria:** *Descriptive statistics of criteria are listed in Table 1.*

- One hundred and eight (N=108) elementary school children (ages 8-11) were recruited from seven public school sites in the local public and charter school systems. Participating schools gave information packets to parents whose children have been identified by the school as struggling readers. Parents who were interested returned completed consents and questionnaires to experimenters and an initial reading screening was completed. Those who met inclusion criteria were offered interventions.
- All children were classified as struggling readers/DD based on scores of 1SD below their age norm expectations (SS < 85) on the Basic or Broad Cluster Scores, or their subtests, of the Woodcock Johnson Fourth Edition (WJ-IV) Tests of Achievement (Schrank, McGrew, Mather, & Wendling, 2014), or Test of Word Reading Efficiency Second Edition (TOWRE-2) Composite Score (Torgesen, Rashotte, & Wagner, 1999). All participants displayed a minimal IQ score of 80 on the Full Scale Wechsler Abbreviated Intelligence Scale (WASI; Wechsler, 1999). All students were present for at least 65 of the 70 hours of explicit intervention (described in detail below) over the course of the school year.
- Were in Grades 3 or 4 during the school year of the study.
- Children in mainstream education or special education could participate based on school or teacher referrals.
- First and primary language must be English.

- Must have corrected vision (>20/40) and adequate hearing (>25 dB at 500+ Hz bilaterally).
- Had no history of being diagnosed with serious emotional/psychiatric disturbances (major depression, psychotic or pervasive developmental disorder).
- Had no history of a chronic neurological or medical condition (like seizure disorder or acquired brain injuries).
- Parents signed parent permission forms for screening and intervention components of the study, and provided answers to questions on child development, educational and medical history, language status, and family demographics using a brief history questionnaire.

Table 1: Descriptive statistics of participants' standard scores on inclusion and diagnostic measures

Characteristics	***DD Participants M (SD)***
WJ-III LW Standard Score	88.30 (8.09)*
WJ-III WA St. Score	88.17 (8.57)
WJ-III RF St. Score	86.73 (10.00)
WJ-III PC St. Score	80.69 (8.28)
WJ-III Basic RC St. Score	87.17 (7.56)
WJ-III Broad RC St. Score	82.37 (9.05)
TOWRE-2 SWE St. Score	76.15 (10.31)
TOWRE-2 PDE St. Score	73.60 (8.20)
WASI-2 FSIQ St. Score	93.83 (10.20)

- Normative average = Standard Score (SS) =100, standard deviation (SD) =15.

1.10 Model Indicators

1.10.1 The Counting Span Task (CT-SPAN)

The CT-SPAN task (Turner & Engle, 1989) is a computer task which requires participants to count specific geometric shapes (e.g., orange squares) while remembering the sequence of counts for each set of items. For example (see Figure 11), participants look at a screen full of shapes and are asked to recall the number of blue squares. On the

next screen they see a different series of shapes and again recall the blue squares. Following a number of such screen presentations, the participant is asked to produce the number of orange squares they saw on each screen in the correct order – this is a sequence. The number of counts to be remembered increases in difficulty. Participants reach a ceiling after making an error on one sequence. Given the stringent ceiling criteria, partial scores (i.e.. the number of correct individual trials) are commonly used as a measure of accuracy and were thus used in the present analysis (see Cowan et al., 2005 for an extensive review). In order to ensure than participants were using verbal in addition to visual and spatial cues, participants were primed to 'count the shapes quietly to themselves' while engaging in the task.

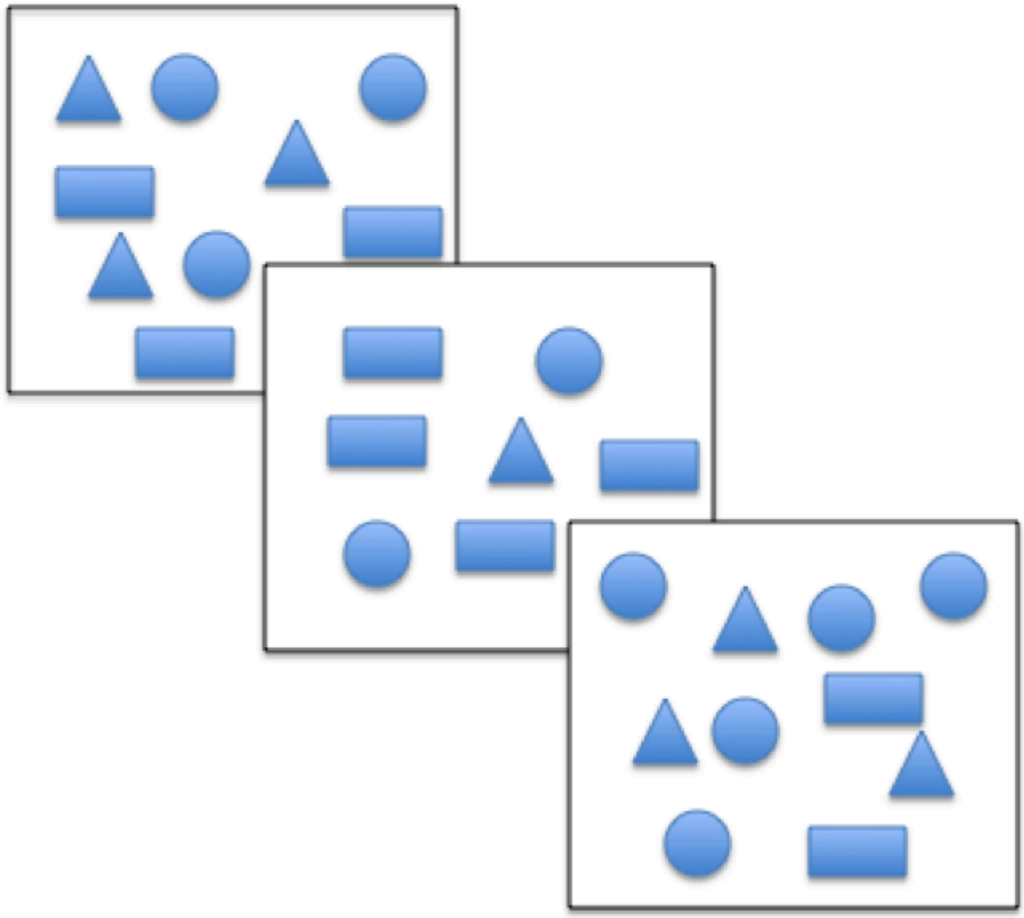


Figure 8: Example of the CT-SPAN task
For each screen, participants were asked to count the number of blue circles while holding the number of blue circles counted on previous screens in memory. Following an increasing number of screens participants were asked to list the number of blue circles counted on each screen in the correct sequence. The correct response in this example would be 3-2-5.

The CT-SPAN task was presented using E-Prime experimental software._The CT-SPAN task has classically been described as an overall measure of central executive capacity in the Baddeley and Hitch (1974) WM system. The task relies on a participant's ability to code and integrate visual, spatial, and verbal information as well as increasingly high demands of storage capabilities as each additional screen (set of shapes) interferes with the previous counting completed on previous screens (Cowan et al., 2005; Dahany,

Windsor, & Kohnert, 2007). Due to the necessity of a high level of attentional control to perform this task, Counting Span is often seen as a proxy for higher order executive functioning abilities (Engle et al., 1999).

This task has been widely used and its internal consistency is 0.77 (Conway et al., 2005; Kane, et al., 2004). The CT-SPAN task also has high convergent validity with other tasks during which attention and manipulation are important (like tasks of language comprehension, following spatial and oral directions, and in context vocabulary learning, Conway et al., 2005). The CT-SPAN task has discriminant validity with tasks that measure automatic processes and do not require manipulation of material (Kane, Bleckley, Conway & Engle, 2001).

For both the WM memory (Model 1) and Engle et al., (1999) models (Model 2) the CT-SPAN task falls under the Central Executive. Since Brown and Hulme's language model does not discriminate among non-language based WM tasks and their general relation to language development, the CT-SPAN task is placed with other 'language free' WM tasks under the latent factor 'Memory Abilities' in this model (Model 3).

1.10.2 Test of Non-word Repetition (TNWR)

The TNWR (Dollaghan & Campbell, 1998) is an experimental measure with 16 nonwords (ranging from 1-4 syllables in length) for 96 total syllables and possible points (1 point per correctly repeated syllable, which was the tabulated raw score used for the present study). The experimenter speaks a word and audio records the child's repetition response. All stimuli begin and end with consonants (no consonant clusters or solely tense vowels). Split-half reliability is high, reported as .85 (Dollaghan & Campbell, 1998).

The NWR test has been widely used in alternative forms, most notably in the CTOPP (Wagner, Torgesen, & Rashotte, 1999) and CTOPP-2 (Wagner, Torgesen, Rashotte, & Pearson, 2013), and is highly correlated to reading fluency, vocabulary and working memory tasks (Coady & Evans, 2008). Due to its relatability across a spectrum of factors, it has been hypothesized to be subsumed by different latent factors in each model: (i) For the WM model (Model 1), NWR is hypothesized as an indicator of the Phonological Loop; (ii) For the Engle component model (Model 2) it is theorized as a simple STM task, as, like Digit Span (described below), it does not rely on higher order attention facility for successful task completion; (iii) For the language theory Brown and Hulme (1996) model (Model 3), NWR is subsumed by the Language Abilities latent factor due to this body of research supporting its close relation to vocabulary acquisition and maintenance (MacDonald, 2016).

1.10.3 The Digit Span Forward Task

The Digit Span Task is an individually administered subtest from the Wechsler Intelligence Scale for Children-IV (WISC-IV) used to measure phonological working memory (Integrated Wechsler, 2004). Participants are asked to repeat a series of numbers in forward order. The length of sequences increases along with difficulty across trials. Ceiling is reached following two consecutive incorrect trials of the same length. One point is awarded for each correct series with a maximum of 14 total points. The Digit Span Forward Task has been shown to correlate highly with other measures of phonological WM (i.e. NWR; see Heitmann, Asbjornsen, & Helland, 2004; and Torgesen, Wagner, Simmons, & Laughon, 1990).

1.10.4 The Digit Span Backward Task

The Digit Span Backward Task was used to test verbal WM requiring increased variable manipulation. For this task, participants are asked to repeat a string of numbers in backward order with a ceiling reached following two consecutive incorrect trials of the same length. The strings of numbers increase in length and thus difficulty across trials. Like the Forward Task, one point is awarded for each correct series with a maximum of 14 total points. In addition to its strong correlation to phonological WM (Heitmann et al., 2004), recent research has further classified the Digit Span Backwards task as a manipulative task of phonological WM requiring participants to use the phonological loop slave system to hold the Digits Span numbers in memory and then reproduce them in the opposite order (Vasic, Lohr, Steinbrink, Martin, & Wolf, 2008). This task has been shown to be specifically challenging for older children with DD (Vasic et al., 2008).

Due to their reliance on verbal rehearsal and output, the Digit Span tasks have been classified as part of the Phonological Loop latent factor in the Baddeley and Hitch model (Model 1). For the Engle et al. (1999) component model (Model 2), these tasks were conceptualized as STM tasks, as, despite the manipulative component of the Backward Span Task, correct responses do not rely on longer-term storage or interference of competing stimuli. For language primary theorists (whose theoretical thinking lies closely with the Brown and Hulme, 1996 model, Model 3), the Digit Span Tasks are challenging to classify due to their reliance both on rote memory and verbal ability. In a comprehensive overview of WM tasks, Cowan (2008) concluded that the Digit Span tasks are conceptually closer to STM or WM tasks than true tests of verbal abilities and

are thus classified as part of the Memory Abilities latent factor in the language primary driven model (Model 3)[2].

1.10.5 The Spatial Span Forward Task

The Spatial Span Forward task (WISC-IV Integrated; Wechsler, 2004) assesses visual-spatial working memory and has been shown to be a strong measure of the visuospatial sketchpad slave system (Bacon, Parmentier, & Barr, 2013; de Jong et al., 2009; Sumner, Griffith, & Mineka, 2010). In this task, participants reproduce sequences of tapped blocks presented by the tester. The task is administered on a three-dimensional model with increasing length of blocks tapped and complexity across trials. Ceiling is reached following two consecutive incorrect trials of the same length. One point is awarded for each correct series with a maximum of 14 points total.

1.10.6 The Spatial Span Backward Task

Similar to the Digit Span tasks, in order to examine a measure of spatial WM with increased attentional control, we compared performance across both the Spatial Span Forward and Backward tasks (the Spatial Span Backward task from the WISC-IV asks participants to reproduce sequences of blocks in the opposite or backward sequence presented by the tester). Like the Digit Span tasks, one point is awarded for each correct series with a maximum of 14 total points.

Spatial Span Backwards has been shown to be particularly challenging for children and adults with DD with hypotheses suggesting either the role of either a faulty or weak visuospatial sketchpad (c.f. Vandierendonck et al., 2004) or perhaps overall executive functioning difficulties taxing the visuospatial system (Bacon et al., 2013).

---

[2] Despite this theoretical reasoning, we also ran the model with the Digit Span tasks as part of the Language Factor, see Appendix A.

As model indicators, both Spatial Span tasks have been theorized to be related to the Visuo-spatial sketchapd latent factor for the WM theory (Model 1). For the component model (Engle et al., 1999, Model 2), they are hypothesized to fall under the STM latent factor as, like Digit Span and NWR, they do not require higher order attentional control or longer-term manipulation like the Counting Span Task. The Spatial Span tasks are not related to any component of the language theory and therefore for Model 3 they are hypothesized to be subsumed by the Memory Abilities factor of the model.

1.10.7 WASI-II Matrix Reasoning (Wechsler Abbreviated Scale of Intelligence Second Edition, Matrix Reasoning)

The WASI-II Matrices is a non-verbal measure commonly used as a proxy IQ task which, particularly when combined with a verbal IQ task. Matrix Reasoning examines one's ability to choose the correct picture to complete a pattern. The WASI-II is a robust and frequently used measure for assessing overall cognitive ability (i.e., IQ).

The WASI-II has been normed on national sample of approximately 2,300 individuals ranging in age from 6:0 to 90:11. Designed to correlate with the Wechsler Intelligence Scale of Children-IV (WISC-IV; Wechsler, 2003) as well as the Wechsler Adult Intelligence Scale-IV (WAIS-IV; Wechsler, 2014), the WASI-II has been shown to have high concurrent validity. Among the child sample (ages 6-16 years), reliability coefficients of scores on the WASI-II range from 0.87 to 0.96. For the present study, raw scores were used.

For the Baddeley and Hitch (1974) WM driven model (Model 1), intellectual abilities are not included in the model and are thus hypothesized to be related but external to

the underlying two-factor model. For the component model (Model 2) indication, the WASI-II Matrix Reasoning is used as a proxy for Intelligence as outlined by the Engle et al. (1999) model. For the language model (Model 3), this intelligence measure is theorized as an external but loosely related indicator to the memory and language factors as it does not include an explicit language or memory component.

1.10.8 The Peabody Picture Vocabulary Test (PPVT-4)

The *Peabody Picture Vocabulary Test (PPVT-4)* was used to measure lexical/vocabulary abilities (Dunn & Dunn, 2007). During test administration, participants were shown four pictures and asked to identify the picture that best matches a word provided to them by the tester. All participants started with practice items to ensure clear understanding of the instructions. Start items vary depending on participant age, and a basal is reached when the participant correctly identifies at least 11 out of a possible 12 items in a set. Ceiling is reached when a participant is incorrect on at least 8 items in the 12-item set. PPVT-4 raw scores are converted to standard scores with an average of 100 and a standard deviation of 15. For these analyses, raw PPVT scores will be used to maximize comparison between and within participants.

The PPVT-4 is a well-established measure of vocabulary skills. The test was normed on a sample of 3540 individuals aged 2 years 6 months to 90+ years to establish age norms, and a subsample of 2003 individuals used for grade norms (kindergarten through grade 12). The distributions and numerical targets for each age groups were matched to 2004 U.S. Census data. The PPVT-4 has been shown to have very high split half reliability (.94 and .95, forms A and B), alternate form reliability (.87 and .93), and test-retest reliability (.92 to .96). The test has high correlations with other tests of vocabulary knowledge (.80 to .84) and

substantially with measures of other aspects of language and reading skill (.37 to .79). Additionally, the PPVT-4 successfully discriminates between children with language delays and disorders and those who are developing typically, however children with disabilities related to speech and hearing do not differ significantly from the general population on the PPVT-4.

For the WM model (Model 1) indication, the PPVT-4 was hypothesized as an external variable loosely related to the model. Due to the high convergent validity between IQ and vocabulary ability (Engle et al., 1999), the PPVT-4 was hypothesized to fall under the Intelligence factor for the component model (Model 2). For the language model (Model 3), the PPVT-4 was hypothesized as falling under the Vocabulary Abilities latent factor.

1.10.9 Comprehensive Tests of Phonological Processing (CTOPP-2)

The *Comprehensive Tests of Phonological Processing (CTOPP-2*; Wagner et al. 1999), was used as a measure of phonological awareness. Three CTOPP-2 subtests, *Elision, Blending*, and *Phoneme Isolation* form a Phonological Awareness Composite score which measures a child's awareness of, and access to, the phonological structure of oral language. In the present study this composite score was used as a measurement of this indicator.

The Elision subtest is a 34-item test which measures the ability to remove individual phonemes from words to form other words (e.g., "say band without saying "/b/", correct response: "and"). The test becomes progressively more difficult until the deleted phonemes can no longer be detected through orthographic knowledge of a word (e.g., say "fixed" without saying "k"). The Blending Words subtest is a 33-item audio-recorded test measuring a child's ability to blend sounds to form words (e.g., "What word do these sounds make: t-oi?"- correct response: "toy"). Finally, the Phoneme Isolation subtest is a 32-item

test which measures the ability to isolate sounds within words (for example: "What is the second sound in the word *island*?"). Items become more difficult as spelling strategy cannot be used.

To move forward on each subtest, each examinee must answer one of the first three items correctly after receiving feedback from the examiner. A ceiling is reached when the examinee has responded incorrectly three times in a row or when the last item has been administered.

The CTOPP-2 has been normed on a sample of 1900 individuals ages 4-24 years, and has shown high test re-test reliability (between .75 to .92), scorer difference reliability (all coefficients exceeded .90), as well as high criterion prediction validity (between .64 - .82 across 9 other phonologically based tests).

For the WM model (Model 1) the CTOPP-2 was hypothesized as an indicator of the Language Abilities latent factor, which, due to its heavy reliance on phonological skill, shares variance with the NWR indicator and the phonological latent factor. For the Engle Model (Model 2), the CTOPP-2 is theorized as an external variable under a phantom latent factor, Phonological Abilities. For the Language Model (Model 3), we would hypothesize that the CTOPP-2 would be an indicator for the Language latent factor.

1.10.10Clinical Evaluation of Language Fundamentals (CELF-4) Recalling Sentences

The *Clinical Evaluation of Language Fundamentals (CELF-4*; Semel, Wiig, & Secord, 2003) is a measure of general language ability, which, upon administration of a core set of subtests produces a Core Language Score (CLS). For the current study, the *Recalling Sentences* subtest was used to measure working memory of complex language form. The

Recalling Sentences subtest tests a student's ability to listen to and repeat spoken sentences increasing in length and complexity.

The standardization of the CELF-4 involved over 4500 children, adolescents and young adults across 12 age groups (age ranges 5-21 years) in 47 states. Average test-retest reliabilities range from a low of .70 (*Recalling Sentences)* to .90. Internal consistency measures ranged from coefficient alpha of .70 – .95, indicating homogeneity of the items within an individual subtest. The CLS has high correlations with many of language indexes and has shown large effect sizes between children diagnosed with language disorders and those who are developing normally. It is sensitive to identifying children who struggle with specific difficulties in language (Semel et al., 2003). For this study, the total raw score was used to measure this model indicator.

Due to its reliance on both core phonological memory abilities and basic language tenets, for the WM model (Model 1), the CELF-4 Recalling Sentences indicator was theorized to fall under both the Language and Phonological Factors. The Engle model (Model 2) would similarly categorize this indicator under both a Language and general WM factor[3]. The Language model (Model 3) would place Recalling Sentences as part of the Language Abilities latent factor and posit that any variability shared with the WM latent factor would be secondary and insignificant.

1.11 Intervention Design

Participants were placed into instructional groups of 4-8 students depending on their performance on the reading screening measures (see below). Intervention sessions lasted between 45 and 60 minutes a day for 70 contact hours, with most schools scheduling them

[3] Although the model was initially run with the CELF Sentence Repetition loading onto both the STM and Language Factors, this model did not converge. Therefore the final model (Figure 13) shows the CELF Sentence Repetition loading only onto the STM factor.

every day of the week. Experienced teachers from the research project implemented the intervention program (PHAST) for each group of children. The reading program provided a variety of reading strategies and have already shown positive effects for elementary and high school students (Lovett, Lacerenza, Palma & Frijters, 2012; Morris et al., 2012) in previous studies. Direct instruction of foundational skills was combined with explicit instruction of reading strategies (word identification and comprehension). The intervention was carefully scaffolded so that before strategies were introduced, the prerequisite skills needed to execute the strategy successfully have been taught and practiced. Strategies are introduced over time and children were taught a metacognitive plan that allowed them to select, apply, monitor, and evaluate strategy application during reading. The intervention components and the overall model have been described in separate publications (Lovett et al., 2000; Lovett et al., 2012; Morris et al., 2012).

The intervention included:

a. Decoding and word identification components that focus on the teaching of five word identification strategies. Most individuals with DD experience persistent decoding problems (as described above) and so decoding instruction was a base onto which the other intervention components were scaffolded. The five decoding strategies were Sounding Out, Rhyming, Peeling Off, Vowel Alert, and Spy. Every lesson allocated time to acquiring the skills and knowledge needed to execute the strategies successfully. Using a metacognitive Game Plan, children selected, applied, monitored and evaluated their application of these strategies. The metacognitive Game Plan can be thought of as related to the Central Executive components of the

WM system in its emphasis on manipulation and monitoring of previously learned information (described by Lovett et al., 2000) which are traits integral to a sound WM system as described by Baddeley and Hitch (1974) and Engle et al. (1999). In this way, the PHAST reading program can be thought of as drawing on specific WM skills although there is no direct instruction of WM, and therefore some relation between WM and reading outcomes following PHAST treatment can be hypothesized.

b. Comprehension components that teach text comprehension strategies (e.g., predicting, setting goals, clarifying, questioning, summarizing) used a combined skills and strategy instruction approach. Pre-skills (explicit instruction on conventions of text structure for narrative, expository, and other text genres) necessary to successfully implement the comprehension strategies were taught directly.

c. Vocabulary instruction that is integrated into the text reading activities to develop deeper word knowledge (Kamil et al., 2008; Klinger et al., 2007). Using new vocabulary that was encountered in the program texts, teachers pre-taught word meanings, and engaged the students in activities such as vocabulary elaboration, semantic mapping, and semantic feature analysis (Beck & McKeown, 2007).

d. Fluency components that develop word identification efficiency and text reading fluency. Fluency exercises provided repeated and varied practice on learned sounds, keywords, affixes, irregular words, and difficult words. Fluency was also emphasized at the text level. Passages that have been read independently and

accurately were reread and children were encouraged to read quickly without errors. Children practiced daily to increase accuracy and reduce reading times.

e. Motivational Components that address the maladaptive attributions and low motivation for reading (Wigfield et al., 2008) often demonstrated by struggling readers—a tendency that compounds with successive years of reading difficulty. To foster reading engagement and accelerate remediation, attribution retraining and motivational reshaping were woven into the strategy dialogue that children acquire during intervention.

Each instructional group was monitored by a lead teacher to insure implementation integrity, and multiple methods were used to teach and test similar constructs to obtain high inter-method reliability and re-enforce reading concepts and tools.

Participants were tested on a series of psychoeducational and cognitive tests prior to receiving any intervention (described in the previous and subsequent sections). Follow up reading related tasks were administered, after 23 hours of instruction, at 45 hours of intervention, and after 70 hours of intervention were completed.

1.11.1 Reading Intervention Gain Indicator:

To establish gains or losses made over the course of the reading intervention, a mathematical computation of the level of reliable change index (RCI) from each participant’s pre- and post-intervention baseline scores was performed. The RCI is calculated by dividing the absolute change (the difference between post-intervention performance (Time 70) and pre-intervention (Time 0) performance) by the standard error of the pre-post difference. This results in a z score RCI for each measure. Based on the

creation of z-scores, each subject's reliable change index (RCI) is distributed around 0 so that positive values represent levels of change above chance for that measure, while negative values still represent some level of change, but with decreasing probability of significance.

In order to develop a more reliable index of change, this study utilized 4 tests: WJ-III Word Attack, WJ-III Word Identification, Test of Transfer and Challenge Words Test to create a composite reliable reading change score for single word reading/decoding (untimed). This composite was created by averaging the z-scores from the 4 measures transformed reliable change index. Brief descriptions of each test are summarized below:

1.11.2 Initial Reading intervention scores composite:

Preliminary research on these intervention data have shown that initial testing scores show strong relations to reading score changes as well as cognitive indicators used in this study (Riggall et al., 2017; Winter, Frijters, Lovett, & Morris, 2016; Winter, Riggall, Branum-Martin, Frijters, & Morris, 2018). As such, a composite of the z-scores across initial testing (Time 0) for each of the previously mentioned tests (WJ-III Word Attack, WJ-III Word ID, Test of Transfer and Challenge Words Test) was also included as an external indicator for each model. Composite z-scores were derived by computing an individual z-score for each participant utilizing the mean and SD of the entire study sample. Z-scores across all four measures were then added and divided by four to yield an average initial reading z-score for each participant.

1.11.2.1 WJ-III Word Attack and Word Identification:

The Woodcock Johnson Test of Achievement-III (WJ-III; Woodcock et al. 2001) assesses academic abilities.

The WJ-III utilizes two tests to measure single word reading/decoding abilities. The **Letter-Word Identification** task assesses reading decoding abilities by asking subjects to identify printed letters and words. **Word Attack** measures non-speeded reading decoding and phonological skills by asking subjects to read phonically regular non-words aloud.

The WJ-III has been normed using approximately 8,800 individuals representative of the U.S. population ages 24 months to 90 years of age (McGrew & Woodcock, 2001). The normative sample was controlled for 10 community and individual variables and 13 socio-economic variables.

Reliabilities for these tasks were calculated using split-half coefficients using an odd and even number split. Reliabilities on speeded subtests were calculated using Rasch analysis procedures. The subtests included in this study all show high reliability and validity. The reliability calculated for the Letter-Word Identification subtest is 0.94, and for Word Attack is 0.87.

1.11.2.2 Test of Transfer

The Test of Transfer measures transfer-of-learning effects and consists of 30 uninstructed words that vary systematically from keyword spelling patterns. For example, the keywords *bug*, *pack*, and *end* are represented by transfer probes *pug*, *puck*, and *endless*. Importantly, items on this test are restricted from intervention materials and activities. Thus, this test targets transfer of learning and generalization in printed language that are demonstrated deficits in reading disorders (Benson, Lovett, & Kroeber, 1997). Specifically, analysis of raw scores on this test allows for evaluation of transfer of rhyme, onset, and letter-sound sub-syllabic segments of the instructed keyword patterns. Additionally, previous research demonstrates that this measure is psychometrically appropriate for human

growth-curve modeling and is a more sensitive index than standardized tasks to assess responsiveness to the various interventions (Lovett, Steinbach, & Frijters, 2000; Morris et al., 2012). It consistently produces 70-hour treatment effect sizes ranging from .65 to .85.

1.11.2.3 Challenge Words Test

The Challenge Words Test consists of 30 uninstructed, multisyllabic words that incorporate common spelling patterns and affixes from keywords taught during treatment. Similar in design to the Test of Transfer and Keyword Test, this test is the most challenging and uses longer and more complicated uninstructed words (e.g., *mistakenly*). Raw scores are used to quantify performance and subjects rarely achieve more than 10 correct among 30 words.

This measure represents a more difficult decoding task that explicitly measures metacognitive transfer of the decoding strategies taught in the curriculum in this study. The Challenge Words Test has also been shown to be a sensitive index of transfer of learning for students with DD (Lovett et al., 2000; Lovett, Borden, DeLuca, Kacerenza, et al., 1994), consistently producing 70-hour treatment effect sizes ranging from .65 to .85.

1.12 Analysis Approach

To analyze the collected data, we first processed the data in SAS to determine means, and check distributions of each variable for skewness, outliers and non-normality. Following data processing, the three proposed models were run in MPLUS using confirmatory factor analysis (CFA) to determine goodness of fit and the relative relation between variables and the relation between these proposed factors and reading intervention outcomes. Given the limitations of the available assessment data, some latent factors were supported by only one indicator, and were thus represented by phantom factors (single-

indicator factors), to insure proper model identification. In our exploration of working memory and its relation to reading intervention outcomes in a group of children with DD, we used MPlus Version 7.3 (Muthén & Muthén, 2014). Better model fit (Hu & Bentler, 1998) was characterized by the following: (1) non- significant chi-square values; (2) SRMR values of less than or equal to .05; (3) RMSEA values less than or equal to .08; and (4) CFI values greater than .90 (Marsh, Hau, & Grayson, 2005; Marsh, Hau, & Wen, 2004).

Based on initial model results, modifications of each model were run to test the direction of the relation between the variables based on the theoretical hypothesis.

1.13 Descriptive Statistics

Descriptive statistics are provided to document each cognitive skill examined in the study. See Table 2 for complete descriptive statistics. Table 2 includes the mean, standard deviation, median, range of raw scores, skew and kurtosis statistics as well as the standard error for the skew and kurtosis statistics for all variables for the final tally of 108 subjects included in this study.

Table 2: *Descriptive Statistics for Whole Sample of Elementary School Students with Developmental Dyslexia (N=108)*

Variable	*M*	*SD*	*Range*	*Skew(SE)*	*Kurtosis(SE)*
TNWR	81.57	6.68	31	-.48(.20)	-.40(.41)
Digit Span FW	7.60	1.68	7	.32(.22)	-.61).43)
Digit Span BW	5.56	1.17	7	.43(.22)	.94(.43)
Spatial Span FW	5.91	1.66	10	.30(.21)	.45(.41)
Spatial Span BW	4.81	2.05	13	.50(.21)	1.08(.41)
CT-SPAN	8.67	6.64	33	1.38(.21)	2.33(.42)
PPVT4	138.78	22.73	116	-.05(.20)	-.12(.41)
CTOPP2 PA	19.28	6.81	30	-.73(.20)	.10(.41)
CELF RS	51.21	12.84	66	.28(.20)	-.15(.41)
WASI2 Matrices	13.14	4.33	20	.02 (.20)	-.59(.40)
Reading Gains	-.13	.60	3.64	-.41(.20)	1.13(.41)
Initial Reading Score	-.004	.91	4	.69(.20)	.28(.41)

Note: M = Mean, SD = Standard Deviation, CT-SPAN = Counting Span, TNWR = Test of Non-word Repetition, Digit Span FW = Digit Span Forward, Digit Span BW = Digit Span Backward, Spatial Span FW = Spatial Span Forward, Spatial Span BW = Spatial Span Backward, PPVT = Peabody Picture Vocabulary Test-4, CTOPP PA = Comprehensive Test of Phonological Processing Phonological Awareness Composite, CELF RS = Clinical Evaluation of Language Fundamentals Recalling Sentences. CTOPP PA is a composite raw score based on three CTOPP subtests described in Methods. Reading Gains and Initial Reading Score are composite z-scores. All other scores are raw scores.

Demographic data including gender, ethnicity, age at first testing time point, and grade are displayed in Table 3.

Table 3: Demographic information for DD sample

Characteristic	***DD participants N=108***
Ethnicity (% White, African American, Asian, Hispanic, Biracial)	17% Caucasian 77% African American <1% Asian 2% Hispanic 4% Biracial
Age at testing M (SD)	9.01 (0.70)
Grade (% 3rd grade)	56% 3rd grade
Gender (% male)	55% male

In an attempt to further understand the WM abilities of our sample as compared to typically developing children, mean scaled scores and their distribution were examined for four tasks theorized to represent WM: Digit Span Forward and Backward and Spatial Span Forward and Backward (Table 4). The data indicate that all scores were normally distributed and that this sample fell within a normal range across all four measures as compared to typically developing children of a similar age where scaled scores of 10 with standard deviations of 3 are typical (Flanagan & Alfonso, 2017). Additionally, this sample showed a wide range of WM abilities indicating that relations between WM and other factors would not be suppressed due to a lack of versatility in the data

Table 4: Means of Scaled Scores for four normed tests of WM

Variable	*M*	*SD*	*Range*	*Skew*	*Kurtosis*
Digit Span FW	9.18	2.61	13	0.44	-0.15
Digit Span BW	8.33	2.30	11	0.20	0.15
Spatial Span FW	9.01	2.89	15	0.44	0.11
Spatial Span BW	8.78	3.22	18	0.38	0.51

Note: M = Mean (of Scaled Scores), SD = Standard Deviation, FW = Forward, Backward = Backward. Distribution of scaled scores on the WISC-V is M = 10, SD = 3 across all four measures.

1.14 Data Screening

Normality was evaluated for each variable. Per Kline (2005), normal probability scatter plots were examined for each variable to determine skew and kurtosis. All variables fell below Kline’s suggested cutoffs of <3 and <10 for skewness and kurtosis respectively and appeared to be normally distributed.

The data were also screened for collinearity by examining correlations between variables which were r>.90. Using this criterion, none of the variables showed collinearity, thus all variables were used in the models.

For outlier analysis, univariate data points more than 3 *SD*s either above or below the mean of each age sample were investigated. All outliers were first checked for data scoring

or entry errors by confirming their values by cross-checking the original protocols. Fewer than one percent of participants met this 3 SDs criteria on a single measure and participants were thus retained for subsequent analysis.

1.15 Correlations

Pearson correlations were conducted as a means of measuring the direction and strength of the relationships between the model variables (Table 5). Moderate correlations (.39 - .75) were found between language and reading measures including Non-Word Repetition, Digit Span Forward, CTOPP, CELF Repeating Sentences and PPVT. Similarly, measures involving visuo-spatial abilities, including Spatial Span Forward and Backward and WASI Matrices also showed moderate correlations between each other. WASI Matrices was also moderately correlated with the PPVT giving credence to the Engle et al. (1999) hypothesis that vocabulary abilities are part of a general intelligence factor. Surprisingly, Digit Span Forward and Backward were not even moderately correlated indicating perhaps that they draw on different cognitive processes in this sample. Initial Reading Scores was moderately correlated with Non-Word Repetition, PPVT, CTOPP, CELF Sentence Repetition, and WASI Matrices lending support for language and fluid intelligence abilities as indicators of initial reading performance in struggling readers.

Table 5: Pearson correlations between variables

	1.	**2.**	**3.**	**4.**	**5.**	**6.**	**7.**	**8.**	**9.**	**10.**	**11.**	**12.**
1. Non word repetition	1.00											
2. Digit Span FW	.36*	1.00										
3. Digit Span BW	.06	.06	1.00									
4. Spatial Span FW	.03	-.05	-.07	1.00								
5. Spatial Span BW	.02	-.15	.10	.44*	1.00							
6. Counting Span	.06	.20	.08	.16	.07	1.00						
7. PPVT	.31*	.07	.08	.24*	.24*	.05	1.00					
8. CTOPP PA	.28*	.05	.09	.19	.10	-.13	.28*	1.00				
9. CELF RS	.35*	.42*	.08	.18	.19	.27*	.54*	.18	1.00			
10. WASI Matrices	.02	-.09	-.02	.32*	.41*	.07	.26*	.17	.12	1.00		
11. Initial Reading Score	.22*	.14	.13	.05	.07	.05	.30*	.30*	.37*	.21*	1.00	
12. Reading Score Gains	.05	-.02	-.09	.13	-.00	.02	.12	.16	.03	-.16	.04	1.00

Note: ** connotes statistically significant (<0.05) correlations.*

RESULTS

Each of the three models outlined in the introduction was tested. During this process, attention was focused on identifying the better fitting model. All model results are presented in Table 6 and individual model results are discussed below.

Table 6: Model results

Model Name	χ^2	*df*	*p*	CFI	SRMR	RMSEA
Model 1: Baddeley and Hitch 'Memory' Model	43.65	33	0.18	0.96	0.05	0.04
Model 2: Engel et al. 'executive' model	97.33	44	<0.01	0.68	0.09	0.11
Model 3: Brown and Hulme 'Language' Model	85.06	47	<0.01	0.78	0.09	0.09

1.16 Model 1: Baddeley and Hitch (1976) WM Model

The Baddeley and Hitch (1976) model (Model 1) included 3 latent factors, each with at least 2 indicators and 2 phantom factors (see Figure 12). The model also included two external indicators, Reading Intervention Gains and Initial Reading Scores as described in the methods section. This model, driven by the tripartite working memory theory, fit the data very well (see Table 7). The model met the Chi square (.18), CFI (.96), RMSEA (.04) and SRMR (.05) goodness of fit criteria. Aside from a single indicator, loadings across all three factors, Phonological Loop, Visuo-spatial sketchpad and Language were all moderate to strong >.39. The Digit Span Backward indicator showed a weak relation to the higher factor structure (Phonological Loop) at .13, consistent with its low correlation to Digit Span

Forward. Among factors, the relation between Language and Intelligence (.36); Phonological Loop and Central Executive (.36), Visuo-Spatial Sketchpad (.55), Phonological Loop and Language (.37); and Visuo-Spatial Sketchpad and Intelligence (.56) were all moderate to high. Initial Reading Scores showed significant relations to Phonological Loop (.29), Language (.44) and Intelligence (.21) factors. No factors were significantly related to the Reading Intervention Gains indicator. This model was also run to include an additional phantom factor 'Episodic Buffer' with CELF Sentence Repetition as its sole indicator. The model and its fit are described in Appendix A.

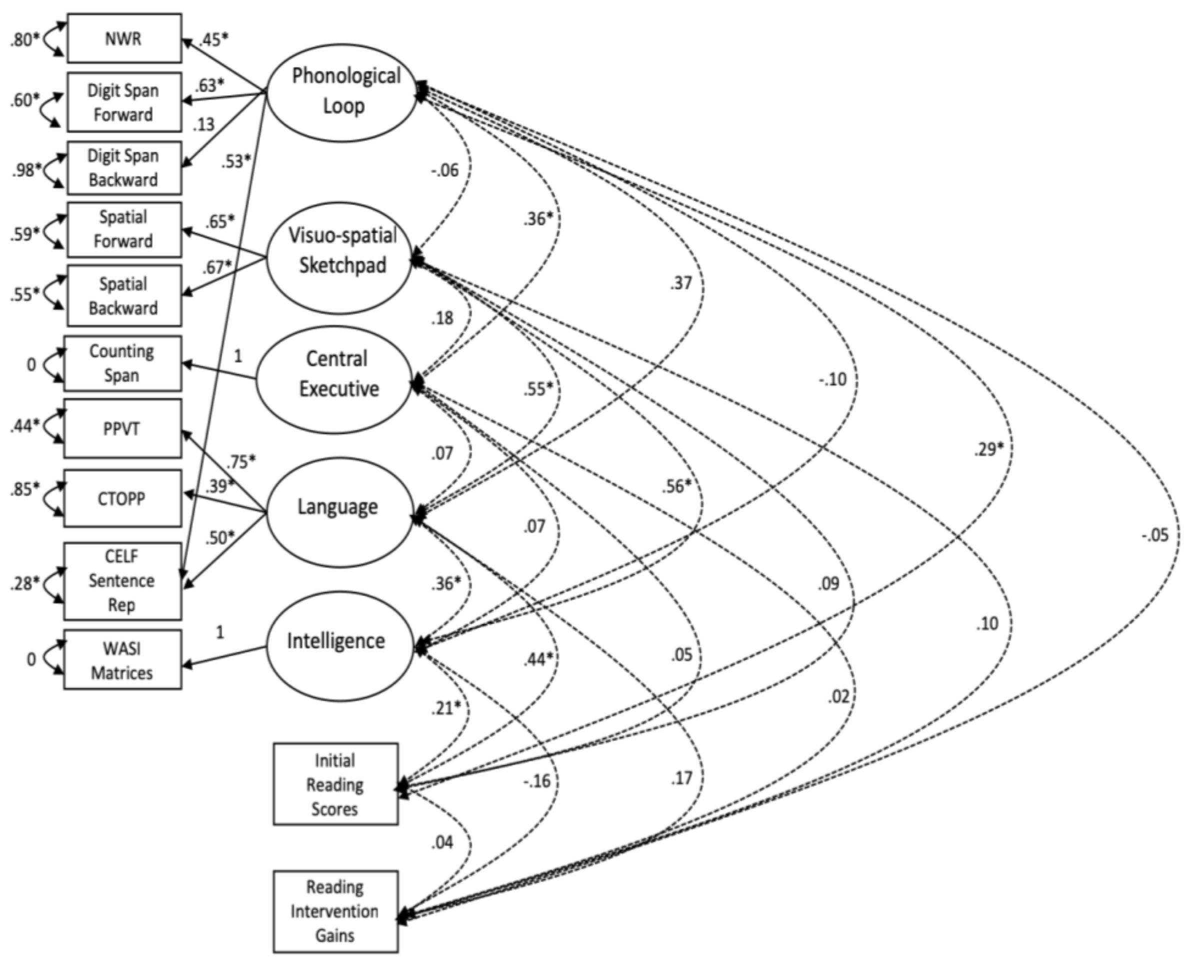


Figure 9: Model 1 findings

Table 7: Standardized model results for Model 1

Parameters	Relation/Variable	Estimate	SE	Ratio	p-value
Loadings	Phonological loop by NWR	0.46	0.10	4.39	<0.01
	Phonological Loop by Digits Forward	0.64	0.12	5.25	<0.01
	Phonological Loop by Digits Backward	0.13	0.11	1.13	0.26
	Phonological Loop by CELF Sentence Rep	0.53	0.11	4.76	<0.01
	Visuo-spatial Sketchpad by Spatial Forward	0.64	0.09	6.89	<0.01
	Visuo-spatial Stetchpad by Spatial Backward	0.68	0.09	7.20	<0.01
	Central Executive by Counting Span	---	---	---	---
	Language by CELF Sentence Rep	0.50	0.13	3.93	<0.01
	Language by PPVT	0.74	0.09	8.56	<0.01
	Language by CTOPP Phonological Abilities	0.38	0.10	3.66	<0.01
	Intelligence by Matrix Reasoning	--	---	---	---
Latent Covariances	Initial Reading Score with Phonological Loop	0.32	0.12	2.54	0.01
	Initial Reading Score with Visuo-spatial Sketchpad	0.09	0.12	0.75	0.45
	Initial Reading Score with Language	0.42	0.11	3.82	<0.01
	Initial Reading Score with Central Executive	0.05	0.10	0.55	0.58
	Initial Reading Score with Intelligence	0.19	0.09	2.06	0.04
	Change Score with Phonological Loop	-0.07	0.12	-0.61	0.54
	Change Score with	0.09	0.12	0.71	0.48

	Visuo-spatial Sketchpad Change Score with Language	0.15	0.11	1.35	0.18
	Change Score with Central Executive	0.02	0.09	0.21	0.83
	Change Score with Intelligence	-0.15	0.09	-1.66	0.09
	Visuo-spatial Sketchpad with Phonological Loop	-0.06	0.18	-0.36	0.72
	Language with Phonological Loop	0.35	0.22	1.57	0.12
	Language with Visual Spatial Sketchpad	0.54	0.14	3.89	<0.01
	Central Executive with Phonological Loop	0.36	0.12	2.98	<0.01
	Central Executive with Visuo-spatial Sketchpad	0.17	0.12	1.42	0.16
	Central Executive with Language	0.08	0.13	0.60	0.55
	Intelligence with Phonology	-0.12	0.13	-0.98	0.33
	Intelligence with Visuo-spatial Sketchpad	0.55	0.10	5.51	<0.01
	Intelligence with Language	0.37	0.11	3.35	<0.01
	Intelligence with Central Executive	0.06	0.10	0.64	0.52
Residual Variances	NWR	0.79	0.10	8.12	<0.01
	Digit Span Forward	0.60	0.15	3.85	<0.01
	Digit Span Backward	0.98	0.03	34.27	<0.01
	Spatial Span Forward	0.59	0.12	4.94	<0.01
	Spatial Span Backward	0.54	0.13	4.29	<0.01
	CTOPP Phonological Awareness	0.86	0.08	10.98	<0.01
	CELF Sentence Rep	0.28	0.11	2.64	<0.01
	PPVT	0.45	0.13	3.46	<0.01

Note: Estimates are considered significant for p-values less than or equal to 0.05.

1.17 Model 2: Engle et al. (1999) Model

The Engle et al. (1999) model (Model 2) consisted of two factors and two phantom factors with the two external indicators used in the previous models (Reading Intervention Gains and Initial Reading Scores; see Figure 13). The model did converge however criteria for Chi-square (<.01), CFI (.68), RMSEA (.11) and SRMR (.09) were not met. The STM factor showed significant loadings for all factor indicators except for Digit Span Backward (.12). The Intelligence factor showed significant loadings for both indicators, however the PPVT showed a much stronger relation to its parent factor than WASI Matrices (.99 compared to .26), although both loadings were significant. The STM factor was significantly related to Central Executive (.32), Language (.24), and Intelligence (.62) factors. The Language and Intelligence factors were also significantly related (.29). The Initial Reading Scores indicator was significantly related to Intelligence (.30), Language (.30), and STM factors (.42), while Reading Intervention Gains did not show relations to any of the factors.

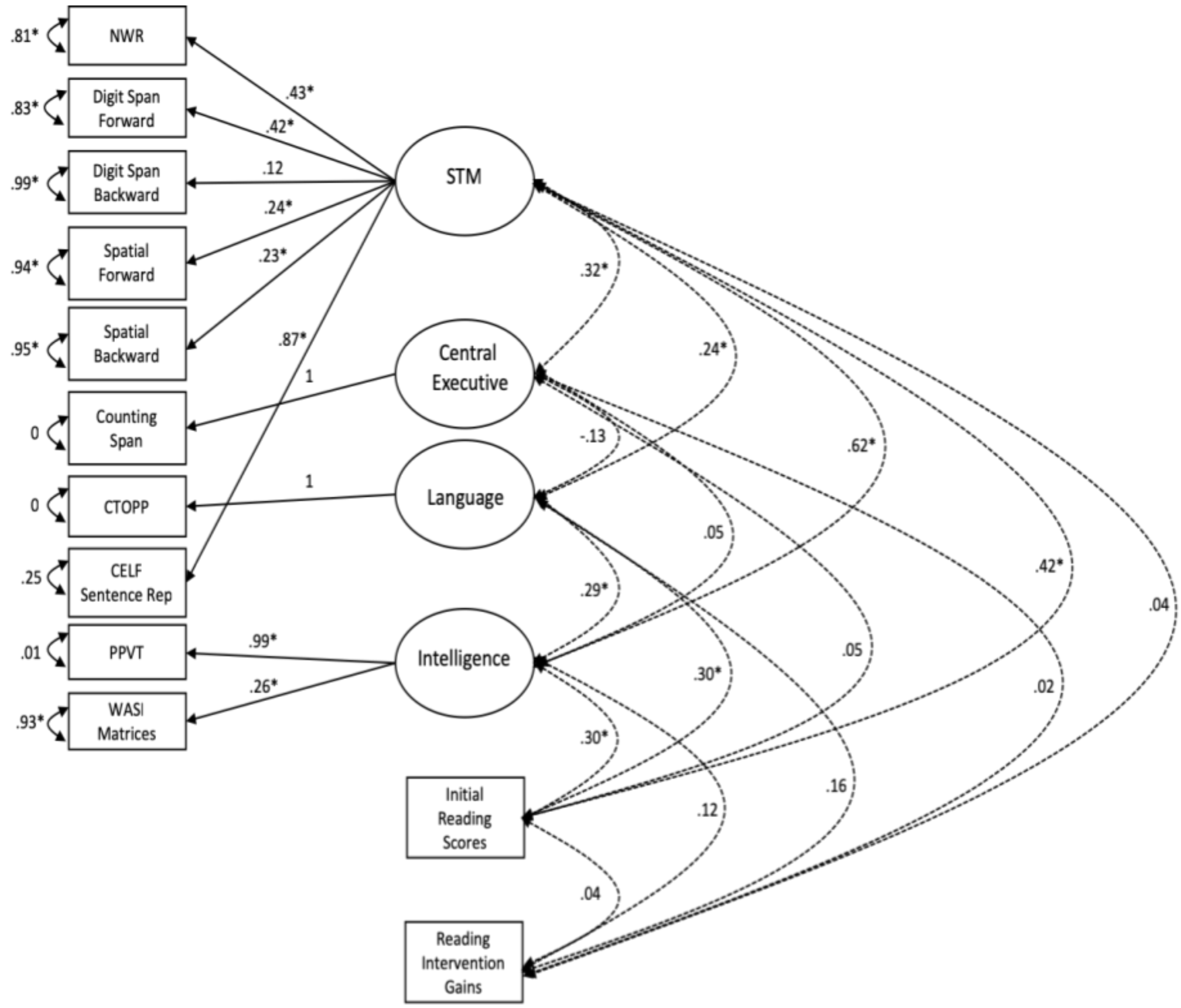


Figure 10: Model 2 findings

Note: *This model was also run with the CELF Sentence Repetition task loading onto both the STM and Language Factors but it did not converge thus subsequently the CELF was defined as part of the STM parent factor only yielding the model presented in Figure 13.*

1.18 Model 3: Brown and Hulme (1996) Language Model

The Brown and Hulme (1996) model (Model 3) was a two-factor model with a single phantom factor and two external indicators: Initial Reading Scores and Reading Intervention Gains (see Figure 14). Although the model converged, it did not meet criteria for Chi square (.01), CFI (.78), RMSEA (.09) and SRMR (.09) goodness of fit, indicating that it is an inferior model to the previously tested Baddeley and Hitch model. Attempts to improve on this model using theoretically driven hypotheses such as causally linking

Language to Working Memory, showed similarly poor fit results. All indicators for the Language Abilities factor were significant. The Memory Abilities factor showed both Spatial Spans Forward (.66) and Backward (.65) as significant indicators, but Counting Span (.19), Digit Span Forward (-.10) and Digit Span Backward (.01) were not significantly related to the overall factor structure. Among factors, Memory Abilities was significantly correlated with both Language Abilities (.41) and Intelligence (.55), and Language Abilities was also significantly related to Intelligence (.26). Both Intelligence (.21) and Language Abilities (.48) factors showed significant relation to Initial Reading Scores, while no factor was significantly related to Reading Intervention Gains. An alternative model showing the Digit Span tasks as part of the Language Abilities factor yielded similar results and can be found in Appendix B.

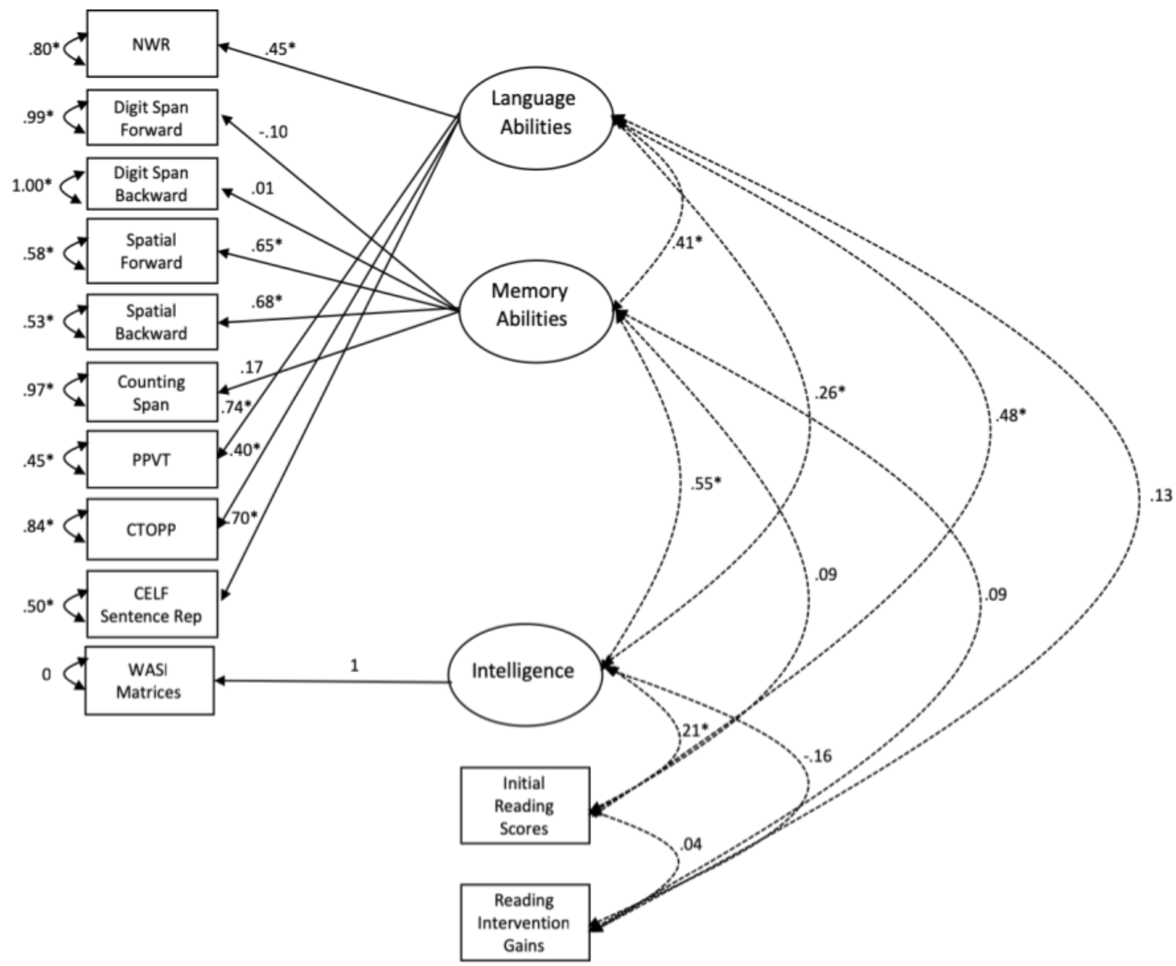


Figure 11: Model 3 findings

1.19 Ancillary Analysis

1.19.1 Factor Structure of STM in the Engle et al. (1999) Model

In order to determine the elements contributing to the poor fit of the Engle et al. (1999) model (Model 2), separate analysis was run on the STM factor to determine its goodness of fit without the variance contributed by the rest of the model factors (Figure 15). Although the model converged it did not pass the Chi-square ($p<0.05$), CFI (.57), RMSEA (.17) and SRMR (.10) fit criteria. Upon further examination, only three out of the six indicators, Non-Word Repetition, Digit Span Forward and CELF Sentence Repetition,

significantly related to the STM factor, indicating that perhaps the indicators would be better fit by at least two factors as displayed in Model 1.

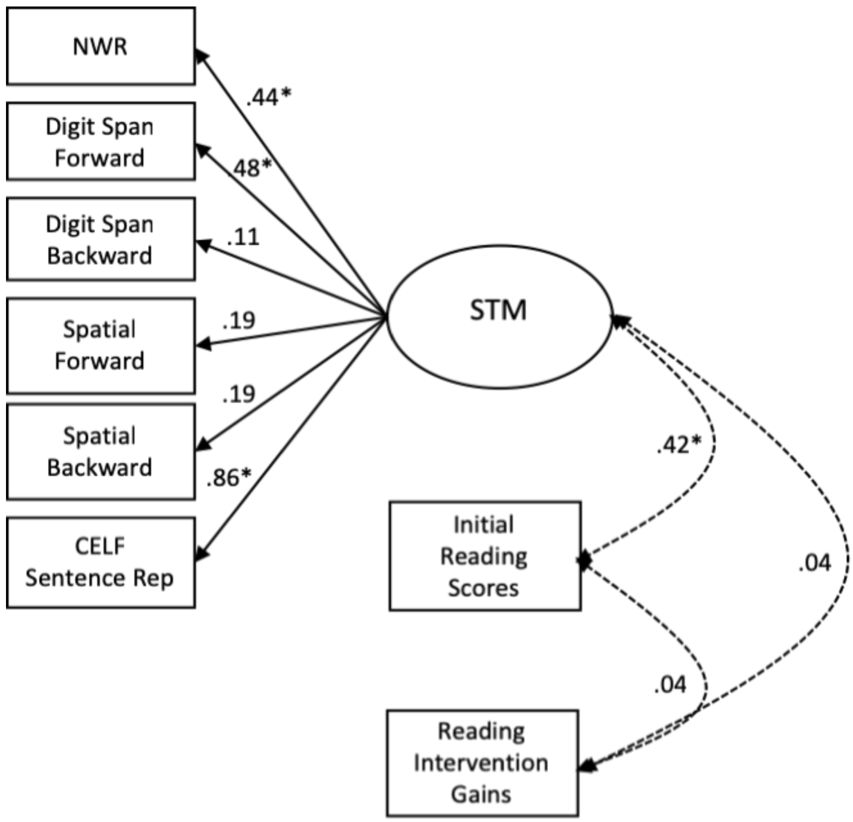


Figure 12: STM single factor model

1.19.2 Exploratory Factor Analysis (EFA)

Following our theory based model testing, we ran a post-hoc exploratory factor analysis (EFA) on the 10 memory and language indicators to determine statistically optimal groupings of the indicators. Although EFAs are generally performed where no theoretical or conceptual model exists, a post-hoc EFA can help uncover indicators which are leading to poorer model fit and also revealing cross-loadings between indicators not previously hypothesized by the theoretical model (Graham, Guthrie, & Thompson, 2003). Our results

showed that models with 1, 2 and 3 factors all converged, but only models with 2 and 3 factors met our previously described goodness of fit model criteria (Table 8). Additionally, in a comparison of Chi squared values the 3-factor model was confirmed as the best-fit model ($p<0.01$) when directly compared to the 2-factor model. The model with 4 factors did not converge and will not be further discussed

Table 8: EFA model estimates for 1, 2 and 3 factor models

							Comparison			
Model Name	χ^2	*df*	*p*	CFI	SRMR	RMSEA	$\Delta\chi^2$	Δdf	*p*	Note
1 factor	96.20	35	<0.01	0.59	0.12	0.13	57.76	9	<0.01	vs. 2 factor
2 factor	38.44	26	0.06	0.92	0.06	0.07	16.63	8	0.03	vs. 3 factor
3 factor	21.81	18	0.24	0.98	0.04	0.04	74.39	17	<0.01	Vs. 1 factor

The 2-factor model (Table 9) showed significant loadings for Spatial Span Forward, Spatial Span Backward, and WASI Matrices loading onto the first factor and Non Word Repetition (NWR), Counting Span and CELF Sentence Repetition loading onto the second factor. Both Digit Span Forward and PPVT cross-loaded on both factors, while CTOPP and Digit Span Backward did not load onto either.

Table 9: EFA Geomin Rotated Loadings for 2-factor model

Variable	***Factor 1***	***Factor 2***
TNWR	-0.12	0.50*
Digit Span Forward	-0.40*	0.60*
Digit Span Backward	0.00	0.11
Spatial Span Forward	0.57*	0.06
Spatial Span Backward	0.69*	0.00
Counting Span	0.00	0.28*
CTOPP	0.18	0.21
CELF Sentence Repetition	0.01	0.85*
PPVT	0.27*	0.52*
WASI Matrices	0.57*	-0.00

(is significant at 0.05)*

The 3-factor model (Table 10) indicated significant loadings for Spatial Span Forward and Backward and WASI Matrices for the first factor. The second factor showed significant loadings for NWR, Counting Span, CELF Sentence Repetition and PPVT, with Digit Span Forward cross-loading across both factors similar to our 2 factor EFA model. Interestingly, the third factor consisted of only NWR, while similar to our 2-factor model, Digit Span Backward and CTOPP did not load significantly onto any of the factors.

Table 10: EFA Geomin Rotated Loadings for 3-factor model

Variable	***Factor 1***	***Factor 2***	***Factor 3***
TNWR	-0.02	0.38*	0.34*
Digit Span Forward	-0.32*	0.58*	0.01
Digit Span Backward	0.03	0.08	0.05
Spatial Span Forward	0.58*	0.03	-0.03
Spatial Span Backward	0.69*	-0.00	-0.11
Counting Span	-0.00	0.38*	-0.34
CTOPP	0.29	-0.00	0.59
CELF Sentence Repetition	0.10	0.86*	-0.02
PPVT	0.36	0.42*	0.24
WASI Matrices	0.58*	-0.05	0.03

(is significant at 0.05)*

DISCUSSION

This study examined the best-fit theoretical model for language and WM data collected on a group of elementary school children who were identified with DD. Within this framework, these language and WM measures were also examined to understand their relation to the children's initial reading scores as well as the changes in reading that were made over the course of a 70-hour reading intervention. More specifically, our study examined: 1. Whether the Baddeley and Hitch (1974) model would have the strongest fit. 2. Within that model, whether the Phonological Loop latent factor would have the strongest relation to and share the most variance with reading intervention gains.

1.20 The best fit model

Results supported our hypothesis that the Baddeley and Hitch (1974) (Model 1) was the best fitting model for data collected on a group of children with DD. Despite the other two models' convergence, they produced poorer fit, allowing us to conclude the statistical strength of Model 1 in describing this sample of children with DD given this set of measures. Additionally, Model 1 (the Baddeley and Hitch model) was also a better

theoretical and statistical fit than any of the EFA models produced. Although this model of WM has been previously explored theoretically (Becker, MacAndrew, & Fiez, 1999; Chien and Piez, 2010; Wilson, 2000 among others), in children with low achievement in a national curriculum (Gatherole & Pickering, 2000), and to test the relation between WM and early writing (Bourke & Adams, 2003), this is the first time it is operationalized and statistically supported for a group of children with DD. Across all three factors of the model, all indicators were significantly (.38 to .75) related to each parent factor with the exception of Digit Span Backward (.13). Additionally, the Phonological Loop factor was significantly related to the Central Executive factor as is predicted by the original paper (Baddeley & Hitch, 1974) and other related memory research (Cowan et al., 1995). These results suggest that phonological processes, as measured by the model indicators, share similarities to a higher order WM process as indicated by the Counting Span task.

Model 1 showed better fit than the two competing models due to its separation of phonological and visual-spatial WM components, where the Engle et al. (1999) model (Model 2) advocated these factors be subsumed into a larger STM component, and the Brown and Hulme (1996) model (Model 3) theorized the convergence of language and phonological memory indicators. Our ancillary analysis of this individual STM factor, as well as the use of EFA, supported a clear separation of these components using this set of indicators. Interestingly, relations between factors were unexpectedly low between Phonological Loop and Visuo-spatial Sketchpad (-.06), and Visuo-spatial Sketchpad and Central Executive (.17), despite the Baddeley and Hitch model advocating for these factors as being distinct but related. Numerous studies have shown that visuo-spatial short-term memory abilities in DD are similar (e.g. Jeffries & Everatt, 2004; Kibby, Marks, Morgan, &

Long, 2004) or superior to age-matched controls (Archibald & Gathercole, 2006), leading these researchers to conclude, similar to our current findings, that verbal and spatial WM within this population are discernibly separate entities. However, others have found decreased performance and inefficiencies on visuo-spatial WM tasks (primarily spatial span tasks; Bacon, Parmentier, & Barr, 2013) for individuals with DD (Menghini et al., 2011; Smith-Spark & Fisk, 2007), lending evidence for a common deficit in both phonological and visual spatial abilities, perhaps based on related executive functioning or cognitive load challenges within this population.

In an attempt to further understand this specific observed deficit, Bacon et al. (2013) tested adults with DD on the process demanding Spatial Span Backward task after supplying participants with a visual strategy prior to taking on the task. They found that with training, DD individuals were able to overcome their deficit and perform similarly to typical individuals, lending evidence to reliance on a faulty verbal system while attempting the original Spatial Span Backward task. These authors concluded that reliance on a faulty verbal system when attempting visual tasks might be partially responsible for their poor performance, and that introduction of a visual strategy helped with successful performance, again lending evidence to related but distinct phonological and visual-spatial systems (Bacon et al., 2013). These authors and others also conclude that failure to adopt a preferable strategy or adapt to task demands is evidence for a more general executive functioning deficit (Bacon et al, 2013; Smith Spark & Fisk, 2007; Vavara, Varuzza, Sorrentino, Vicari, & Menghini, 2014), indicating that our current findings of a non-significant relation between visual-spatial WM and executive functioning (.18) likely reflects the strong use of verbal over spatial abilities in the Central Executive Counting Span

task. Indeed, these results show a significant relation (.36) between the Phonological Loop and Central Executive, which is likely task specific.

Model 1 was also superior to the language driven model (Brown & Hulme, 1996; Model 3) in its separation between language and phonological WM, as well as its separation between visual spatial memory and memory requiring higher order manipulation (i.e., the Central Executive indicated by the Counting Span task). Indeed in Model 3, while the Spatial Span forward and backward tasks both loaded significantly onto the Memory Abilities factor, the Counting Span task did not, indicating a distinction within these tasks. Additionally, within the Baddeley and Hitch model (Model 1), while the Phonological Loop factor showed significant relations to the Central Executive, the Language factor did not (.08), indicating that aside from shared linguistic components between phonology and executive functioning, a higher order processing is shared between phonological WM tasks and Central Executive which is distinct from the language tasks used as indicators.

Importantly though, despite the clear preference for the model indicators to separate themselves into Phonological, Visuo-spatial and Central Executive latent factors, the Language latent factor contributed significantly to the model by sharing strong relations with the Phonological Loop, Intelligence and interestingly, the Visuo-spatial Sketchpad latent factors. Relations between phonological working memory and language have been robustly replicated (Acheson & MacDonald, 2009; Baddeley, 1966a; Brown et al., 1996; Dollaghan & Campbell, 1998; and others), as have relations between language and intelligence (described by Engle et al., 1999). In fact, the shared variance found between language and our perceptual intelligence indicator, Matrices, might shed light on our findings of a relation between language and visuo-spatial working memory. Intelligence

tests measured by tasks like Matrix Reasoning have been thought to tap into the one-dimensional construct of fluid intelligence (G*f*; Horn, 1983; Raven 2000), however recent research has also placed increased importance on the visuo-spatial tasks in their contribution to G*f*, indicating that G*f* as measured by Matrix Reasoning tests like those on the WASI or Ravens might be pulling for a stronger visuo-spatial component than uniform a-modal fluid intelligence attribute (Stephenson & Halpern, 2013). Fluid intelligence is one theoretical part of general intelligence (G) the second of which is crystallized intelligence (G*c*) classically described by Cattell (1951) as encompassing vocabulary and general knowledge.

Given that these two types of intelligence together form a proxy for G as measured by IQ tests like the WAIS, WISC or WASI, it would follow that language tasks such as receptive vocabulary (i.e., the PPVT in our study) and phonological awareness (i.e., the CTOPP in our study) would show strong relations to visuo-spatial tasks (i.e., Corsi Blocks in our study). In light of the above finding, the minimal overlap (-.06) between our Visuo-spatial and Phonological factors indicates the that WM components of visuo-spatial and phonological tasks are distinct (as previously noted by Baddeley 2000) and points to the specialized relation between language and spatial tasks which is separable from memory.

1.21 Relations between Factors and Initial Reading Abilities and Intervention Gains in Reading

The statistical significance found between numerous factors in our study and *initial* reading scores has been replicated by numerous researchers. For example, tests of phonological abilities have long been used to diagnose DD and have been highly predictive of later reading problems in young children (Saygin et al., 2013; Torgesen, 1997; 2005). However, some have even argued that the utility of such specified cognitive tests wears off

once children have been in school for a number of years, and that replacing such tests with achievement assessments mimicking what they learn at school offers much stronger content validity of assessment (Fletcher & Miciak, 2017). For example, in their paper analyzing the utility of progress monitoring slopes and change scores in predicting reading outcome, Tolar, Barth, Fletcher, Francis and Vaughn (2014) assert that the best predictor of future reading performance in older elementary and middle school children is an assessment that a) has occurred closest in time to the present reading evaluation and b) shows significant content validity with the present reading evaluation (e.g., a comprehension measure taken soon before a comprehension exam). Following this model, the strongest predictor of future reading abilities in our sample would be our initial reading measures' average z score.

Our results across all three models show moderate relations (.21 - .44) between phonology, language and intelligence factors and initial reading scores. These findings are strongly supported by previous literature findings of general language ability, phonological awareness and IQ all being strongly related to reading abilities, but are not predictive of reading changes following intervention (Burns, 2016; Naglieri, 2001; Stuebing et al., 2015; Vellutino, Scanlon, & Lyon, 2000). These results are further supported by a lack of relation found for any of the executive, STM, or visuo-spatial factors, and initial reading scores.

Across all three models, our results show that none of our WM, Language or Intelligence factors relate significantly to gains made over the course of a reading intervention, but that many factors were strongly related to initial reading scores. These results have found support in recent literature where Stuebing et al. (2015) have argued that cognitive predictors show very small, if any (1 – 2%), relations to reading gain scores or growth curve models, and that baseline reading scores remain the strongest predictor of

future reading performance (see also Vellutino et al., 2006). In fact, the very incremental added value of cognitive measures relative to baseline reading skills has thus far been replicated by numerous studies (e.g. Burns et al., 2016; Miciak et al., 2016), leading Fletcher and Miciak (2017) to argue for the minimization of comprehensive and lengthy cognitive assessment for characterizing learning disorders such as DD, as well as not using such tests to drive intervention plans. Despite this controversial recommendation, our current results support this assertion and, based on the strong relation between many of our cognitive based tasks and initial reading scores, call into question the very utility of these cognitive tests in their assertion for predicting reading changes (in our case).

1.22 Limitations and Future Directions

Our study had a few limitations of note. First, our sample was limited only to children with DD who typically show lower scores across various reading and language measures used in this study. These limited range of scores thus impact the results of SEM and other statistical tests. Adding a 'group-type' indicator allowing for the inclusion of typically developing children might increase the robustness and clarity of the models. Second, many of the factors included in the models are just identified, while ideally SEM modeling calls for over-identified factors (Kline, 1998). In this case, our study was limited by both sample size, which restricted the number of indicators we could include, as well as time constraints and the strain of testing on young children with DD. This is most notable for our Central Executive factor which could only be used as a phantom factor due to its representation by a single indicator, Counting Span. Third, due to time constraints and practicality considerations, we were unable to test the sample on all types of language measures, thus at least one factor (Spelling) originally theorized in Model 3 (Brown and

Hulme) was without a corresponding indicator and remains untested in the current model. Additionally, given our null findings of relations between model factors and reading gain scores, it might be informative to re-run the models based on response to intervention valence, that is, whether or not cognitive assessments could predict reading change for those who responded well to the reading intervention vs. those who responded poorly.

CONCLUSION

In conclusion, the WM and language indicators from this reading intervention study are best described by the Baddeley and Hitch (1974) theoretically driven model where phonological, visuo-spatial and executive WM are all separate but loosely related factors. Although our language, intelligence and phonological factors shared variance with scores measured prior to reading intervention (initial reading scores), none of these nor other model factors were related to reading changes made over the course of the intervention prompting discussion of the defensibility of using these cognitive tasks as predictors of reading change or future reading performance among struggling readers.

REFERENCES

Acheson, D. J., & MacDonald, M. C. (2009). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological bulletin*, *135*(1), 50.

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.

Anderson, J. R. (1990). Cognitive psychology and its implications. New York: Freeman.

Archibald, L. M., & Gathercole, S. E. (2006). Short‐term and working memory in specific language impairment. International Journal of Language & Communication Disorders, 41(6), 675-693.

Archibald, L. M., & Gathercole, S. E. (2006). Visuospatial immediate memory in specific language impairment. Journal of Speech, Language, and Hearing Research. 49(2), 265 - 277

Bacon, A. M., Parmentier, F. B., & Barr, P. (2013). Visuospatial memory in dyslexia: Evidence for strategic deficits. *Memory*, *21*(2), 189-209.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in cognitive sciences*, *4*(11), 417-423.

Baddeley, A. (2003). Working memory and language: An overview. *Journal of communication disorders*, *36*(3), 189-208.

Baddeley, A. D. (1966a). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly journal of experimental psychology*, *18*(4), 362-365.

Baddeley, A. D. (1966b). The influence of acoustic and semantic similarity on long-term memory for word sequences. *Quarterly Journal of Experimental Psychology*, *18*(4), 302-309.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In *Psychology of learning and motivation* (Vol. 8, pp. 47-89). Academic press.

Becker, J. T., MacAndrew, D. K., & Fiez, J. A. (1999). A comment on the functional localization of the phonological storage subsystem of working memory. *Brain and cognition*, *41*(1), 27-38.

Benson, N. J., Lovett, M. W., & Kroeber, C. L. (1997). Training and transfer-of-learning effects in disabled and normal readers: Evidence of specific deficits. *Journal of Experimental Child Psychology*, *64*(3), 343-366.

Bollen, K. A., & Bauldry, S. (2011). Three Cs in measurement models: Causal indicators, composite indicators, and covariates. *Psychological methods*, *16*(3), 265.

Bourke, L., & Adams, A. M. (2003). The relationship between working memory and early writing assessed at the word, sentence and text level. *Educational and Child Psychology*, *20*(3), 19-36.

Brooks, A. D., Berninger, V. W., & Abbott, R. D. (2011). Letter naming and letter writing reversals in children with dyslexia: Momentary inefficiency in the phonological and orthographic loops of working memory. *Developmental neuropsychology*, *36*(7), 847-868.

Brown, G. D., Hulme, C., & Gathercole, S. E. (1996). Nonword repetition, STM, and word age-of-acquisition: A computational model. *Models of short-term memory*, 129-148.

Burns, M. K. (2016). Effect of cognitive processing assessments and interventions on academic outcomes: Can 200 studies be wrong. *NASP Communiqué*, *44*(5), 1.

Chein, J. M., & Fiez, J. A. (2010). Evaluating models of working memory through the effects of concurrent irrelevant information. *Journal of Experimental Psychology: General*, *139*(1), 117.

Coady, J. A., & Evans, J. L. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairments (SLI). *International Journal of Language & Communication Disorders*, *43*(1), 1-40.

Coolidge, F. L., & Wynn, T. (2005). Working memory, its executive functions, and the emergence of modern thinking. *Cambridge archaeological journal*, *15*(1), 5-26.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in brain research*, *169*, 323-338.

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive psychology*, *51*(1), 42-100.

Danahy, K., Windsor, J., & Kohnert, K. (2007). Counting span and the identification of primary language impairment. *International journal of language & communication disorders*, *42*(3), 349-365.

Dawes, E., Leitão, S., Claessen, M., & Nayton, M. (2015). A profile of working memory ability in poor readers. Australian Psychologist, 50(5), 362-371.

de Jong, C. G., Van De Voorde, S., Roeyers, H., Raymaekers, R., Oosterlaan, J., & Sergeant, J. A. (2009). How distinctive are ADHD and RD? Results of a double dissociation study. Journal of Abnormal Child Psychology, 37(7), 1007-1017.

Dehaene, S. (2009). Reading in the brain: The new science of how we read. Penguin.

Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. Journal of Speech, Language, and Hearing Research, 41(5), 1136-1146.

Dunn, L. M., & Dunn, D. M. (2007). PPVT-4: Peabody picture vocabulary test. Pearson Assessments.

Ehri, L. C. (2014). Orthographic mapping in the acquisition of sight word reading, spelling memory, and vocabulary learning. Scientific Studies of Reading, 18(1), 5-21.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of experimental psychology: General*, *128*(3), 309.

Flanagan, D. P., & Alfonso, V. C. (2017). *Essentials of WISC-V assessment*. John Wiley & Sons.

Fletcher, J. M., & Miciak, J. (2017). Comprehensive cognitive assessments are not necessary for the identification and treatment of learning disabilities. *Archives of Clinical Neuropsychology*, *32*(1), 2-7.

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.

Gathercole, S. (2011). Working memory and learning. *International Journal of Psychology*, 47: 592.

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition*, *23*(1), 83-94.

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of memory and language*, *28*(2), 200-213.

Gathercole, S. E., & Pickering, S. J. (2000). Working memory deficits in children with low achievements in the national curriculum at 7 years of age. *British Journal of Educational Psychology*, *70*(2), 177-194.

Gupta, P., & Tisdale, J. (2009). Does phonological short-term memory causally determine vocabulary learning? Toward a computational resolution of the debate. *Journal of memory and language*, *61*(4), 481-502.

Heitmann, R. R., Asbjørnsen, A., & Helland, T. (2004). Attentional functions in speech fluency disorders. *Logopedics Phoniatrics Vocology*, *29*(3), 119-127.

Henry, L. A. (2010). The episodic buffer in children with intellectual disabilities: An exploratory study. *Research in developmental disabilities*, *31*(6), 1609-1614.

Hoff, E., Core, C., & Bridges, K. (2008). Non-word repetition assesses phonological memory and is related to vocabulary development in 20-to 24-month-olds. *Journal of Child Language*, *35*(4), 903-916.

Jeffries, S., & Everatt, J. (2004). Working memory: Its role in dyslexia and other specific learning difficulties. *Dyslexia*, *10*(3), 196-214.

Kamil, M. L., Borman, G. D., Dole, J., Kral, C. C., Salinger, T., & Torgesen, J. (2008). Improving Adolescent Literacy: Effective Classroom and Intervention Practices. IES Practice Guide. NCEE 2008-4027. *National Center for Education Evaluation and Regional Assistance*.

Kane, M. J., Bleckley, M. K., Conway, A. R., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, *130*(2), 169.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: a latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189.

Kibby, M. Y., Marks, W., Morgan, S., & Long, C. J. (2004). Specific impairment in developmental reading disabilities: A working memory approach. *Journal of learning disabilities*, *37*(4), 349-363.

Kline, R. (1998). Principles and practice of SEM. *New York: The Guilford Press. Koenig, HG, Kvale, JN, & Ferrel, C.(1988). Religion and well-being in later life. The Gerontologist*, *28*, 18-28.

Lovett, M. W., Borden, S. L., DeLuca, T., Lacerenza, L., Benson, N. J., & Brackstone, D. (1994). Treating the core deficits of developmental dyslexia: Evidence of transfer of learning after phonologically-and strategy-based reading training programs. *Developmental psychology*, *30*(6), 805.

Lovett, M. W., Frijters, J. C., 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.

Lovett, M. W., Lacerenza, L., & Borden, S. L. (2000). 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*(5), 458-476.

Lovett, M. W., Lacerenza, L., De Palma, M., & Frijters, J. C. (2012). Evaluating the efficacy of remediation for struggling readers in high school. *Journal of Learning Disabilities*, *45*(2), 151-169.

Lovett, M. W., Steinbach, K. A., & Frijters, J. C. (2000). Remediating the core deficits of developmental reading disability: A double-deficit perspective. *Journal of learning disabilities*, *33*(4), 334-358.

MacDonald, M. C. (2016). Speak, act, remember: The language-production basis of serial order and maintenance in verbal memory. *Current Directions in Psychological Science*, *25*(1), 47-53.

McGrew, K. S., & Woodcock, R. W. (2006). *Woodcock-Johnson III Technical Manual: WJ III*. Riverside Publishing.

Menghini, D., Finzi, A., Carlesimo, G. A., & Vicari, S. (2011). Working memory impairment in children with developmental dyslexia: is it just a phonological deficity? *Developmental neuropsychology*, *36*(2), 199-213.

Miciak, J., Stuebing, K. K., Vaughn, S., Roberts, G., Barth, A. E., & Fletcher, J. M. (2014). Cognitive attributes of adequate and inadequate responders to reading intervention in middle school. *School psychology review*, *43*(4), 407-427.

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.

Naglieri, J. A. (2001). Do ability and reading achievement correlate? *Journal of Learning Disabilities*, *34*(4), 304.

Nicolson, R. I., Fawcett, A. J., & Baddeley, A. D. (1992). Working memory and dyslexia. *Report LRG*, *3*, 91.

O'Shaughnessy, T. E., & Lee Swanson, H. (2000). A comparison of two reading interventions for children with reading disabilities. *Journal of learning disabilities*, *33*(3), 257-277.

Riggall, E., Winter, R., Branum-Martin, L., Sevcik, R., Morris, R.M., Frost, S. (2017) *Relation between Implicit Sequence Learning and Response to Reading.* Posted presented at the Southeastern Psychological Association, Atlanta, GA March 2017

Saygin, Z. M., Norton, E. S., Osher, D. E., Beach, S. D., Cyr, A. B., Ozernov-Palchik, O., ... & Gabrieli, J. D. (2013). Tracking the roots of reading ability: white matter volume and integrity correlate with phonological awareness in prereading and early-reading kindergarten children. *Journal of Neuroscience*, *33*(33), 13251-13258.

Schrank, F. A., McGrew, K. S., Mather, N., Wendling, B. J., & LaForte, E. M. (2014). *Woodcock-Johnson IV tests of achievement*. Riverside Publishing Company.

Shallice, T., & Vallar, G. I. U. S. E. P. P. E. (1990). The impairment of auditory-verbal short-term storage. *Neuropsychological impairments of short-term memory*, 11-53.

Smith-Spark, J. H., & Fisk, J. E. (2007). Working memory functioning in developmental dyslexia. *Memory*, *15*(1), 34-56.

Smith-Spark, J., Fisk, J., Fawcett, A., & Nicolson, R. (2003). Investigating the central executive in adult dyslexics: Evidence from phonological and visuospatial working memory performance. *European Journal of Cognitive Psychology*, *15*(4), 567-587.

Snowling, M. J. (2000). Language and literacy skills: who is at risk and why. *Speech and language impairments in children: Causes, characteristics, intervention and outcome*, 245-259.

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.

Sumner, J. A., Griffith, J. W., & Mineka, S. (2010). Overgeneral autobiographical memory as a predictor of the course of depression: A meta-analysis. *Behaviour research and therapy*, *48*(7), 614-625.

Tolar, T. D., Barth, A. E., Fletcher, J. M., Francis, D. J., & Vaughn, S. (2014). Predicting reading outcomes with progress monitoring slopes among middle grade students. *Learning and Individual Differences*, *30*, 46-57.

Torgesen, J. K. (2005). Recent discoveries from research on remedial interventions for children with dyslexia. *The science of reading*, 521-537.

Torgesen, J. K., Rashotte, C. A., & Wagner, R. K. (1999). *TOWRE: Test of word reading efficiency*. Austin, TX: Pro-ed.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Burgess, S., & Hecht, S. (1997). Contributions of phonological awareness and rapid automatic naming ability to the growth of word-reading skills in second-to fifth-grade children. *Scientific studies of reading*, *1*(2), 161-185.

Torgesen, J. K., Wagner, R. K., Simmons, K., & Laughon, P. (1990). Identifying phonological coding problems in disabled readers: Naming, counting, or span measures? *Learning Disability Quarterly*, *13*(4), 236-243.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of memory and language*, *28*(2), 127-154.

Vallar, G., & Baddeley, A. D. (1984). Phonological short-term store, phonological processing and sentence comprehension: A neuropsychological case study. *Cognitive Neuropsychology*, *1*(2), 121-141.

Varvara, P., Varuzza, C., Padovano Sorrentino, A. C., Vicari, S., & Menghini, D. (2014). Executive functions in developmental dyslexia. *Frontiers in human neuroscience*, *8*, 120.

Vasic, N., Lohr, C., Steinbrink, C., Martin, C., & Wolf, R. C. (2008). Neural correlates of working memory performance in adolescents and young adults with dyslexia. *Neuropsychologia*, *46*(2), 640-648.

Vellutino, F. R., Scanlon, D. M., & Reid Lyon, G. (2000). Differentiating between difficult-to-remediate and readily remediated poor readers: More evidence against the IQ-achievement discrepancy definition of reading disability. *Journal of Learning Disabilities*, *33*(3), 223-238.

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

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2013). *Comprehensive test of phonological processing: CTOPP2*. Pro-ed.

Wechsler, D. (2003). *Wechsler intelligence scale for children-WISC-IV*. Psychological Corporation.

Wechsler, D. (2014). *Wechsler adult intelligence scale–Fourth Edition (WAIS–IV*). San Antonio, Texas: Psychological Corporation.

Wechsler, D. *Wechsler abbreviated scale of intelligence. (1999)*. San Antonio, TX: Psychological Corporation.

White, S., Milne, E., Rosen, S., Hansen, P., Swettenham, J., Frith, U., & Ramus, F. (2006). The role of sensorimotor impairments in dyslexia: A multiple case study of dyslexic children. *Developmental science*, *9*(3), 237-255.

Wigfield, A., Guthrie, J. T., Perencevich, K. C., Taboada, A., Klauda, S. L., McRae, A., & Barbosa, P. (2008). Role of reading engagement in mediating effects of reading comprehension instruction on reading outcomes. *Psychology in the Schools*, *45*(5), 432-445.

Wilson, M. (2001). The case for sensorimotor coding in working memory. *Psychonomic Bulletin & Review*, *8*(1), 44-57.

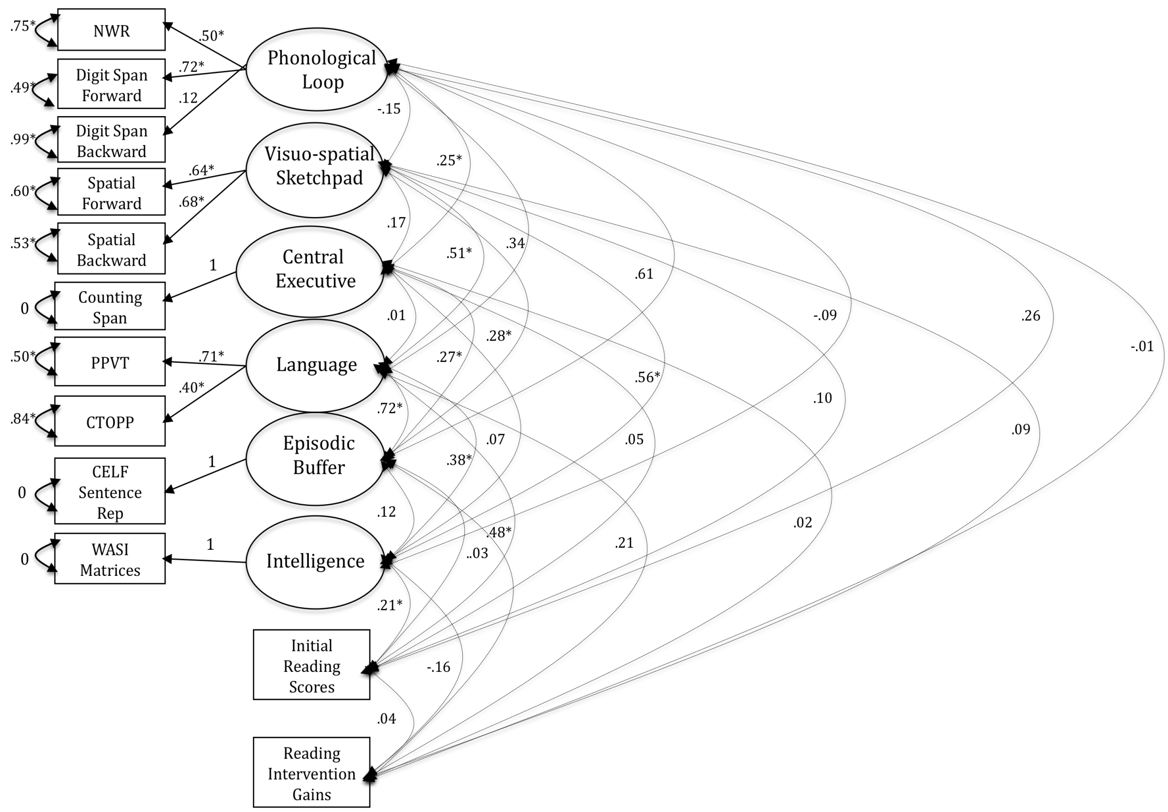
Winter, R., Frijters J. C., Lovett, M. W. & Morris, R. M. (2016) *The Relationship Between Phonological Short Term Memory, Working Memory, and Changes in Reading Scores in Children with Dyslexia.* Poster presented at International Neuropsychological Society, Boston MA February 2016

Winter, R., Riggall, E., Branum-Martin, L., Frijters, J., Morris, R.M. *Working Memory Related to Decoding Gains Across Reading Intervention in Children with Developmental Dyslexia.* Poster presented at the International Neuropsychological Society, Washington, DC February 2018.

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

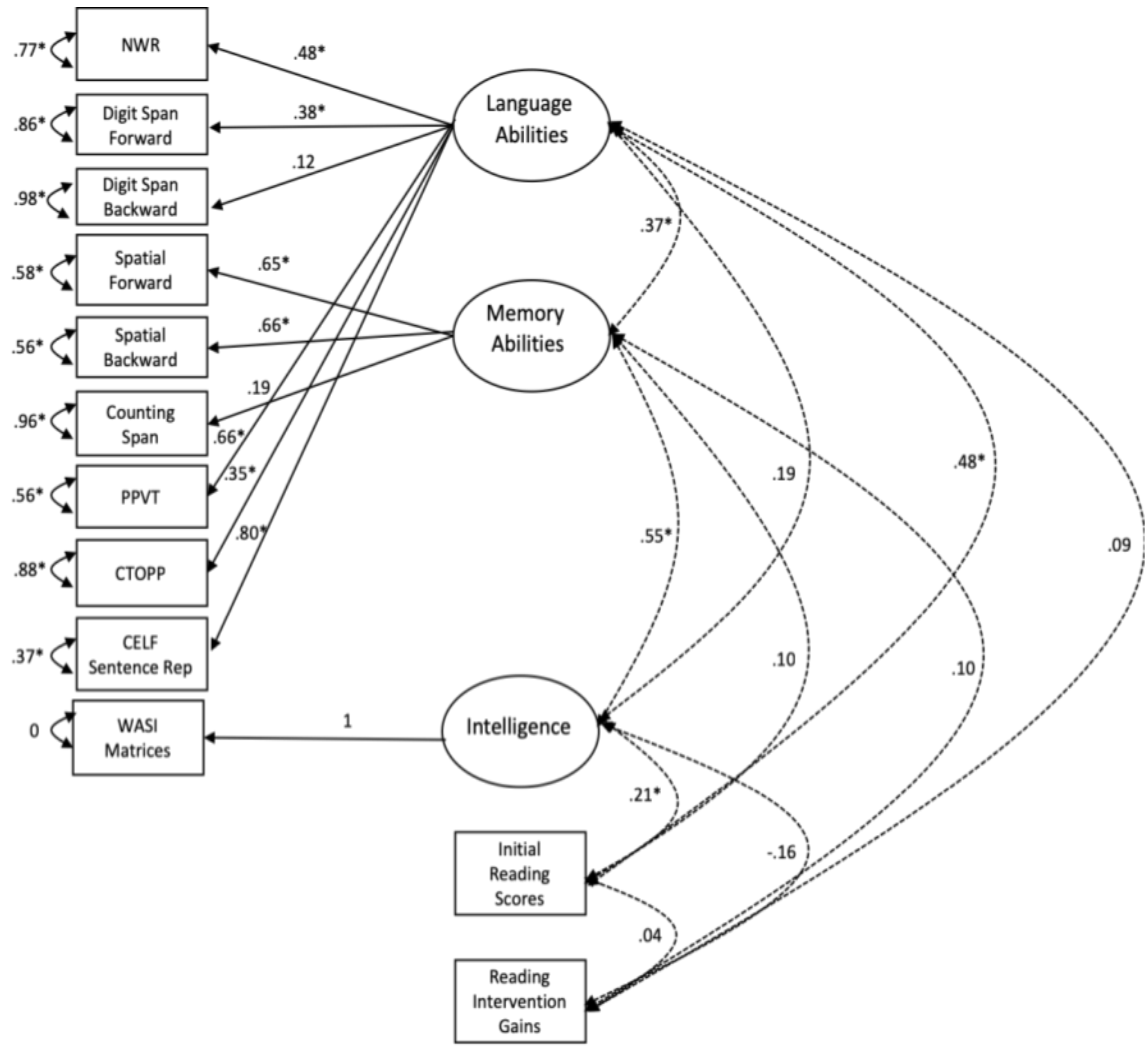
APPENDICES

APPENDIX A: MODEL 1 (BADDELEY AND HITCH) WITH EPISODIC BUFFER



Here a single indicator represents the Episodic Buffer: CELF Sentence Repetition. The overall model yielded a fair fit: Chi-square 37.79, df = 31, p=0.19; RMSE: 0.04; CFI: 0.96; SRMR: 0.05, similar to that seen in the original Model 1 (Figure 12). Since the original Model 1 is more elegant and parsimonious (i.e., the results are similar with fewer factors) it is the primary model discussed in the Discussion.

APPENDIX B: ALTERNATIVE MODEL 3 (BROWN AND HULME)



Here, both Digit Span tasks are subsumed by the Language Abilities factor. The overall model yields similarly poor fit results to the originally tested model: Chi-square 77.96, df = 47, <0.05; RMSE: 0.08; CFI: 0.83; SRMR: 0.08. Additionally, the indicators seem to fit similarly, with the exception of Digit Span Forward showing significant relation to the Language Abilities factor in this model (.38) and almost no relation (-.01) to the Memory Abilities Factor in the original model.