A Multifactorial Model of Risk for Dyslexia

Gabrielle O'Brien

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

Adrian K.C. Lee, Chair

Jason D. Yeatman

Richard Wright

Program Authorized to Offer Degree:
Speech and Hearing Sciences

University of Washington

Abstract

A Multifactorial Model of Risk for Dyslexia

Gabrielle O'Brien

Chair of the Supervisory Committee:
Professor Adrian K.C. Lee
Speech and Hearing Sciences

There are many competing theories about the causes of dyslexia, a learning disability in which an individual without apparent neurological or circumstantial barriers struggles to become a fluent reader relative to their peers. Many modern hypotheses about dyslexia promote the idea of a "core deficit" originating either in sensory or phonological processes. This dissertation describes four psychophysical studies, both visual and auditory, of children ages 8-12 ($n = 205$) designed to investigate the claims of various hypotheses about reading disability. We find that indices of categorical behavior on a phoneme labeling task, as well as summary measures of performance on a random dot motion discrimination task, are correlated with reading skill. However, effect sizes and patterns of correlations between multiple predictors of reading skill do not support the framework of a core deficit in either sensory or phonological processes. Instead, results are consistent with an additive risk factor model, in which several distinct causal pathways are required to explain variance in reading skill at the population level.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This document exists thanks to a number of people.

I would like to thank my mentors at all stages of my graduate work: Jason Yeatman, KC Lee, Richard Wright and Jay Rubinstein. Also thanks to the faculty members whose efforts make the auditory science community at the University of Washington excellent: David Perkel, Ed Rubel, Ione Fine, and many others.

I've benefited many times over from the support of friends who have become like family: Wendy Parkinson, Kelly Jahn, Lindsey DeVries, Shae DiNino and Liesbeth Gijbels. My labmates Emily Kubota and Patrick Donnelly were incredible guides and compatriots at various stages of my dissertation work.

Finally, I thank my parents and grandparents, my partner Jamie, and the Brew family for believing in my endeavors.

DEDICATION

for Bruno

Chapter 1

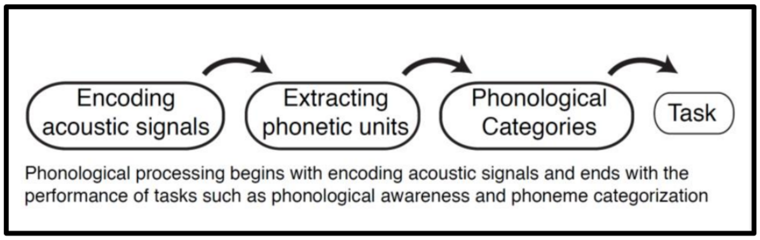
INTRODUCTION

There are many conflicting theories of dyslexia, a learning disorder that specifically affects reading in 5-17% of the population and cannot be explained by cognitive, sensory, or circumstantial factors (Shaywitz, 1998; Snowling, 2000). It is broadly, but not universally (Ramus and Szenkovits, 2008; Pennington et al., 2012), held that dyslexia that dyslexia is characterized by impaired phonological processing, which includes *understanding that speech can be decomposed into phonemes* and *efficiently manipulating, remembering, and accessing phonemic information* (Wagner and Torgesen, 1987; Snowling, 1998). However, it is unclear to what extent the maturation of these skills is a cause or consequence of literacy training or why individuals with dyslexia struggle to master phonological awareness, memory, and automaticity. In a general framework, we might imagine that the first disruption in the phonological processing could occur at one of three stages in the auditory pathway: at the level of extracting relevant information for speech perception from incoming acoustic signals, at the phonemic level (mapping symbols and sounds to discrete phonemes), or at the level of general cognitive processes such as working memory and attention (Serniclaes et al., 2004). This framework is schematized in Figure 1.1

In the past decades, several theories have been presented that posit a core deficit of dyslexia occurring somewhere in the auditory pathway. Typically, the core deficit is framed as a measurable impairment in some aspect of auditory, linguistic, or cognitive processing that is sufficient to cause dyslexia and accounts for the majority of cases. The purpose of this introduction is to briefly survey the evidence supporting the various core deficit theories and to discuss the strengths and weaknesses of these arguments.

It is useful to consider a list of desirable properties that a core deficit model of dyslexia

Figure 1.1: A schematic of several stages at which phonological processing might by disrupted.



should have:

- The deficit must be **causally related to dyslexia**. An intervention that remedies the deficit in an individual should measurably improve their reading skill.

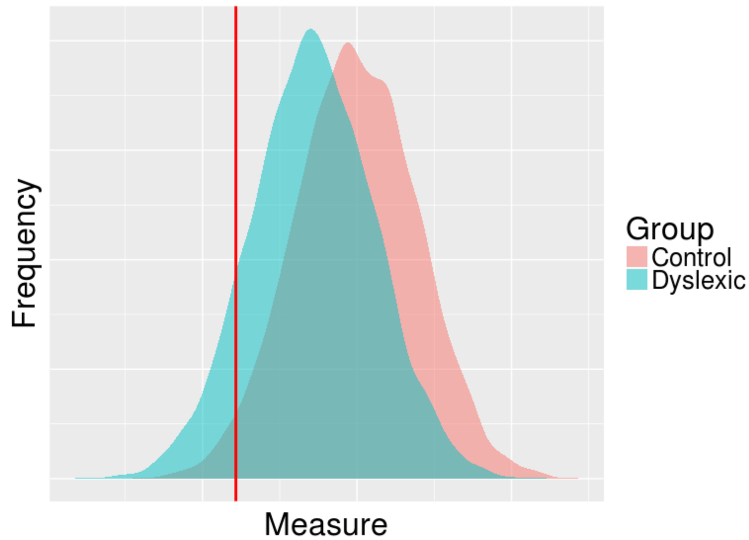
- The deficit must have **explanatory power as a sensitive and specific mechanism**. In other words, the proposed mechanism should be able to account for the observed impairments in phonological processing, as well as observations from behavioral auditory measures, without simultaneously predicting impairments that individuals with dyslexia do not have.

- To be considered a core deficit, the degree of impairment should **correlate with the degree of reading difficulty**. In other words, in a group of individuals with dyslexia, individuals with the greatest impairment should also have the most disordered reading (in agreement with Rosen (2003)).

- **Most individuals with dyslexia should possess the deficit.** Although clinical populations often yield noisy data, we should expect that any impairment warranting the "core deficit" label is broadly applicable. If a deficit does not occur in most individuals with dyslexia, then it should not be framed as the core deficit of dyslexia.

- Most individuals with dyslexia should **fall outside the normal range observed in**

Figure 1.2: Simulated distributions of a measure in control and dyslexic groups.



In this example, the average value of the measure differs by group, but most measures from the dyslexic group fall within the control distribution. The red line indicates the lower bound of the 95% confidence interval on the Control group. Above, an effect size of 0.7 is shown (Cohen's d). Assuming the measure is reasonably reliable, this pattern may indicate a risk factor or symptom of dyslexia. However, it would not explain many cases of typical and disordered reading.

individuals with typical reading ability. For example, if individuals with dyslexia perform worse on a given measure but fall overwhelmingly within the 95% confidence interval of the non-dyslexic population, then that measure likely does not reflect a deficit that in isolation causes dyslexia (Ramus, 2003)—assuming the measure is not dominated by noise. Otherwise, any individual with that deficit would have dyslexia. This pattern may indicate a vulnerability that interacts or adds with other factors to influence reading ability (see Figure 1.2). But a core deficit should not be present in a substantial proportion of the greater population.

In practice, some of these criteria may be difficult to meet. For example, because a sample of dyslexics is heterogenous in terms of the remedial treatment participants may have received before the study, and because children are unique population with limited capacity for psychophysics, correlations with any measure in a cohort of young dyslexics may be especially noisy. However, we should aspire to hold a high standard of evidence because of the immense effort and financial resources children, families, and teachers will invest in treatments that may appear backed by science.

We begin with a brief review of several prominent theories of dyslexia. Then, we identify several outstanding questions that this dissertation will address.

1.1 A review of theories of dyslexia

The field of dyslexia research is more than a century old, and it is not possible to discuss every theory that influences modern thinking about reading disability. The following hypotheses represent some of the most-cited and frequently discussed ideas today. Note that because this dissertation is primarily considered with auditory aspects of reading development, we do not give equal space to predominantly visual mechanisms here.

1.1.1 The rapid temporal processing hypothesis

In 1980, Paula Tallal demonstrated that dyslexics as a group performed worse on a temporal order judgment task with short interstimulus intervals (ISIs) but not long ISIs. She concluded that temporal auditory processing is a unique challenge for dyslexics. The proposed consequence of this alleged deficit would be in processing rapid aspects of speech, included formant transitions and voice onset times. By extension, children would struggle to develop a clear internal model of the sounds of their language, which would hinder their ability to learn the mappings between letters and phonemes. This hypothesis was incredibly influential in driving attention to auditory processing in dyslexics, and the basic tenet—that subtle auditory processing deficits could jeopardize the development of phonological awareness—still proliferates in many of the auditory theories of dyslexia that have followed Tallal's.

Although this hypothesis has been the subject of immense study and discussion for the last few decades, a great deal of evidence has been aggregated against the conclusion that rapid temporal processing deficits are at the core of dyslexia (see reviews: Farmer and Klein, 1995; Rosen, 2003). In brief, four separate studies have clarified that Tallal's initial finding can be explained by control participants reaching ceiling performance in one task condition (Reed, 1989; Nittrouer, 1999; Marshall et al., 2001; Waber et al., 2001), implying that poor readers do not perform as controls even at long ISIs. Furthermore, there is now abundant evidence that individuals with dyslexia perform (on average) poorly on many auditory tasks that do not involve the processing of brief sounds (Hämäläinen et al., 2013; Rosen, 2003; Ramus, 2003; Amitay et al., 2002), including slow ($\sim$ 2 Hz) amplitude and frequency modulation detection (Witton et al., 1998; Lorenzi et al., 2000; Stuart et al., 2006). Additionally, the intervention program created on the basis of the rapid temporal processing hypothesis, called Fast ForWord, has not shown replicable benefits in randomized control trials conducted by investigators unaffiliated with the product (see Strong et al. (2011) for review). In sum, there appears to be a criticial mass of literature demonstrating the untenability of the rapid temporal processing hypothesis as originally stated, both in terms of the specificity of the mechanism and its clinical value as a target for remediation.

There are two reasons why the rapid temporal processing hypothesis is still pertinent to today's research. First, it popularized the concept that a subtle perceptual deficit could act as a "roadblock" for children learning to map the sounds of their language to letters. In this view, children with a fuzzy percept of formant transitions—such as the cue that differentiates /ba/ from /da/—might struggle to learn how those ambiguous categories relate to discrete letters. While it is difficult to establish whether this proposed causal pathway is correct, the concept is still frequently referenced in discussions of potential auditory mechanisms of reading disability (Goswami, 2015; Calcus et al., 2016; Ziegler et al., 2009).

The second reason is that the rapid temporal processing hypothesis, in some form, is still discussed as a candidate mechanism of dyslexia today. Some researchers claim that rapid temporal processing difficulties occur as the consequence of a more general disruption in neural development (such as entrainment to speech envelope (Goswami, 2015) or abnormal noise in cortical neurons (Hancock et al., 2017)). Others propose that in spite of the conflicting findings in literature, a temporal processing explanation for dyslexia is still viable

(Vandermosten et al., 2010, 2011). Some proponents of the magnocellular theory of dyslexia, which hypothesizes that abnormal visual motion processing in the magnocellular pathway is a key feature of the disorder, contend that a similar fundamental mechanism impairs rapid temporal processing in both the auditory and visual domains (Stein, 2018; Casini et al., 2018; Witton et al., 2002). It is unclear whether these renewed theories will be able to answer any of the original criticisms of the rapid temporal processing hypothesis.

1.1.2 The temporal sampling hypothesis

A recent flavor of the temporal processing hypothesis is a shift towards features at the syllabic scale, called the "temporal sampling hypothesis" (Goswami, 2011). This is a relatively recent hypothesis inspired by the temporal sampling framework of audition posed by Poeppel (Poeppel, 2003; Giraud and Poeppel, 2012). In Poeppel's model, neural oscillations sample incoming auditory information. Neural oscillations in the 4-10 Hz frequency band (called Theta waves) are thought to be important for coding syllabic information because these endogenous oscillations have been observed to phase lock to syllables in speech. This process was hypothesized to be disrupted in dyslexics by Goswami (2011).

The hypothesis of the temporal sampling framework is that dyslexics abnormally integrate phonetic features with syllabic-rate information, and this somehow leads to irregularly specified phonemes that are difficult to map to graphemes. The mechanisms of integration across temporal timescales to construct auditory objects are still largely unknown, so the proposed pathway from abnormal entrainment to syllabic-level cues in speech to disordered phonemes remains vague.

The sampling theory predicts in particular that psychophysical measures of rise time processing, as well as amplitude modulation discrimination and detection at low modulation rates, should be impaired in individuals with dyslexia. A recent review of auditory psychophysical studies estimated that experiments involving some aspect of rise time perception were associated with an average effect size of 0.8 (Hämäläinen et al., 2013). While this meta-analysis suggests that reliable group differences exist, it also implies that Control and Dyslexic groups would be more than 50% overlapping in a typical study. Said another way, only a minority of individuals with dyslexia could be considered to have disordered rise time processing relative to the general population: setting a liberal threshold of "disordered processing" at 1 standard deviation below the mean in the Control population, 42.1% of individuals with dyslexia (and by definition, 15.8% of individuals without dyslexia) could be considered disordered. Setting a more stringent cutoff at the bottom 5% of the Control population (or 1.96 standard deviations), only 12.3% of dyslexia could be considered disordered. Thus, the literature does not appear to support the view that most individuals with dyslexia perform especially poorly on psychophysical experiments intended to target syllable-rate rise time detection.

In all, the temporal sampling theory has led to an intriguing body of work showing that rise time impairments occur in some dyslexics as measured by a set of psychophysical tasks. Based on estimates from the current literature, this deficit likely affects the minority of dyslexics. There is only limited evidence that individuals with dyslexia having special difficulty with slow amplitude modulation rates (Amitay et al., 2002; Lorenzi et al., 2000; Stuart et al., 2006; Witton et al., 2002) and contradictory evidence from studies of speech in noise processing suggesting slow modulation sensitivity is intact (Ziegler et al., 2009; Dole et al., 2012; Calcus et al., 2016). Researchers have also raised concerns that the temporal sampling theory fails to predict many of the experimental conditions in which individuals with dyslexia perform like their typically developing peers (Ramus and Ahissar, 2012). While this theory continues to attract much discussion and makes relatively strong claims about the centrality of rise-time processing impairments to dyslexia, there is presently little evidence in aggregate that a consistent profile of auditory processing difficulties occur across disabled readers.

1.1.3 Deficits in speech robustness

Yet another auditory framework of dyslexia proposes that the core deficit is not auditory sensitivity, but difficulty extracting informative cues from ethological signals due to a lack of robustness to distortion—for example, speech in the presence of noise (Ziegler et al., 2009). This theory is appealing because it purports to explain why group differences in measures of speech perception tend to be small: most psychophysics are performed in ideal listening conditions. As such, true impairments may only be detectable in degraded listening conditions.

In the literature so far, effect sizes between 0 and 1 are typical (Messaoud-Galusi et al., 2011; Hazan et al., 2009; Boets et al., 2007; Dole et al., 2012; Poelmans et al., 2011; Calcus et al., 2016; Ziegler et al., 2009); thus the speech robustness hypothesis faces the same challenge as other auditory hypotheses: the majority of individuals with dyslexia appear to perform in the typical range.

Another challenge to the speech robustness hypothesis comes from its ability to predict individual differences in other aspects of auditory processing. Recently, Calcus et al. (2016) addressed the question of whether speech in noise perception predicts the ability to categorically label speech sounds among struggling readers. If individuals with dyslexia with speech in noise deficits are particularly unable to extract phonetic information from complex acoustic signals, then these individuals would be expected to have shallower phoneme identification functions. The authors did not find a significant association between categorical labeling and speech in noise perception. In fact, the authors observed that across several different noise conditions, individual performance among disabled readers was highly variable. This observation was not consistent with a uniform mechanism of impairment. Another study by Calcus et al. (2018) also found that individual profiles of performance in several noise conditions were inconsistent, seeming to rule out the idea that a general deficit in utilizing degraded acoustic information is universal in dyslexia. These experiments have substantially challenged the idea that speech in noise deficits are predictive of reading skill on an individual level, even though there may be reliable group-level differences.

1.1.4 Deficits in categorical perception of speech

A great deal of literature has focused on the "categorical perception[1] deficit" in individuals with dyslexia; in other words, that groups of disabled readers tend to show shallower identification functions on phoneme labeling tasks as well as more variable discrimination functions (Manis et al., 1997; Mody et al., 1997; Joanisse et al., 2000; Blomert and Mitterer, 2004; Hakvoort et al., 2016; Serniclaes et al., 2004). Some researchers theorize that this impairment is the core deficit of dyslexia (Serniclaes et al., 2001, 2004; Bogliotti et al., 2008; Noordenbos and Serniclaes, 2015), and many other theories of dyslexia emphasize the role of categorical perception in phonological development (Chiappe et al., 2001; Zhang et al., 2012; Robertson et al., 2009; Cheung et al., 2009; Manis et al., 1997; de Gelder and Vroomen, 1998; Paul et al., 2006). For example, to test the hypothesis that individuals with dyslexia are specifically impaired when processing brief dynamic cues in speech, Vandermosten et al. (2010; 2011) tested categorical labeling of speech tokens containing dynamic and static phonetic cues. Indeed, most arguments for a low-level acoustic encoding deficit contend that compromised encoding leads to "fuzzy" representations of phonemes in the brain.

The primary source of evidence that phonemes are "fuzzy" comes from studies of identification and discrimination task performance. As with many other deficits reviewed so far, there does seem to be evidence for a reliable group difference between a sample of dyslexics and a sample of typically developing controls on these measures: the 2015 Noordenbos and Serniclaes meta-analysis of 34 identification and 13 discrimination studies calculated average effect sizes of 0.66 and 0.86 respectively. Although these effect sizes are clearly meaningful, they imply moderately overlapping distributions of dyslexic and typical readers.

---

[1] It is not clear in much of the literature whether *categorical perception* refers to the classical model of Liberman and colleagues (Liberman, 1970; Liberman and Mattingly, 1985), in which discrimination of speech sounds is proposed to be entirely determined by an individual's identification function, or if the term implies only a tendency to label phonemes consistently on phoneme identification tasks. Pisoni et al. (1974; 1982) demonstrated that an individual can produce sharp psychometric functions on a phoneme identification task without having categorical perception in the classical sense, but this distinction between sharp psychometric functions and true categorical perception is not explicitly made in the literature I've cited above. Thus, different research groups may not entirely agree what phoneme identification tasks are measuring.

There are several important considerations in interpreting studies of categorical labeling and discrimination. First, the notion that categorical labeling slope is closely related to phonological awareness—the ability to identify and segment phonemes in a word—has been challenged by several recent studies that have found only small (and in some cases, non-significant) mediating effects of phonological awareness on the relationship between indices of categorical labeling and reading skill (Hakvoort et al., 2016; O'Brien et al., 2018; Snowling et al., 2019). Although further study is warranted, the close agreement of these results should elicit a deeper look into the purported relationship between phonological awareness and categorical labeling.

Second, identification functions are known to change shape throughout development (Nittrouer, 1992). They are likely closely linked to age and cognitive maturation. However, because they are typically measured in literate adults and children undergoing literacy training, it is difficult to disentangle the extent to which they are changed by phonological awareness. Serniclaes has argued that reliable categorical labeling is not a consequence of reading ability, because a group of French illiterate adults showed mostly normal performance on a /ba/∼/da/ labeling task (Serniclaes et al., 2005). Further study about the causal direction of the relationship between categorical labeling and reading skill is warranted. However, the variance explained among struggling readers by their categorical labeling performance is already relatively small; shrinking our estimate of the degree to which categorical labeling limits reading ability may well bring the power of this marker to explain reading skill below clinically relevant levels. Indeed, several authors maintain that categorical labeling and discrimination already explain too little variance to be useful (Rosen and Mangari, 2001; Hazan et al., 2009; Messaoud-Galusi et al., 2011).

A theoretical question that must be addressed is whether non-categorical perception is actually harmful to reading acquisition. At the core of the categorical perception deficit hypothesis is the notion that dyslexics struggle because they are deluged with irrelevant acoustic information, instead of representing sounds in the simpler code of the phonemes of their language (Serniclaes et al., 2004). This is reminiscent of classical arguments about categorical perception that suggest a downsampling process is employed by speech-specific neural circuits, converting a rich acoustic signal into phonemes and discarding phonetic information. Although not all discussions of fuzzy phonemes in the dyslexia literature make such claims explicitly, they must lead to similar conclusions implicitly if they accept that a categorical perception deficit is a key feature of dyslexia.

Yet, sub-phonemic variability is known to assist, rather than interfere, with lexical acceses in many contexts (McLennan et al., 2003; Connine, 2004; Whalen, 1991; Marslen-Wilson and Warren, 1994; McQueen et al., 1999; Dahan et al., 2001). Not only are typically developing listeners sensitive to within-category variation of speech sounds, they in fact benefit from such information in the context of word recognition. Furthermore, there is evidence that typical listeners likely do not categorize phonemes in the process of recognizing words, but instead maintain probabilistic representations of several candidate phonemes on the basis of incoming phonetic information (Andruski et al., 1994; McMurray et al., 2009). While it is often taken as a natural starting point in dyslexia research, the notion that typical listeners rely on a primarily phonemic code during ethological listening conditions is not a consensus among speech scientists (Cleary and Pisoni, 2001; Port and Leary, 2005; Port, 2007; Holt and Lotto, 2010).

It will be a challenge for future dyslexia researchers to discern how categorical labeling of phonemes in an experimental setting relates to phonological awareness and reading skill at an individual level. Currently, it is clear that less categorical labeling of phonemes and reduced phonological awareness are associated with dyslexia, but not that these two impairments always co-occur or can predict reading skill at an individual level.

1.1.5 The statistical learning hypothesis

Several theories of deficits that are not specifically auditory or linguistic in nature have emerged in response to the inconsistent evidence for lower-level processing disorders. In this framework, sensory processing is intact and phonemes are represented normally in the brains of individuals with dyslexia, but cognitive processes such as working memory and attention are disordered.

The statistical learning (or anchoring) hypothesis, first posed by Ahissar (2006; 2007), posits that individuals with dyslexia fail to benefit from stimulus repetitions, whereas typically developing individuals are able to form “perceptual anchors” around repeated stimuli. This mechanism was suggested following a comprehensive study in which disabled readers participated in a battery of auditory tasks designed to probe low- and high-level auditory mechanisms (Ahissar, 2007).

A strength of the anchoring hypothesis is that it offers an appealing explanation for the oft-repeated result that individuals with dyslexia are more likely to perform abnormally on auditory psychophysical tasks, but do not have consistent profiles of impairment. For example, in the speech in noise literature, Dyslexic groups were more likely to do poorly on speech-in-noise tasks, but few individuals performed consistently poorly when presented with a battery of such tasks (Calcus et al., 2016, 2018). Whether or not the specific mechanism of perceptual anchoring can explain such findings remains to be decided.

There has been immense interest in whether individuals with dyslexia have general difficulties with statistical learning (i.e., taking advantage of the statistics of their environment). Recent studies have shown that individuals with dyslexia, on average, appear to have some difficulties with statistical learning in multiple contexts: learning the statistics of noise (Agus et al., 2013), transitional probabilities in streams of syllables and pure tones (Jaffe-Dax et al., 2016), and novel probabilistic categories (Gabay et al., 2015). However, there are many studies that fail to find such differences (Gabay and Holt, 2018; Staels and Van den Broeck, 2015; Gould and Glencross, 1990; Inácio et al., 2018; Jiménez-Fernández et al., 2011; Samara and Caravolas, 2017; Du and Kelly, 2013). The authors of a 2018 study about category learning (Gabay and Holt, 2018) concluded that the current results underscore that it is too strong to conclude that individuals with dyslexia have a general impairment in processing probabilistic information.

Further study is required to determine whether this hypothesis can explain the inconsistent psychophysical literature on dyslexia. Still, it is not clear how the proposed mechanism (statistical learning) will change our understanding of dyslexia as a disorder. This framework has the power to explain why some disabled readers perform poorly on certain psychophysical tasks, but many (and possibly, most) dyslexics do not particularly struggle with these tasks. For example, although anchoring has been proposed to explain deficits in categorical labeling, recall the meta-analysis of 34 phoneme identification studies that predicted an average effect size of 0.66 meaning, only 9.7% of dyslexics fall below the 95% confidence interval of the typical population (Noordenbos and Serniclaes, 2015). Moreover, while proponents of the statistical learning hypothesis argue that the mechanism can explain reduced phonological awareness, there are currently no studies (that I know of) directly assessing the relationship between such measures.

In conclusion, the statistical learning hypothesis poses a fascinating avenue for new research into statistical learning in dyslexia that could, with further refinement and exploration, explain some or most of the perplexing patterns from basic auditory psychophysics. It is not clear that most individuals with dyslexia have a domain-general statistical learning impairment, and future iterations of this hypothesis will have to be sufficiently nuanced to account for contexts where disabled readers appear to have normal sensitivity to the statistics of acoustic stimuli. At present, it is premature to say whether or not this hypothesis explains any of the phonological processing difficulties individuals with dyslexia face. However, considering the substantial number of individuals with dyslexia who do not appear to be candidates for a statistical learning deficit, there is still room for questions about whether this mechanism could be a core or causal attribute of dyslexia.

1.2 Outline of the dissertation

This dissertation addresses the role of categorical labeling, sensory processing, and domain-general mechanisms such as attention and statistical learning of speech sounds in literacy acquisition. These questions are of practical significance pertinent to our understanding of reading disorders and their treatment, but are also relevant to theories of speech perception and reading disability in general.

Four experiments are presented in this dissertation, involving a total of 205 unique participants[2] between the ages of 8 and 12:

- **Clarifying whether individuals with dyslexia are specifically impaired at categorizing speech sounds that differ on the basis of spectrotemporal modulations.** This work responds to claims that individuals with dyslexia are impaired at categorizing speech sounds that differ on the basis of a spectrotemporal cue, but not a steady-state cue.

- **Investigating whether stimulus duration affects how disabled readers categorize speech sounds.** This study examines whether longer exposure to speech cues enables children with dyslexia to behave more categorically on a standard phoneme labeling task.

- **Characterizing the relationship between sensory, cognitive, and phonological predictors of dyslexia.** This study uses a mathematical model of sensory decision making, the drift diffusion model, to model how children of various reading levels perform on a random dot motion discrimination task. Statistical methods are used to explore the correlational structure of predictors to elucidate the validity of a "core deficit" model of dyslexia.

- **Investigating the contributions of statistical learning, stimulus adaptation, and fatigue to categorical labeling.** This study is designed to gauge the influence of task features on how disabled and strong readers label phonemes.

[2]Note that some participants are involved in more than one study.

Chapter 2

READING ABILITY AND PHONEME CATEGORIZATION

This chapter is published as O'Brien et al. (2018) with co-authors Daniel McCloy and Jason Yeatman.

Dyslexia is associated with abnormal performance on many auditory psychophysics tasks, particularly those involving the categorization of speech sounds. However, it is debated whether those apparent auditory deficits arise from (a) reduced sensitivity to particular acoustic cues, (b) the difficulty of experimental tasks, or (c) unmodeled lapses of attention. Here we investigate the relationship between phoneme categorization and reading ability, with special attention to the nature of the cue encoding the phoneme contrast (static versus dynamic), differences in task paradigm difficulty, and methodological details of psychometric model fitting. We find a robust relationship between reading ability and categorization performance, show that task difficulty cannot fully explain that relationship, and provide evidence that the deficit is not restricted to dynamic cue contrasts, contrary to prior reports. Finally, we demonstrate that improved modeling of behavioral responses suggests that performance does differ between children with dyslexia and typical readers, but that the difference may be smaller than previously reported.

2.1 Introduction

Dyslexia is a learning disability that affects between 5% and 17% of the population and poses a substantial economic and psychological burden for those affected (Lyon et al., 2003; Shaywitz, 1998; Snowling, 2000). Despite decades of research, it remains unclear why so many children without obvious intellectual, sensory, or circumstantial challenges find written word recognition so difficult.

One popular and persistent theory is that dyslexia arises as a result of an underlying auditory processing deficit (Farmer and Klein, 1995; Goswami, 2011; Van Ingelghem et al., 2005; Snowling, 1998; Steinbrink et al., 2014; Tallal, 1980). According to this theory, a low-level auditory processing deficit disrupts the formation of a child's internal model of speech sounds (phonemes) during early language learning; later, when young learners attempt to associate written letters (graphemes) with phonemes, they struggle because their internal representations of phonemes is compromised (Poelmans et al., 2011).

In line with this hypothesis, many studies report group differences between individuals with dyslexia and typical reading control participants in auditory psychophysical tasks including amplitude modulation detection (Hämäläinen et al., 2013; McAnally and Stein, 1997; Menell et al., 1999; Rocheron et al., 2002; Witton et al., 1998), frequency modulation detection (Boets et al., 2007; Dawes et al., 2009; Gibson et al., 2006; Stoodley et al., 2006; Witton et al., 2002), rise time discrimination, and duration discrimination (Banai and Ahissar, 2004, 2006; Goswami, 2011; Thomson et al., 2006; Thomson and Goswami, 2008). Moreover, attributing dyslexia to an auditory deficit is appealing because one of the most effective predictors of reading difficulty is poor phonemic awareness—the ability to identify, segment and manipulate the phonemes within a spoken word (Bus and van IJzendoorn, 1999; Hulme et al., 2002). If a child's phoneme representation were abnormal due to auditory processing deficits, it could be a common factor underlying both poor performance in auditory phoneme awareness tests and difficulties learning to decode written words.

In one variant of this auditory hypothesis, dyslexia is thought to involve a deficit specifically in the processing of rapid modulations in sound, usually referred to as a "rapid temporal processing" deficit (Merzenich et al., 1996; Tallal, 1980; Tallal et al., 1996). This idea has been controversial (see Rosen (2003) for review), but also remains as one of the widely cited accounts of auditory deficits in dyslexia. One source of controversy is that the rapid temporal processing deficit is far from universal; in Tallal's 1980 study that first proposed a causal relationship between rapid temporal processing and reading ability, only 8 out of 20 dyslexic children showed a deficit on a temporal order judgment task. Similarly, in a more comprehensive study of 17 dyslexic adults, only 11 had impaired performance on an extensive battery of auditory psychophysical tasks (Ramus, 2003). Moreover, in that study, tasks requiring temporal cue sensitivity were not systematically more effective than non-temporal tasks at separating dyslexics from controls. Several studies have also shown that dyslexic participants have heightened sensitivity to allophonic speech contrasts (differences in prounciation that do not change the identity of the speech sound) (Bogliotti et al., 2008; Noordenbos and Serniclaes, 2015; Serniclaes et al., 2004), including contrasts primarily marked by temporal differences in the stimuli. This contradicts the hypothesis that the dyslexic brain lacks access to fine-grained temporal information. Furthermore, other studies have shown that individuals with dyslexia have normal abilities to resolve spectrotemporal modulations in noise, suggesting that apparent speech perception difficulties result from higher-level aspects of processing beyond encoding sounds. (Calcus et al., 2018; Dole et al., 2012; Ziegler et al., 2009) More recently, the rapid temporal processing hypothesis has been reframed as a deficit in processing dynamic, but not necessarily rapid, aspects of speech (Boets et al., 2011; Law et al., 2014; Poelmans et al., 2011).

On the whole, there is ample evidence that abnormal performance on auditory psychophysical tasks is frequently associated with dyslexia (Hämäläinen et al., 2013), but much less evidence that a specific deficit in temporal processing is causally related to dyslexia. But regardless of whether the alleged deficit involves all modulations of the speech signal or only rapid ones, there are at least two major reasons why the theory that auditory temporal processing deficits cause the reading difficulties seen in dyslexia has been called into question. First, although correlations have been found between certain psychoacoustic measures and reading skills (Goswami et al., 2002; Vandermosten et al., 2010; Witton et al., 1998; Talcott et al., 2000), there have been no direct findings relating performance on psychoacoustic tasks to phonological representations (Rosen, 2003). On the contrary, there is growing evidence that poor performance on psychophysical tasks may be partially or totally explained by the demands of the test conditions and not the stimuli themselves (Banai and Ahissar, 2006). In this view, it is not fuzzy phonological representations, but rather poor working memory—which often co-occurs with dyslexia (Siegel and Ryan, 1989; Swanson, 1993; Vargo et al., 1995; Wang and Gathercole, 2013)—that drives deficits in both psychoacoustic task performance and in reading ability (Amitay et al., 2002; Banai and Ahissar, 2004). Proponents of this hypothesis argue that the phonological deficit associated with dyslexia is only observed when considerable memory and time constraints are imposed by the task; evidence for this view comes from studies in which listeners perform different kinds of discrimination paradigms on the same auditory stimuli (Ahissar, 2007; Ahissar et al., 2006; Banai and Ahissar, 2006; Ziegler, 2008). A neuroimaging study indicating that individuals with dyslexia have similar neural representations of speech sounds as controls, but differences in connectivity between speech processing centers, also lends support to the idea of "intact but less accessible" representations of phonetic features (Boets et al., 2013). Additionally, many individuals with dyslexia perform normally on speech perception tasks when provided ideal listening conditions but exhibit a deficit when speech is presented in noise. This finding has been interpreted as indicating a deficit in higher-level auditory processing such as the ability to attend to perceptually relevant cues and ignore irrelevant aspects of the signal (Calcus et al., 2018; Dole et al., 2012; Ziegler, 2008). Dyslexia is also known to be comorbid with ADHD, and numerous reviews have suggested that, in addition to working memory, differences in attention may be another key driver of the observed group differences (Ramus, 2003; Roach et al., 2004). However, it is not clear how this hypothesis accounts for several studies where dyslexics actually perform better than controls discriminating fine-grained acoustic cues (Bogliotti et al., 2008; Calcus et al., 2018; Goswami, 2011).

A second reason to question the link between auditory processing deficits and dyslexia is that the standard technique for data analysis in psychophysical experiments is prone to severe bias. Specifically, in many phoneme categorization studies, participants identify auditory stimuli that vary along some continuum as belonging to one of two groups, a psychometric function is fit to the resulting data, and the slopes of these psychometric functions are compared between groups of dyslexic and control readers (for reviews, see Vandermosten et al., 2011; Noordenbos and Serniclaes, 2015). In this approach, a steep slope indicates a clear boundary between phoneme categories, whereas a shallow slope suggests less defined categories (i.e., fuzzy phonological representations), possibly due to poor sensitivity to the auditory cue(s) that mark the phonemic contrast. Unfortunately, most studies in the dyslexia literature fit the psychometric function using algorithms that fix its asymptotes at zero and one, which is equivalent to assuming that categorization performance is perfect at the extremes of the stimulus continuum (i.e., assuming a lapse rate of zero). This assumption is questionable in light of the evidence that dyslexics may be less consistent categorizing stimuli across the full continuum (Noordenbos and Serniclaes, 2015), as well as evidence that attention, working memory, or task difficulty, rather than stimulus properties, may underlie group differences between readers with dyslexia and control subjects.

The zero lapse rate assumption is particularly problematic given that fixed asymptotes at zero and one leads to strongly downward-biased slope estimates even when true lapse rates are fairly small (Wichmann and Hill, 2001a,b). In other words, if a participant makes chance errors on categorization, and these errors happen on trials near the ends of the continuum, the canonical psychometric fitting routine used in most previous studies will underestimate the steepness of the category boundary. Thus, a tendency to make a larger number of random errors (due to inattention, memory, or task-related factors) will be wrongly attributed to an indistinct category boundary on the contrast under study. Since it is precisely the subjects with dyslexia that more frequently show attention or working memory deficits, research purporting to show less distinct phoneme boundaries in readers with dyslexia may in fact reflect non-auditory, non-linguistic differences between study populations.

Unfortunately, most studies in the dyslexia literature that use psychometric functions to model categorization performance appear to suffer from this bias. Although most do not report their analysis methods in sufficient detail to be certain, we infer (based on published plots, and the lack of any mention of asymptote estimation) that the bias is widespread (e.g., Reed (1989); Manis et al. (1997); Breier et al. (2001); Chiappe et al. (2001); Maassen et al. (2001); Bogliotti et al. (2008); Zhang et al. (2012)). A few studies report using software that in principle supports asymptote estimation during psychometric fitting, but do not report the parameters used in their analysis (e.g., Vandermosten et al., 2010, 2011). In one case, researchers who re-fit their psychometric curves without data from the continuum endpoints found a reduced effect size, prompting the authors to wonder whether any effect would remain if an unbiased estimator of slope were used. Only a few studies have fit psychometric functions with free asymptotic parameters or investigated differences in lapse rates (Collet et al., 2012; Noordenbos et al., 2013; Noordenbos and Serniclaes, 2015); however, it is worth noting that their methods did not constrain the lapse rate, which can lead to upward-biased slope estimates (Schütt et al., 2015). This is a consequence of the fact that slope and lapse parameters trade off in the optimization space of sigmoidal function fits (Treutwein and Strasburger, 1999).

In light of all these problems—the inconsistency of findings, confounding influences of experimental task design, and bias introduced by analysis choices—it is reasonable to wonder whether children with dyslexia really have any abnormality in phoneme categorization, once those factors are all controlled for. The present study addresses the relationship between reading ability and phoneme categorization ability, and in particular, whether children with dyslexia show a greater deficit on phoneme contrasts that rely on temporally varying (dynamic) cues, as opposed to static cues. This study avoids the aforementioned methodological problems by using multiple paradigms with different attentional and memory demands, and by analyzing categorization performance using Bayesian estimation of the psychometric function (Schütt et al., 2015). In this approach, the four parameters of a psychometric functionthe threshold, slope, and two asymptote parametersare assigned prior probability distributions, which formalize the experimenter's assumptions about their likely values. By allowing the asymptote parameters to vary, the slope is estimated in a less biased way than in traditional modeling approaches with fixed or unconstrained asymptotes. However, because fitting routines trade off between asymptote parameters and the slope parameter of a logistic model in optimization space (Treutwein and Strasburger, 1999), it can be difficult to estimate both accurately at the same time. To address this difficulty, we first performed cross-validation on the prior distribution of the asymptote parameters to determine the optimal model to fit the data.

This paper presents data from 44 children, aged 8-12 years, and a wide range of reading abilities. Our experimental task is based on the design of Vandermosten et al. (2011; 2010), which assessed categorization performance for two kinds of stimulus continua: those that differed based on a spectrotemporal cue (dynamic), and those that differed based on a purely spectral cue (static). In the original study, the authors concluded that children with dyslexia are specifically impaired at categorizing sounds (both speech and non-speech) that differ on the basis of dynamic cues. However, although the dynamic and static stimuli in their study were equated for overall length, the duration of the cues relevant for categorization were not equal: in the dynamic stimuli, the cue (a vowel formant transition) was available for 100 ms, but in the static stimuli, the cue (the formant frequency of a steady state vowel) was available for 350 ms. This raises the question of whether cue duration, rather than the dynamic nature of the cue, was the source of apparent impairment in categorization among participants with dyslexia. The present study avoids this confound by changing the static cue stimuli from steady-state vowels to fricative consonants (a (/ʃa/∼/sa/ continuum), so that the relevant cue duration is 100 ms in both the static (/ʃa/∼/sa/) and dynamic (/ba/∼/da/) stimulus continua. Additionally, Vandermosten and colleagues used a test paradigm in which listeners heard three sounds and were asked to decide if the third sound was more like the first or second (an ABX design). Here we included both an ABX task and a single-stimulus categorization task, to see whether the memory and attention demands of the ABX paradigm may have played a role in previous findings. Thus, by (a) assessing categorical perception of speech continua with static and dynamic cues, (b) varying the cognitive demands of the psychophysical paradigm, and (c) empirically determining the optimal parameterization of the psychometric function with cross-validation, we aim to clarify the role of auditory processing deficits in dyslexia.

2.2 Methods

2.2.1 Participants

A total of 52 native English-speaking school-aged children with normal hearing were recruited for the study in the study. Children ages 8-12, without histories of neurological or auditory disorders, were recruited from a database of volunteers in the Seattle area (University of Washington Reading and Dyslexia Research Database). Parents and/or legal guardians of all participants provided written informed consent under a protocol that was approved by the University of Washington Institutional Review Board. All experiments were performed in accordance with relevant guidelines and regulations of the Institutional Review Board. All subjects had normal or corrected-to-normal vision. Participants were tested on a battery of assessments, including the Woodcock-Johnson IV (WJ-IV) Letter Word Identification and Word Attack sub-tests (Schrank et al., 2014), the Test of Word Reading Efficiency (TOWRE) (Torgesen et al., 2011), and the Wechsler Abbreviated Scale of Intelligence (WASI-II) (Wechsler, 2011). All participants underwent a hearing screening to ensure pure tone detection at 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz and 8000 Hz in both ears at 25 dB HL or better. A total of six subjects who were initially recruited did not pass the hearing screening and were not entered into the study, and two more were unable to successfully complete training (described below). Note that our screening was not designed to estimate pure tone thresholds, and as such we cannot be sure whether the participants who failed had a sensory limitation, did not comply with instructions, or could not complete the task for another reason. In total, a total of 44 children participated in the full experiment.

2.2.2 Demographics

In order to understand the relationship between phoneme categorizatino ability and reading ability, we selected our cohort of participants to encompass both impaired and highly skilled readers. Although we treat reading ability as a continuous covariate in our statistical analyses, for the purpose of recruitement and data visualization we defined three groups

Table 2.1: Summary statistics and group differences on various demographic and behavioral measures.

	Above Average	Below Average	Dyslexic	Significance: Above Avg. v. Dys.	Significance: Below Avg. v. Dys.
	$n = 13$ 3♂, 10♀	$n = 16$ 9♂, 7♀	$n = 15$ 5♂, 10♀		
WASI-II					
FS-2	124 (15.4)	109.8 (12)	99.9 (14.3)	0.029	< 0.001
Nonverbal IQ	59.6 (8.6)	54.1 (9.5)	46 (8.2)	0.008	< 0.001
Woodcock-Johnson IV					
Basic Reading Score	117.1 (11.7)	94.6 (4.4)	75.1 (6.8)	< 0.001	< 0.001
Nonword	115.2 (12.3)	97.5 (4.8)	84.2 (6.8)	< 0.001	< 0.001
Real Word	116.5 (12.6)	93 (6.8)	69.7 (10.5)	< 0.001	< 0.001
TOWRE 2					
TOWRE Index	113.3 (9.1)	85.4 (10.6)	68.5 (7.6)	< 0.001	< 0.001
Nonword	109.8 (10.9)	84 (9.3)	71.9 (4.5)	< 0.001	< 0.001
Real Word	115.3 (11.5)	88.4 (12.4)	68.2 (13.1)	< 0.001	< 0.001
CTOPP 2					
Phonological Awareness	107.2 (13.3)	95.8 (14.6)	86.6 (9.5)	0.11	< 0.001
Phonological Memory	107.8 (17.5)	93.5 (10.1)	88.1 (17.5)	0.63	0.006
Rapid Naming	98.6 (9.1)	82.6 (8.7)	75 (8.4)	0.068	< 0.001

based on the composite Woodcock-Johnson Basic Reading Score (WJ-BRS) and TOWRE Index. The "Dyslexic" group comprised participants who scored 1 standard deviation or more below the mean (i.e., scores ≤ 85) on both the WJ-BRS and TOWRE Index; Above Average readers were defined as those who scored above the mean on both tests; and Below Average readers were defined as participants who fell between the Dyslexic and Above Average groups. We used both the WJ-BRS and TOWRE Index in our criterion to improve the confidence of our group assignments, though they are highly correlated measures in our sample ($r = 0.89, p < 0.001$). There were 15 subjects in the Dyslexic group, 16 in the Below Average group, and 13 in the Above Average group. There were no significant differences in age between groups (Kruskal-Wallis rank sum test, $H(2) = 1.048, p = 0.592$), nor was there a significant correlation between age and WJ-BRS score ($r = 0.02, p = 0.890$). We did not exclude participants with ADHD diagnoses from the study because ADHD is highly comorbid with dyslexia (Light et al., 1995; Stevenson et al., 2005), so this inclusion contributes to a more representative sample of children. However, we did account for the presence of ADHD diagnosis in our statistical models. Of our 44 participants, 7 had a formal diagnosis of ADHD: 2 in the Above Average group, 1 in the Below Average group, and 4 in the Dyslexic group.

Table 2.1 shows group comparisons on measures of reading and cognitive skills, and the distributions of each variable are illustrated in Supplementary Figure A.1 in Appendix A. Pairwise correlations between behavioral measures are shown in Supplementary Figure A.2. All subjects had an IQ of at least 80 as measured by the WASI-II FS-2. Therefore, although there was a significant difference in IQ scores across groups, we were not concerned that abnormally low cognitive ability would prevent any child from performing the experimental task. While we expected IQ scores to be higher among stronger readers due to the verbal component of IQ assessment, we also noticed a significant difference in nonverbal IQ across the groups as measured on the WASI-II Matrix Reasoning test. To be certain our results were not confounded by this difference, we also included nonverbal IQ as a covariate in our statistical analyses to confirm the specificity of the relationship with reading skills as opposed to IQ.

2.2.3 Stimuli

Two 7-step speech continua were created using Praat v. 6.0.37 (Boersma and Weenink, 2019), a /ba/∼/da/ continuum and a /ʃa/∼/sa/ continuum, chosen to probe categorization on the basis of dynamic and static auditory cues, respectively. In the /ba/∼/da/ continuum, the starting frequency of the second vowel formant (F2) transition was varied. Individuals who are insensitive to this spectrotemporal modulation should have shallow psychometric functions on the /ba/∼/da/ categorization task. The other continuum was /ʃa/∼/sa/, in which the spectral shape of the fricative noise was varied between tokens (but was static for any given token). Both the /ʃa/∼/sa/ fricative duration and the /ba/∼/da/ F2 transition duration were 100 ms. Endpoint and steady-state formant values for the /ba/∼/da/ continuum were measured from recorded syllables from an adult male speaker of American English. The spectral peaks of the /ʃa/∼/sa/ continuum were chosen from the spectral peaks measured in a recorded /sa/ spoken by the same talker. Spectrograms of the endpoints of both continua are shown in Supplemental Figure A.3 in Appendix A.

The /ba/∼/da/ continuum

Synthesis of the /ba/∼/da/ continuum followed the procedure described in Winn and Litovsky (2015). Briefly, this procedure involves downsampling a naturally produced /ba/ token, extracting the first four formant contours via linear predictive coding, altering the F2 formant contour to change the starting frequency from 1085 Hz (/ba/) to 1460 Hz (/da/) in seven linearly-spaced steps, linearly interpolating F2 values for the ensuing 100 ms to make a smooth transition to the steady-state portion of the vowel (1225 Hz), and regenerating the speech waveform using Praat's source-filter synthesis (with a source signal extracted from a neutral vowel from the same talker). Because this procedure eliminates high-frequency energy in the signal at the downsampling step, the synthesized speech sounds are then low-pass filtered at 3500 Hz and combined with a version of the original /ba/ recording that had been high-pass filtered above 3500 Hz. This improves the naturalness of the synthesized sounds, while still ensuring that the only differences between continuum steps are in the F2 formant transitions.

The /ʃa/∼/sa/ continuum

The /ʃa/∼/sa/ continuum was created by splicing synthesized fricatives lasting 100 ms onto a natural /a/ token excised from a spoken /sa/ syllable. The duration of /a/ was scaled to 250 ms using Praat's implementaion of the PSOLA algorithm (Moulines and Charpentier, 1990). Synthesized fricatives contained three spectral peaks centered at 3000, 6000, and 8000 Hz. The bandwidths and amplitudes of the spectral peaks were linearly interpolated between continuum endpoints in seven steps, and the resulting spectra were used to filter white noise. To improve the naturalness of the synthesized fricatives, a gentle cosine ramp over 75 ms and fall over the final 20 ms was imposed on the fricative envelope. Aside from this onset/offset ramping (which was applied equally to all continuum steps), the contrastive cue (the amplitudes and bandwidths of the spectral peaks) was steady throughout the 100-ms duration of each fricative.

2.2.4 Procedure

Stimulus presentation and participant response collection was managed with PsychToolbox for MATLAB (Brainard, 1997). Auditory stimuli were presented at 75 dB SPL via circumaural headphones (Sennheiser HD 600). Children were trained to associate sounds from the two speech continua with animal cartoons on the left and right sides of the screen and to indicate their answers with right or left arrow keypresses. For the /ʃa/∼/sa/continuum, participants selected between pink and purple snakes. For the /ba/∼/da/ continuum, participants selected between two different sheep cartoons. Throughout all blocks, each cartoon was always associated with the same stimulus endpoint. Two experimental conditions were presented: an ABX condition and a single-interval condition.

The ABX condition

In ABX blocks, participants performed a two-alternative forced-choice identification task in which they heard three stimuli with a 400 ms ISI and indicated whether the third stimulus (X) was more like the first (A) or second (B) stimulus. Stimuli A and B were always endpoints of the continuum, and during the endpoint presentations, the animals associated with those sounds lit up. Based on pilot data with adult participants, varying endpoint order trial-by-trial (i.e., interleaving /ba/-/da/-X and /da/-/ba/-X trials in the same block) was deemed very difficult, so for our young listeners we fixed the order of endpoint stimulus presentation within each block. The order of blocks was counterbalanced across participants. Each block contained 10 trials for each of the seven continuum steps, for a total of 70 trials per block.

The single interval condition

In single interval blocks, participants heard a single syllable and decided which category it belonged to by selecting an animal. No text labels were used, so participants learned to associate each animal with a continuum endpoint during a practice round. Two subjects (neither in the Dyslexia group) lost track of the animals associated with the endpoints after one successful practice round and were given brief additional instruction before successfully completing the task. Each block contained 10 trials for each step on the continuum, for a total of 70 trials.

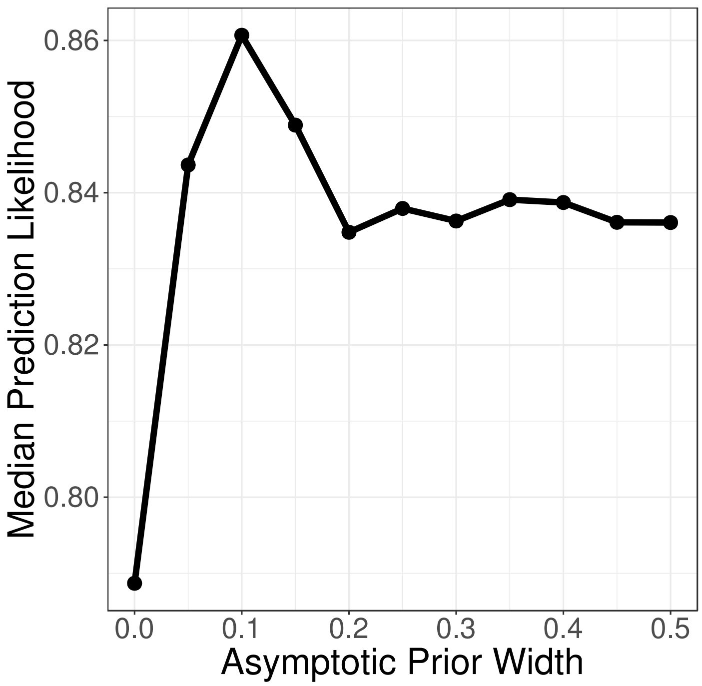
There were thus six blocks total: an ABX /ba/∼/da/ block, an ABX /ʃa/∼/sa/ block, a second ABX /ba/∼/da/ block, a second ABX /ʃa/∼/sa/ block, a single-interval /ba/∼/da/ block, and finally a single-interval /ʃa/∼/sa/ block. Other than the counterbalancing of ABX endpoint order mentioned above, the order of blocks was the same for all participants. Practice rounds were administered before the first ABX round and before each of the single-interval blocks. In practice rounds, participants were asked to categorize only endpoint stimuli and were given feedback on every trial. Participants had to score at least 75% correct on the practice round to advance to the experiment and were allowed to repeat the practice blocks up to three times. Two children did not meet this criterion and were not included in the study (one from the Above Average group and one from the Dyslexic group).

2.2.5 Psychometric Curve Fitting

Modeling of response data was performed with Psignifit 4.0 (Schütt et al., 2015), a MATLAB toolbox that implements Bayesian inference to fit psychometric functions. We fit a logistic curve with four parameters, modeling the two asymptotes, the width of the psychometric, and the threshold. The width parameter was later transformed to the slope at the threshold value. Four-parameter fits of each experimental block were performed with each of 11 possible priors: a prior that fixed the lapse rate at zero (in line with the approach in most previous studies of auditory processing in dyslexia), and ten uniform distribution priors with lower bound zero and upper bound ranging from 5% to 50% in steps of 5% (e.g., a lapse rate of 20% corresponds to constraining the lower asymptote between 0 and 0.2, and the upper asymptote between 0.8 and 1). Next, the optimal prior was chosen using leave-one-out cross-validation. Specifically, for each of the 261 test blocks across all participants, psychometric curves were fit to 69 of the 70 data points in that block, and the likelihood of the participant's selection on the held-out data point was calculated under the model. This process was repeated for each of the 70 data points in each block, for each of the 11 possible prior widths. The estimated likelihoods of the held-out data points were used as a goodness-of-fit metric and pooled across blocks and cross-validation runs to determine the median likelihood for each prior width. The optimal prior was determined to be a maximum lapse rate of 10%, as seen in Figure 2.1; note that when the asymptotes were fixed at 0 and 1, the psychometric fits had the poorest fit to the data.

Having determined the optimal prior for the asymptote parameters, models were re-fit to obtain final estimates of the four parameters, for each combination of participant, cue continuum, and test paradigm. Additionally, 2-parameter fits (with asymptotes fixed at 0 and 1) were also recomputed, for later comparison with the optimal 4-parameter models. Any cases where the best-fit threshold parameter was not in the range of the stimulus continuum

Figure 2.1: The median likelihood of held-out data points from all cross-validation trials.



The median likelihood of held-out data points from all cross-validation trials is plotted against the width of the asymptotic parameter prior distributions (leave-one-out cross validation). The optimal prior width is clearly shown to be one that allows lapse rate to vary between 0 and 0.1. However, in all cases the 4-parameter model fit the data better than the 2-parameter model.

steps (1 through 7) were excluded from further analysis. Of the 261 total psychometric functions fit in the study, 12 of the 2-parameter fits and 10 of the 4-parameter fits were excluded on these grounds.

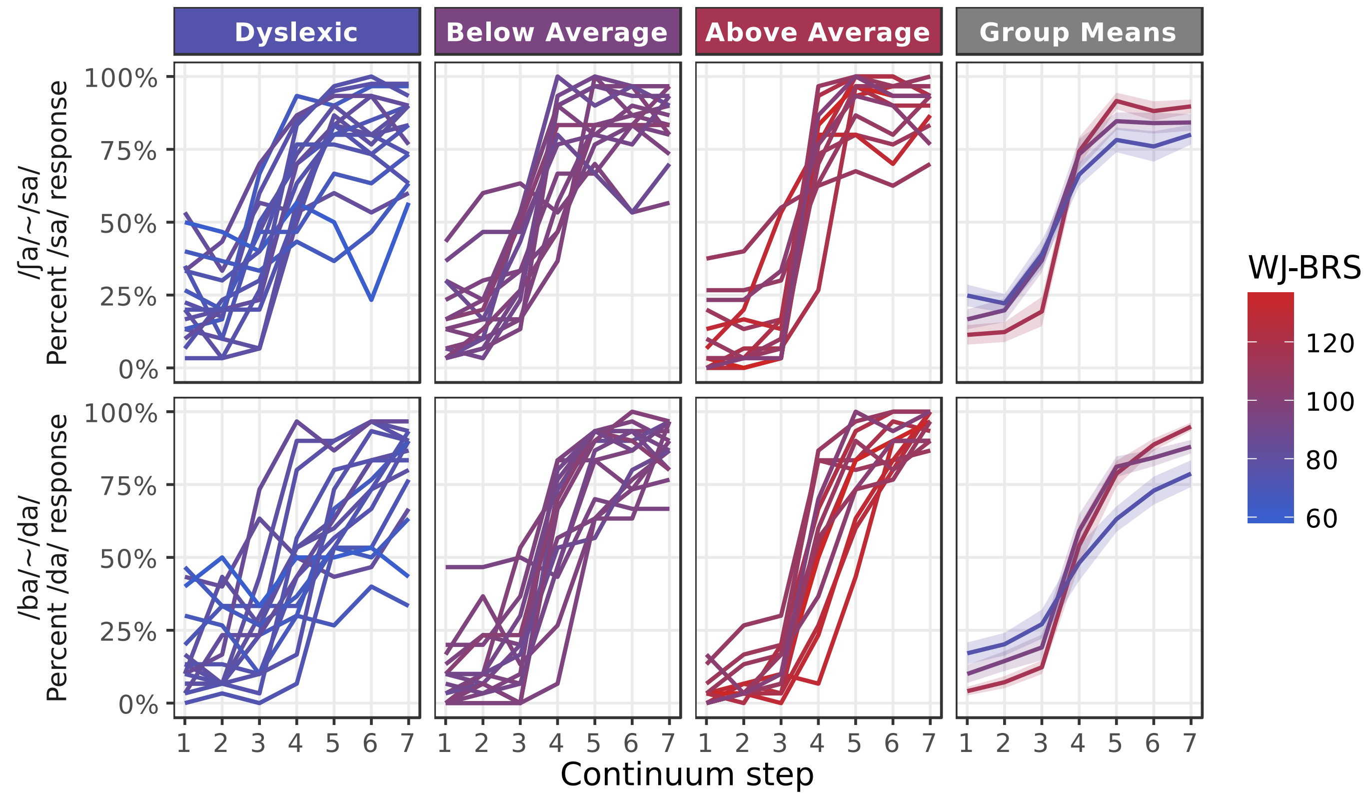
2.2.6 Reaction Time Analysis

Post-hoc analysis of reaction time data was also performed. To analyze reaction time, we first removed any trials with a reaction time shorter than 200 ms or longer than 2 s. The lower cutoff was based on the minimum response time to an auditory stimulus (Fry, 1975); such responses are assumed to be spurious or accidental button presses, and resulted in the exclusion of ~9% of trials. The upper cutoff was chosen based on the histograms of reaction times to exclude trials where subjects became distracted and were clearly temporarily disengaged from the experiment, and resulted in the exclusion of ~6% of trials. We fit each individual's remaining data as a function of continuum step with a 2^{nd} order orthogonal polynomial basis (see Results for rationale), reading ability, and stimulus continuum with a mixed-effects regression model. For visualization, the reaction time data from each experimental block for each participant were reduced by fitting a χ^2 distribution to the reaction times at each continuum step, and using the peak of the fitted distribution as our summary statistic (because reaction time distributions tend to be right-skewed, making the mean a poor measure of central tendency).

2.3 Results

Response functions for each subject on the two stimulus continua (aggregated across task paradigms) are shown in Figure 2.2. Subjects were divided into three groups (Dyslexic, Below Average, and Above Average readers) to be consistent with previous studies that have reported group comparisons, and reading score (as indexed by the Woodcock Johnson Basic Reading Skills standard score (WJ-BRS)) was also treated as a continuous variable in line with the perspective that dyslexia represent the lower end of a continuous distribution (Shaywitz et al., 1992a).

Figure 2.2: Response functions for each subject and each speech continuum.



Results are averaged over responses from both the ABX and single-interval paradigms. The color of each line maps to the subject's reading ability based on the Woodcock Johnson Basic Reading Skills Composite score (WJ-BRS).

For Above Average readers, response functions show the typical sigmoid shape, with a steep slope marking the boundary between categories and consistent labeling within categories; for readers in the Below Average and Dyslexic groups, response functions are much more heterogeneous, with some subjects exhibiting inconsistent performance even at continuum endpoints. The group means are suggestive of differences in both psychometric slope and lapse rate on both the static (/ʃa/∼/sa/) and dynamic (/ba/∼/da/) cue continua.

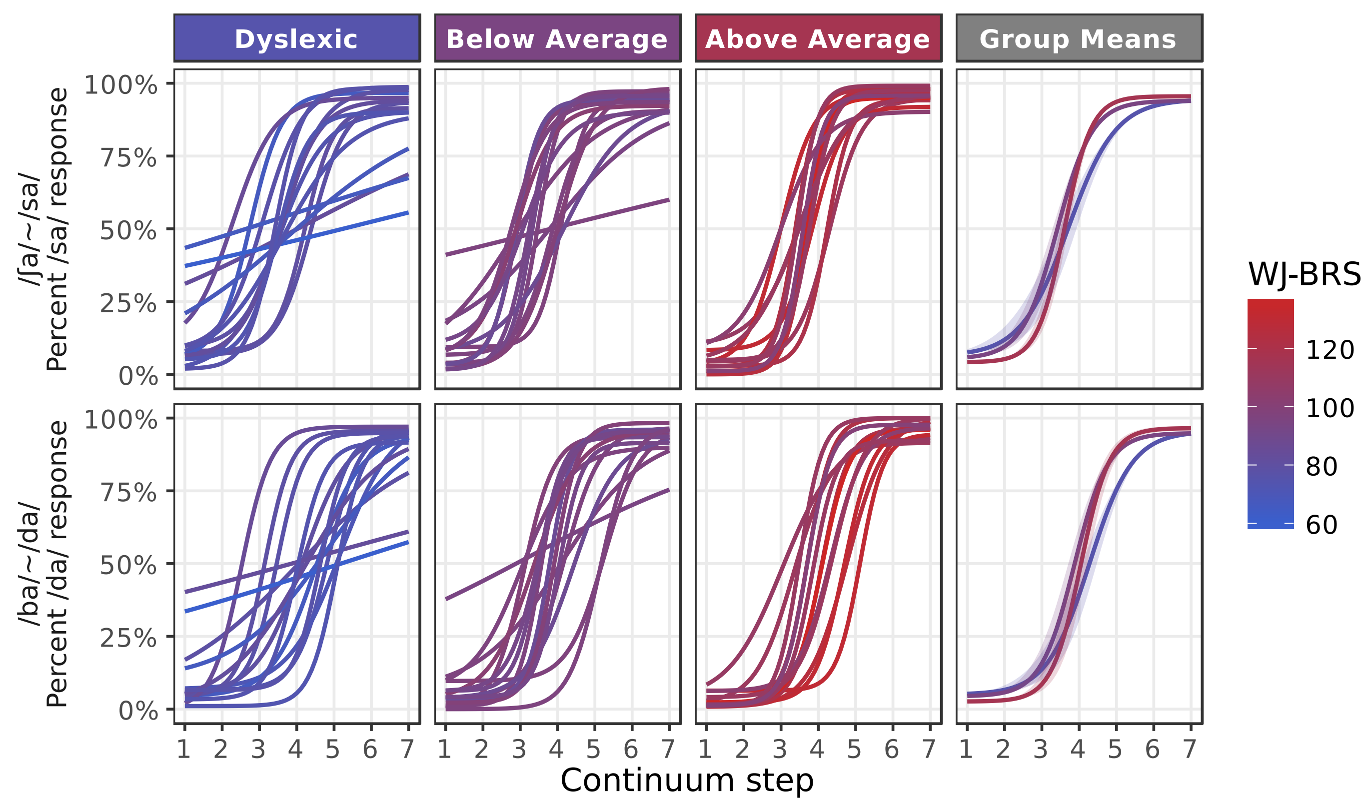
Figure 2.3 shows the psychometric functions fit to each subject's response data on the two continua (aggregated across paradigms). The fitted curves reflect the same pattern seen in the raw data: subjects with high reading scores tend to show consistent categorization at stimulus endpoints and steep category transitions near the center of the stimulus continuum, whereas subjects with poor reading scores show a diversity of curve shapes. For some subjects with poor reading scores the psychometric curves appear quite similar to those of the skilled readers, while others exhibit shallow slopes at the category transition and/or high lapse rates at the continuum endpoints. Estimates of the category boundary (the threshold parameter of the psychometric) also appear to span a wider range of continuum values among poor readers compared with readers in the above average group.

2.3.1 High reading score predicts consistent phoneme identification

The group-level patterns of slope and lapse rate estimates are summarized in Figure 2.4. There is a clear trend of increasingly steeper slopes for children with higher reading scores, across both stimulus continua and both experimental paradigms. The pattern of lapse rate estimates is less consistent, though the lapse rate estimates tended to be higher for poor readers. Slope and lapse rate estimates were each modeled using mixed effects regression, as implemented in the `lme4` library for R. Fixed-effect predictors with deviation coding were used for both the continuum (static /ʃa/∼/sa/ versus dynamic /ba/∼/da/) and paradigm (ABX versus single-stimulus) variables, and reading ability (WJ-BRS score) was included as a continuous fixed-effect predictor. A random intercept by participant was also included. Fixed-effect predictors were also included to control for the effects of ADHD diagnosis and nonverbal IQ (age-normalized WASI Matrix Reasoning score). Because two blocks of the ABX test were administered per subject, slope or lapse rate parameter estimates were averaged over the two blocks for each individual before regression modeling.

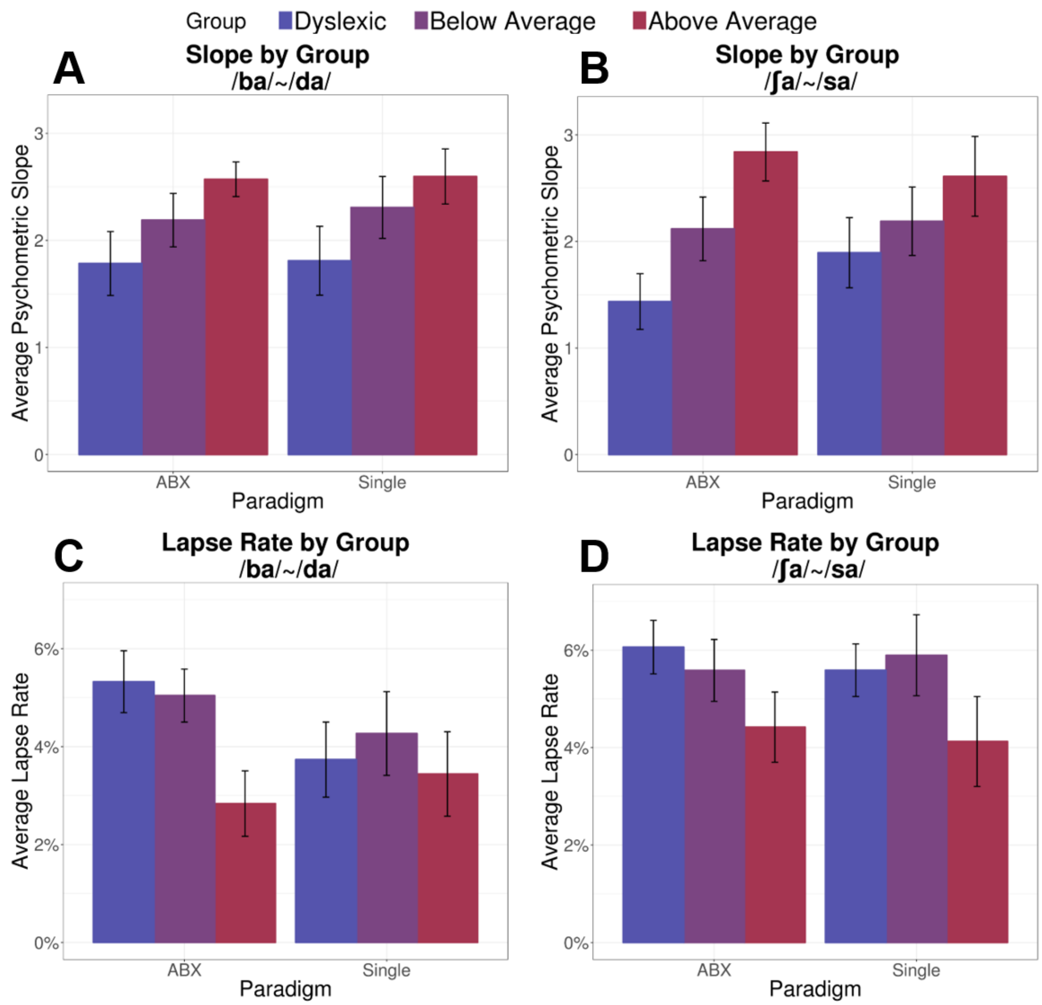
The fully-specified model of psychometric slope estimates indicated that only one predictorWJ-BRS reading scorehad a significant effect on slope estimates; there was no reliable difference in psychometric slopes between the static and dynamic cue continua, nor was there a reliable difference between ABX blocks and single-interval blocks. Stepwise nested model

Figure 2.3: Fitted psychometric functions for each subject and each continuum.



Parameters (slope, threshold, and the two asymptotic parameters) were averaged over estimates from both the ABX and single-interval paradigms. The color of each line maps to the subject's reading ability (WJ-BRS score).

Figure 2.4: Average psychometric function parameters by group.



Average psychometric slope within groups for the /ba/∼/da/ and /ʃa/∼/sa/ continua are shown in panels A and B, separated by paradigm. Average lapse rate parameters for both continua are shown in Panels C and D. Error bars represent one standard error of the mean.

Table 2.2: Summary of four selected linear mixed effects models.

Model	Parameter	β	SE	p
4-parameter slope	WJ-BRS	0.023	0.006	0.0005
4-parameter lapse	WJ-BRS	-0.0004	0.0001	0.002
	Continuum (static)	0.010	0.003	0.007
2-parameter slope	WJ-BRS	0.017	0.004	0.0004
	Continuum (static)	-0.167	0.078	0.039
1st principal component	WJ-BRS	0.021	0.006	0.0007
	WJ-BRS*Paradigm	0.012	0.006	0.048

comparisons were then used to eliminate irrelevant model parameters and yield the most parsimonious model (see Appendix A for details). The most parsimonious model of psychometric slope contained only a continuous predictor of reading ability (WJ-BRS) and a random intercept for each participant (see Table 2.2). The effect size of reading ability on psychometric slope was relatively modest; a 1-unit improvement in WJ-BRS score was associated with an increase of 0.023 in the slope of the psychometric function (slope is measured in units of probability (of selecting /da/ or /sa/) per step along the 7-step continua). This indicates that poor readers tend to have shallower psychometric slopes ($p = 0.0005$). For comparison, increasing age by 1 year was associated with an increase of 0.256 in the slope of the psychometric function. In other words, it takes an 11-point increase in reading score to approximate the change in slope associated with a 1-year difference in age.

To model the relationship between lapse rate and reading score, we first combined the two asymptote estimates from the psychometric fits, by averaging the asymptote deviation from zero or one (i.e., a lower asymptote of 0.1 and an upper asymptote of 0.9 both correspond to a lapse rate of 0.1; the upper and lower lapse rates were averaged for each psychometric fit). The same initial model specification and simplification procedure used with the slope model was also used for the lapse rate model. The final model of lapse rate contained a continuous predictor for reading ability, a categorical predictor for continuum, and a random effect for participant (see Table 2.2); no reliable effect of task paradigm was detected. Higher reading ability was associated with smaller lapse rates (about 0.04% smaller per 1-unit increase of WJ-BRS score), and the static /ʃa/∼/sa/ continuum was associated with larger lapse rates (about 1% higher than the lapse rate estimates for the /ba/∼/da/ continuum). Thus, compared to strong readers, poor readers show shallower psychometric slopes and are also more likely to make errors even near the continuum endpoints. There was a general tendency for all participants to make more errors near continuum endpoints with the static-cue stimuli than with the dynamic-cue stimuli.

We next analyzed these relationships treating reading level as a categorical rather than a continuous variable. Using a mixed model analysis with group (Dyslexic, Below Average, and Above Average) as a categorical predictor with Above Average as the reference group, and subject as a random factor, and following the same model selection procedure as before, we confirmed the relationship between group and psychometric slope (F(2,40.72) = 6.614, $p = 0.003$). The selected model did not include effects of continuum or paradigm. To assess the separability of the groups, we calculated Cohen's d for a group comparison where each individual's slope estimate is averaged across the six test blocks (combining paradigms and continua). We found a large effect size for the comparison between the Dyslexic and Above Average groups ($d = 1.48$) and a medium effect size for the comparison between Dyslexic and Below Average groups ($d = 0.66$). Indeed a linear discriminant analysis (LDA) classification model was able to accurately classify subjects as Dyslexic versus Above Average readers 72.4% of the time (prediction accuracy assessed using leave-one-out cross validation). The same model achieved 51.5% accuracy when classifying Dyslexics versus Below Average readers.

An analysis of lapse rates across reading groups found a significant main effect of group (F(2,41.09) = 5.298, $p = 0.009$). Belonging to the Below Average group was associated with an increased lapse rate by 1.48% ($p = 0.0140$) and belonging to the Dyslexic group was associated with an increased lapse rate by 1.76% ($p = 0.004$) compared to the Above Average group. Using a quadratic discriminant analysis (QDA) with both slope and lapse rate did not improve cross-validated prediction accuracy for classifying Dyslexics versus Above Average readers (72.4%) but did improve the ability to classify Dyslexics versus Below Average readers (64.5%).

2.3.2 Traditional model fitting misrepresents effect of cue continuum

For comparison with the methods commonly used in prior studies, we also fit a mixed effects model using slope estimates from a traditional 2-parameter psychometric function fit (with asymptotes fixed at 0 and 1, see Methods), and performed the same model simplification procedure. In this case, the most parsimonious model indicated a significant effect of stimulus continuum on psychometric slope, in addition to the effect of reading ability seen in the model of 4-parameter fits (see Table 2.2). The magnitude of the estimated effect of continuum on psychometric slope is quite large (difference in slope of 0.167 between dynamic and static cue continua, comparable to a difference of about 10 points on the WJ-BRS), with lower slopes associated with the static-cue (/ʃa/∼/sa/) continuum. Comparing this to the models of 4-parameter fits, we see reasonably similar estimates of the effect of reading ability on psychometric slope, but the effect of stimulus continuum is allocated differently depending on the underlying assumptions of the psychometric fitting routine.

2.3.3 Principal component analysis reveals an effect of task paradigm

Although slope and lapse rate are often thought of as representing separate cognitive processes (category discrimination and attention, respectively), in reality it is difficult to disentangle what is being modeled by each parameter: the two are not truly independent in the psychometric fitting procedure. Moreover, in speech continua, slope and lapse rate may be influenced by the same underlying processes: the endpoints are determined by natural speech tokens, which may not be reliably categorized 100% of the time by all subjects, due to the complex relationships among speech cues, talker/listener dialect, context, etc. Abnormal categorical perception that manifests as shallower psychometric slope may also affect the way stimuli at the endpoints of the continuum are labelled (Serniclaes et al., 2004). For these reasons, we also performed a principal components analysis that combined slope and asymptote estimates into a single index of psychometric curve shape. The first principal component was a linear combination of the slope and both asymptotes (slope: 0.678; lower asymptote: -0.268; upper asymptote: -0.685) that explains about 41% of variance in the data. Using an identical modeling procedure to the analysis of slope and lapse rate, we iteratively eliminated predictors from a fully specified mixed model of the first principal component. The most parsimonious model showed a main effect of reading ability (WJ-BRS score) and a significant interaction between reading ability and experimental paradigm, but no effect of stimulus continuum (see Table 2.2). The main effect indicated that poor readers tend toward a combination of higher lapse rates and shallower slopes, while strong readers tend to have lower lapse rates and steeper slopes. The effect is more pronounced in the ABX paradigm than in the single-interval paradigm. Thus, in line with previous hypotheses (Banai and Ahissar, 2006; Ramus and Szenkovits, 2008), this model suggests that the specific paradigm used in a psychophysical experiment can amplify deficits in people with dyslexia. However, struggling readers still perform poorly on phoneme categorization tasks irrespective of the paradigm used in the experiment.

2.3.4 Poor readers struggle even with unambiguous stimuli at continuum endpoints

The lapse rate results discussed above suggest that children with dyslexia may have considerable difficulty even in categorizing unambiguous stimuli at the continuum endpoints. We performed post-hoc analyses of accuracy and reaction time to further investigate this possibility. In particular, we were interested in whether the high lapse rate estimates resulted from poor readers performance degrading over the course of an experimental block, indicating an inability to stay focused and pay attention for the duration of the experiment. We were also interested in whether the pattern of reaction times across different parts of the stimulus continuum varied systematically with reading ability.

Accuracy at continuum endpoints

Because our young subjects ranged considerably in attention span and motivation, we analyzed the stationarity of response accuracy at continuum endpoints, to ascertain whether task performance declined from start to finish. Following the design of Messaoud-Galusi et al. (2011), for each test block (6 total per subject) we computed the number of correct assignments of the stimulus endpoints. There were twenty endpoint stimulus presentations per block.

We then modeled correct endpoint stimulus categorization as a function of trial number (1 through 70), reading score, stimulus continuum, and test paradigm (ABX or Single-stimulus), plus covariates for age and ADHD diagnosis and a random effect of subject. WJ-BRS, age, and trial number were centered for clear interpretation of interaction terms. Following the same model fitting and simplification procedures described above, we found the most parsimonious model included only main effects of WJ-BRS (β = 0.004, SE = 0.001, p = 0.001) and trial number (β = 0.0008, SE = 0.0003, p = 0.022). The main effect of WJ-BRS indicated that poor readers were more likely than typical readers to err when categorizing unambiguous stimuli, echoing the earlier analysis of lapse rate. The main effect of trial number indicates that all readers tended to decline in accuracy as the test block progressed. The lack of a significant interaction between trial number and WJ-BRS indicates that on average, poor readers were no more likely than other subjects to decline in performance due to fatigue or inattention as the test block wore on. Thus, inability to stay focused over the course of a psychophysics experiment does not explain phoneme categorization deficits in our sample of poor readers.

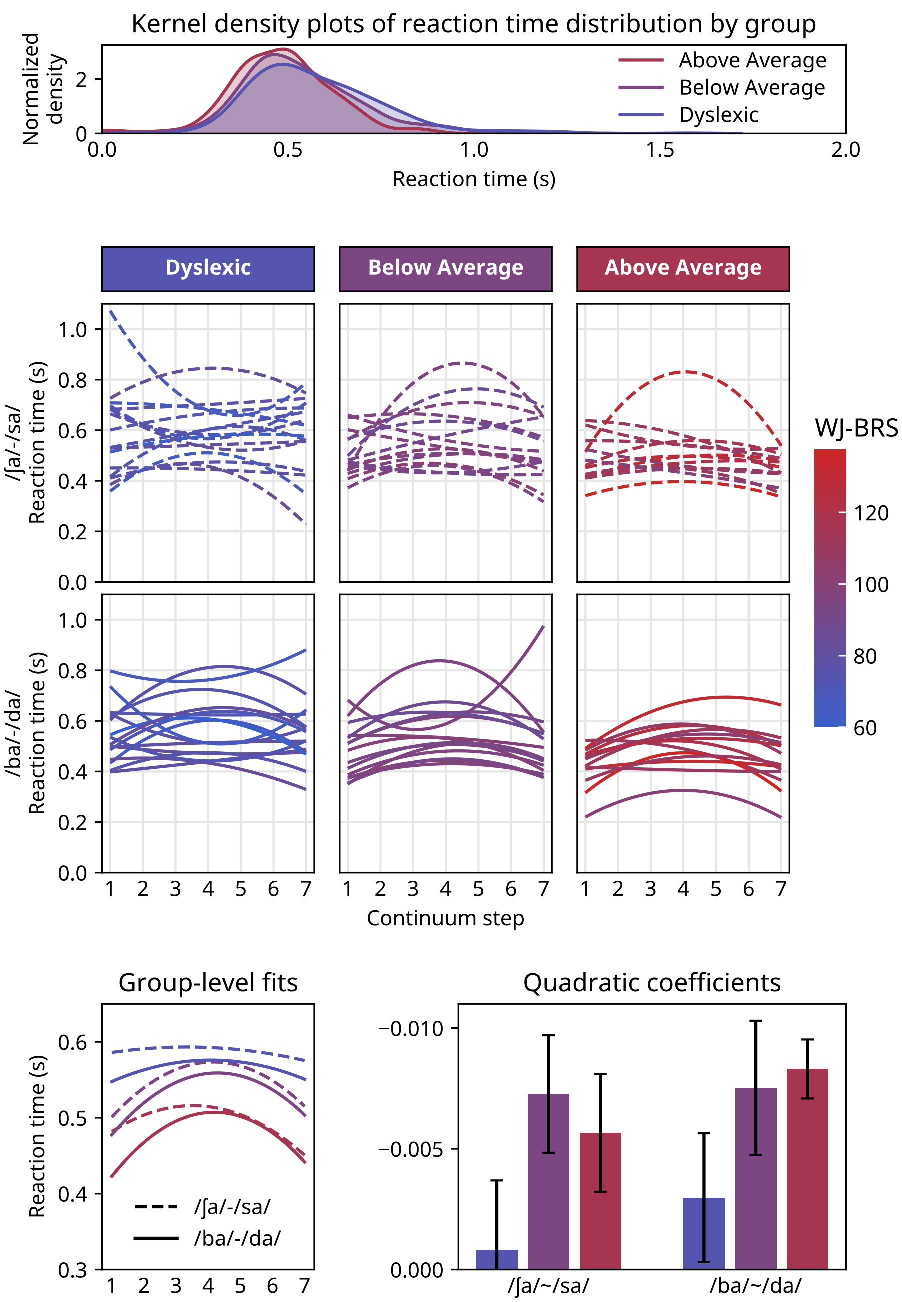
2.3.5 Reaction time analysis

With stimulus continua that span a categorical boundary, it is expected that reaction times will be longer for stimuli near the category boundary (Pisoni and Tash, 1974; Repp, 1981), reflecting the perceptual uncertainty and resulting decision-making difficulty associated with ambiguous stimuli. Although our participants were not under time constraints in this study, they were encouraged to respond as soon as they could after stimulus presentation, and reaction times were recorded. To see whether the children we tested showed the expected pattern (faster reaction times on unambiguous stimuli at or near continuum endpoints, and slower reaction times on ambiguous stimuli near continuum centers), we modeled reaction times across continuum steps with a second-order polynomial basis (and its interaction with reading ability).

Kernel density estimates of the reaction time distributions for the three groups are shown in Figure 2.5A, individual and group-level polynomial fits are shown in Figure 2.5B and C, and the quadratic coefficients of the polynomial fits are shown in Figure 2.5D. The figure separates data from the two stimulus continua for easier comparison with Figures 2.2 and 2.3, but since we had no specific hypothesis about reaction time differences between continua, we did not include that comparison in the post-hoc analysis.

Model results showed an overall relationship between reading ability and reaction time ($\beta = 0.0034$, SE $= 0.0009$, $p = 1.69e-4$) indicating an average 2.5 ms speed-up in reaction time per unit change in WJ-BRS score, although there was substantial overlap in reaction time distributions across groups (cf. Figure 2.5A). The quadratic polynomial coefficient was also significant ($\beta = 2.58$, SE $= 0.447$, $p = 7.83e-9$) showing a general tendency for faster responses at continuum endpoints. Finally, there was a significant interaction between the quadratic polynomial coefficient and WJ-BRS score ($\beta = 0.082$, SE $= 0.024$, $p = 5.29e-4$), indicating a more downward-curving polynomial for subjects with higher reading scores (cf. Figure 2.5B-D). In other words, good reading ability was associated with the expected pattern of reaction times along the stimulus continuum, whereas poor readers were less likely to show an expected pattern; many individual fits showed nearly flat fitted curves or even an inverse pattern to that of the above-average readers. This analysis indicates that children with dyslexia may experience synthesized speech continua in a qualitatively different way from typical readers. Specifically, children with dyslexia are more likely to respond endpoint stimuli and ambiguous stimuli in more or less the same way, whereas above-average readers

Figure 2.5: Reaction times by group.



A: Gaussian kernel density plots (75 ms bandwidth) of raw reaction times pooled across continua and paradigms for each group. B: polynomial fits to reaction times for both speech continua. C: polynomials fit to aggregated participant reaction times for each of the two speech continua. D: mean quadratic coefficients (± 1 standard error of the mean) of the polynomial fits for subjects in each group for the two speech continua.

linger over ambiguous stimuli while categorizing endpoint stimuli more rapidly.

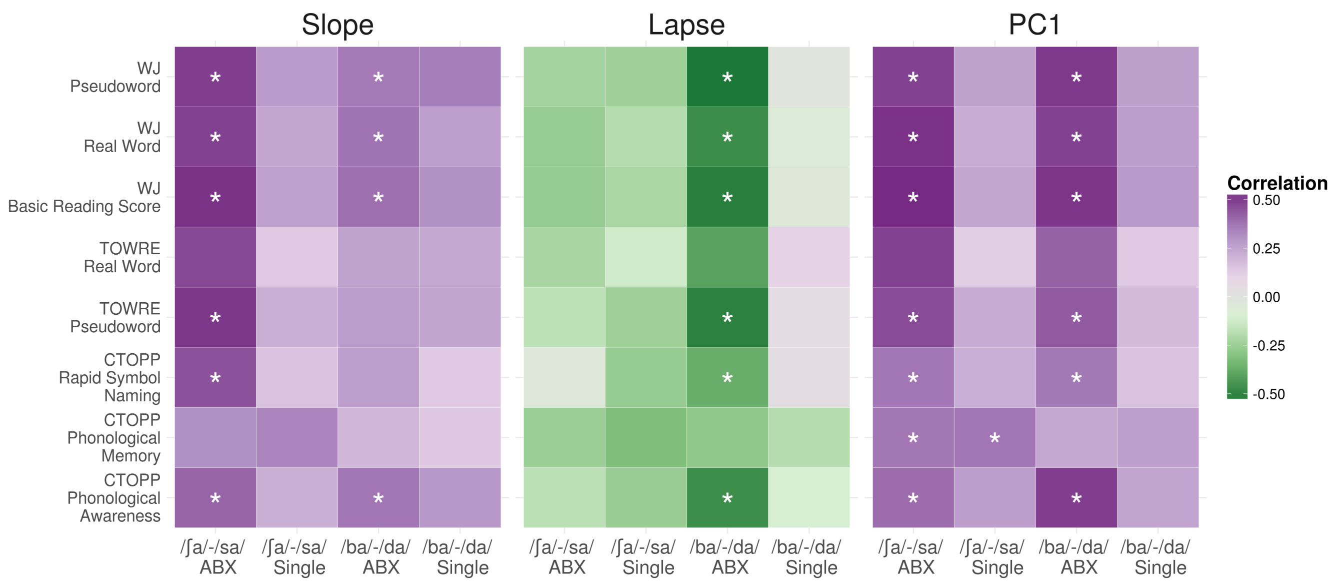
2.3.6 Other measures of reading ability

As a final post-hoc test, we explored which specific components of reading skill might underlie the relationship with performance on speech categorization tasks. To do this, we performed Pearson correlation tests between our estimates of psychometric slope, lapse rate, or the principal component that combines slope and asymptote parameters, and several tests targeting specific aspects of reading ability. Those correlations are summarized in Figure 2.6. The most striking pattern is that the majority of significant correlations between test scores and our model parameters correlate with results from the ABX blocks, but not the single-stimulus blocks. Moreover, it appears that correlations between test scores and lapse rate are clustered on the dynamic-cue /ba/∼/da/ continuum, whereas correlations with psychometric slope are clustered on the static-cue /ʃa/∼/sa/ continuum. Perhaps unsurprisingly, the correlations between test scores and the principal component combining slope and asymptote parameters look something like a union of the correlation patterns for slope and lapse rate. These patterns suggest a striking difference between the ABX paradigm and the single-stimulus paradigm, in how performance on such tasks relates to reading ability.

Pairwise correlations of psychometric function parameters with tests targeting phonological processing were similar to correlations with real word and pseudoword reading measures (Figure 2.6). However, a mixed model predicting psychometric slope as a function of phonological awareness (CTOPP Phonological Awareness) showed that after controlling for nonverbal IQ, the main effect of phonological awareness only showed a weak relationship to psychometric slope ($\beta = 0.018$, SE $= 0.009$, $p = 0.054$). Unsurprisingly, phonological awareness, as well as two other measures of phonological processing (phonological memory and rapid symbol naming) were significantly correlated with WJ-BRS in our sample (Figure A.2).

We therefore wanted to test the frequently discussed hypothesis that phonological awareness mediates the relationship between speech perception and reading ability. In other words, we asked whether our data supported the theory that abnormal labeling of phonemes is as-

Figure 2.6: Correlation between psychometric function parameters and behavioral measures.



Pearson correlations between behavioral measures and slope, lapse, and first principal component (PC1) of the psychometric functions estimated from behavioral responses. Color gradient indicates the magnitude of the correlation and asterisks indicate significant correlations ($p < 0.05$, corrected for multiple tests using false discovery rate estimation).

sociated with poor phonological awareness, which in turn drives reading difficulty.

Following the mediation analysis approach of Baron and Kenny (1986), we tested this hypothesis using the average psychometric slope measured for each individual as a predictor, the WJ-BRS as an outcome measure, and the CTOPP phonological awareness score as a mediator. Using the R `mediate` package with 2000 bootstrapped simulations, the estimated indirect coefficient describing the mediation effect was significant ($\beta = 3.85$, SE $= 2.11$, $p = 0.011$). This analysis indicates that phonological awareness is a partial mediator (36.4% mediation) of the relationship between slope and WJ-BRS. However, when nonverbal IQ was included as a covariate, the estimated indirect coefficient did not meet the criterion of significance ($\beta = 1.6984$, SE $= 1.485$, $p = 0.071$), corresponding to a 25.5% mediation in the relationship between slope and WJ-BRS. Model comparison provided strong evidence for the inclusion of nonverbal IQ as a covariate ($F(1,41) = 9.66$, $p = 0.003$). Therefore, we did not observe clear evidence that phonological awareness mediates the relationship between slope and reading ability once nonverbal IQ is controlled for, but this effect deserves further investigation with a larger sample.

On the other hand, conducting this same analysis with the CTOPP Rapid Symbol Naming measure as the mediator between psychometric slope and reading ability indicated a larger mediation effect: even with the inclusion of nonverbal IQ as a covariate, this measure of rapid naming provided a 52.2% mediation of the relationship between slope and WJ-BRS ($\beta = 3.476$, SE $= 1.745$, $p = 0.021$).

One interpretation of these analyses is that, contrary to what has been argued in the literature (Boets et al., 2010; Joanisse et al., 2000; Manis et al., 1997), the phoneme categorization task we used is not in fact a reliable probe of phonological awareness (whereas the various test batteries are). The phoneme categorization task may more strongly reflect the contributions of rapid access to phoneme labels, not awareness of phonemes themselves. If, on the other hand, one accepts that (psychometric models of) children's performance on phoneme categorization tasks do reflect differences in phonological processing, then these results call into question the specificity of the component tests of these various diagnostic batteries.

2.4 Discussion

Many previous studies have claimed to show a relationship between dyslexia and poor categorical perception of speech phonemes (particularly for contrasts that rely on dynamic spectrotemporal cues), while others have suggested that the apparent auditory or linguistic processing impairments are in fact the result of general attention or memory deficits that manifest because of task difficulty. Our data unambiguously show a relationship between reading ability and phoneme categorization performance: higher reading scores were associated with steeper categorization functions, lower lapse rates, faster response times, and greater tendency to respond more quickly for unambiguous stimuli. But we also set out to answer two more specific questions: first, do poor readers struggle to categorize sounds on the basis of dynamic auditory cues to a greater extent than static cues? And second, are poor readers affected by task difficulty more than typically developing children? We also took care to show the ramifications of methodological choices when modeling categorization data as a psychometric function (particularly the common assumption of zero lapse rate). We discuss each of these points in turn.

It has been argued that dyslexics suffer primarily from an auditory processing deficit that specifically affects the processing of temporal information, leading to unique impairments in processing dynamic temporal and spectrotemporal cues such as formant transitions (Vandermosten et al., 2010, 2011) or modulations on the timescale of 2 to 20 Hz (Law et al., 2014). We found that poor readers showed, on average, less steep categorization functions regardless of whether they were categorizing on the basis of static or dynamic speech cues compared to strong readers. This was true regardless of whether reading ability was treated as a continuous variable, or if the Dyslexic children were compared to Below Average or Above Average reading groups. However, estimated lapse rate was related to cue type in our data (though in the opposite direction than what would be predicted from the literature: higher lapse rates were associated with the static cue continuum, not the dynamic one). These results run counter to the findings of Vandermosten and colleagues, who showed a greater deficit in categorization of speech continua involving dynamic cues compared to static cues (although note that they did not report reaction times or lapse rates, only psychometric slopes). We believe the key difference arises from our choice to equalize the duration of the cues in all stimulus tokens. Vandermosten and colleagues used a 100 ms dynamic cue in their /ba/∼/da/ continuum, and a 350 ms static cue in their /i/∼/y/ continuum, which raised the possibility that integration of sensory signals over time, not the static or dynamic nature of the speech cue, was the source of their observed difference.

One interpretation of our results is that categorizing phonemes may be more difficult when judgments must be made on the basis of very brief cues (whether static or dynamic), and that children with dyslexia are especially susceptible to this difficulty. However, most evidence suggests that slowing down speech does not make it more discriminable for reading and language impaired children (Bradlow et al., 1999; Stollman et al., 1994). Another possibility is that dyslexics have some impairment specifically in processing syllable-length sounds, as has been recently suggested (Goswami, 2011; Law et al., 2014; Lehongre et al., 2011), which cannot be ruled out based on our findings. What we can conclude is that modulations on the order of 10 Hz (the rate of formant transition changes in our /ba/∼/da/ continuum) do not pose a unique difficulty for poor readers, when compared to static speech cues of comparable duration.

In our data, estimates of psychometric slope and lapse rate were unrelated to whether the task was the more difficult ABX paradigm or the easier single-stimulus paradigm. Additionally, the covariates for ADHD diagnosis and nonverbal IQ were never shown to contribute to better model fits of the relationship between standardized reading ability and psychometric shape, a contribution we would have expected if task difficulty were a serious issue for our participants. This result, in conjunction with our findings that poor readers have less release from effort at the endpoints of the continua (measured by reaction time) and do not show signs of becoming distracted during the course of trial blocks any more so than the other participants, argue against interpreting the categorical perception deficit primarily as a manifestation of inattention or ADHD. Our results are therefore in agreement with another study that showed abnormal identification curves in dyslexic children both with and without ADHD (Breier et al., 2001). To the extent that task difficulty may contribute at all to performance differences, our results are more in line with the hypothesis that complex tasks burden working memory in poor readers rather than that they drive inattention.

Our PCA analysis of overall psychometric curve shape provides some evidence that poor readers were more affected by task difficulty than strong readers, and the correlations between various tests of reading ability and psychometric slope or lapse rate were nearly always greater in the ABX paradigm than in the single-stimulus paradigm (see Figure 2.6). However, the relationship between reading ability and both slope and lapse were best modeled without considering test paradigm, indicating that the deficit we observe in poor readers is not primarily driven by task difficulty. It is possible that our ABX and single-stimulus conditions were simply not different enough in difficulty to reveal an effect of task difficulty in our statistical models, despite the apparent difference suggested by the post-hoc correlations with test battery scores. Pertinent to that possibility is a recent meta-analysis of categorical perception studies among children with dyslexia (Noordenbos and Serniclaes, 2015), which found that there is generally a larger effect size reported for discrimination tasks than identification tasks. Indeed, during initial piloting of this study, subjects also performed same/different discrimination task with the same stimulus continua, but performance was so uniformly poor that the discrimination task was dropped from the final experiment. In any case, our findings reinforce claims that in order to fully understand what stimulus-dependent difficulties are faced by poor readers, the stimuli should be tested in more than one context to separate the effects of stimulus properties from the effects of task demands (Banai and Ahissar, 2006).

Our results did not provide clear evidence for the hypothesis that phonological awareness mediates the relationship between reading ability and the ability to label speech sounds categorically. However, our findings are in line with several other recent results. In a 2016 study of 49 dyslexic children and 86 controls, Hakvoort et al. performed a similar mediation analysis to test whether phonological awareness mediated the relationship between a measure of categorical perception and reading ability (Hakvoort et al., 2016). They did not find evidence for this relationship. Like us, they found that rapid automatic naming partially mediated the relationship between speech and reading measures. Additionally, another study of categorical perception deficits in dyslexic children found no correlation between phonological awareness and performance on phoneme discrimination and identification tasks (Robertson et al., 2009). Also relevant is a study which showed that dyslexics were both worse at perceiving speech in noise and worse at labeling phonemes, but did not find evidence that the categorical labeling mediates the relationship between speech-in-noise impairments and reading ability (Calcus et al., 2016). In sum, our study adds another piece of evidence to the view that the classic model relating speech perception to phonological awareness to reading ability may be too simplistic, insufficiently specified, or possibly even incorrect.

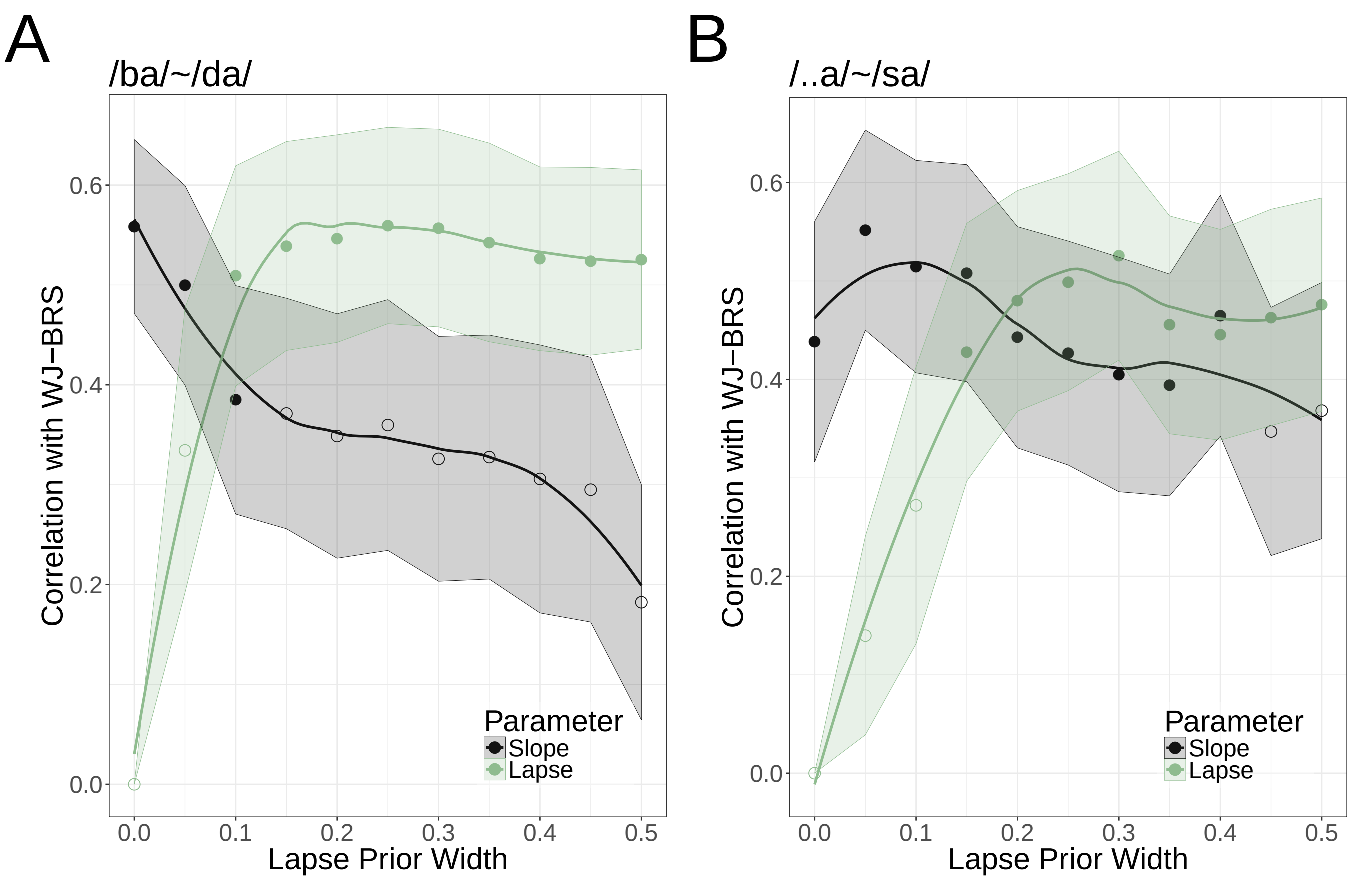
One important point that our study clarifies is the influence of methodological choices on psychometric fitting results. Specifically, we compared 4-parameter models of participants psychometric functions with traditional 2-parameter models that ignore lapse rate by fixing the asymptotes at zero and one. Our cross-validation of the asymptote prior distributions make it clear that a 2-parameter model is a worse fit to the data than any of the possible 4-parameter models we tested (see Materials & Methods for further details). More importantly, using a 2-parameter model in spite of its poor fit to the data leads to incorrect conclusions: 2-parameter models indicated that slopes were lower for the static cue continuum, whereas 4-parameter models indicated that slope was no different between the two continua, but lapse rate was higher on the static cue continuum. In either case, the main findinga clear relationship between reading ability and psychometric sloperemains. This is, to a degree, reassuring about previous conclusions drawn from 2-parameter models. Many prior studies are still probably correct in finding that children with dyslexia tend to have shallower categorization function slopes on particular types of stimuli. However, by failing to model lapse rate those studies may have overestimated the effect size when comparing children with dyslexia against typical readers.

To further illustrate this point, we explored how the conclusions of our study would have changed if we had changed the assumptions of our psychometric fitting routine. Specifically, we varied the width of the Bayesian priors governing the two asymptotic parameters from 0 (effectively fixed) to 0.5 in steps of 0.05 and observed the effects on the correlation of reading ability and slope. We fit psychometric functions for each subject at each of the eleven prior widths, and at each prior width, we computed the correlations between reading score and slope or lapse rate on the ABX trials. Figure 2.7 illustrates the result; the apparent correlation between reading ability and psychometric slope is greatest when the priors on the asymptotes are strongest, and the correlation diminishes when the asymptotes are allowed to vary more freely. Thus, it is critical to determine the optimal parameterization of the psychometric function before interpreting the resulting parameters of the fit.

What is still not clear, and is not within the scope of this study, is the extent to which low-level sensory abilities drive the relationship between phoneme categorization performance and low reading ability. Gap detection thresholds (the gold standard for measuring temporal resolution) have been shown to be normal in dyslexics (Adlard and Hazan, 1998; Boets et al., 2007; McAnally and Stein, 1997; Schulte-Körne et al., 1998), so poor auditory temporal resolution is not likely to be the underlying cause for the correlation between reading ability and categorization performance. There is some evidence that backward masking is more severe in dyslexics (Wright et al., 1997), which could potentially mask the consonant cue preceding the vowel in /ba/∼/da/ and /ʃa/∼/sa/ stimuli, and could explain the difference between our findings and those of Vandermosten and colleagues regarding static-cue stimuli. However, Rosen and Manganari (2001) showed that differences in backward masking did not affect categorical perception, so there is to date no clear relationship between any measured low-level deficit among poor readers and their ability to categorize speech sounds.

It has been long established that perceptual boundaries typically sharpen as a function of linguistic experience (Garnica, 1973), and categorical labeling of speech sounds in pre-lingual children has been shown to predict later success in reading (Bradley and Bryant, 1983). Thus, impairment on categorization tasks among children with dyslexia may represent the

Figure 2.7: Effects of altering priors in the psychometric fitting routine.



Correlations between slope, lapse rate, and reading ability (WJ-BRS score) as a function of the asymptotic prior width. Correlations are for psychometric functions fit to ABX trials. Error ribbons are computed by 10,000 bootstrap simulations. Filled dots indicate significant correlations ($p < 0.05$)

lower end of a distribution of literacy and categorical perception abilities that occurs in the general population, rather than a specific deficit that is unique to dyslexia (Shaywitz et al., 1992a). Our inclusion of children with a large range of reading abilities, rather than just two groups of children widely separated in literacy, suggested that there is no clear and distinct boundary between children with dyslexia and below-average readers that do not meet the criterion for reading disability. Our attempt to predict the group label for out-of-sample behavioral responses with QDA and LDA indicated that the groups were not easily separable by psychometric function shape. Although the mean slope and lapse parameters differed significantly by group, it was difficult to accurately predict the group an individual belonged to on the basis of their estimated slope or lapse. Moreover, variation in our models of children's perceptual boundaries was moderately related to reading ability, and the effect size and variability across our sample argues against interpreting the group-level effect as a unifying feature of dyslexia. Some poor readers had apparently normal psychometric functions, and there was considerable overlap in performance between poor readers in the Dyslexic group and merely Below Average readers. This aspect of our findings is in line with a multiple deficit model of dyslexia, whereby multiple underlying mechanisms, including auditory, visual and language deficits, all confer risk for reading difficulties (Joo et al., 2017a,b; Pennington, 2006; Peterson and Pennington, 2015a).

Our results invite several questions for further exploration. First, although we found that children with dyslexia showed poor performance categorizing speech on the basis of (100 ms) steady-state spectral cues, Vandermosten et al. showed that a large sample of dyslexic adults and children were not significantly impaired on a similar task with (350 ms) steady-state vowels. A natural extension of this work would be to parametrically vary cue duration, to look for a duration-dependent difference between readers with and without dyslexia. Such a difference could reflect a difference in sensory integration, which has been probed at the group level in dyslexics for visual stimuli, but not in relation to other sensory modalities (Talcott et al., 2000). Another question raised by our research is how directly auditory phoneme categorization relates to the common diagnostic measure phonological awareness. Although it has been hypothesized that poor phoneme categorization reflects fuzzy underlying phonological representations, which in turn mediate reading ability, our study showed only weak correlations between categorization function slope and phonological awareness, compared to the relationship between slope and a compound measure of reading skill. However, our mediation analysis did suggest that measures of phonological awareness and/or rapid naming may partially mediate the relationship between phoneme categorization and reading skill. Therefore, it is worthwhile to perform a more targeted analysis of the relationship between categorical perception of speech, phonological awareness and reading ability.

Chapter 3

CATEGORICAL PHONEME LABELING IN CHILDREN WITH DYSLEXIA DOES NOT DEPEND ON STIMULUS DURATION

This chapter is published as O'Brien et al. (2019) with co-authors Daniel McCloy and Jason Yeatman.

It is established that individuals with dyslexia are less consistent at auditory phoneme categorization than typical readers. One hypothesis attributes these differences in phoneme labeling to differences in auditory cue integration over time, suggesting that the performance of individuals with dyslexia would improve with longer exposure to informative phonetic cues. Here, the relationship between phoneme labeling and reading ability was investigated while manipulating the duration of steady-state auditory information available in a consonant-vowel syllable. Children with dyslexia obtained no more benefit from longer cues than did children with typical reading skills, suggesting that poor task performance is not explained by deficits in temporal integration or temporal sampling.

3.1 Introduction

A popular hypothesis is that dyslexia, a learning disability that affects the development of reading skills, is in fact the result of a subtle impairment in the way speech sounds are processed (Farmer and Klein, 1995; Van Ingelghem et al., 2005; Poelmans et al., 2011; Tallal, 1980). It is well-established that both adults and children with dyslexia tend to perform worse than their non-dyslexic peers on phoneme categorization tasks, in that they tend to be less consistent at labeling speech sounds in a stimulus continuum, even for the category exemplar sounds at the continuum endpoints (Brandt and Rosen, 1980; Hakvoort et al., 2016; O'Brien et al., 2018; Serniclaes et al., 2004; Zhang et al., 2012). This yields psychometric models of their performance that exhibit shallower slope than models of typical readers' performance. Across studies employing a variety of experimental paradigms, moderate effect sizes on psychometric slope are typically found (estimated to be 0.66 standard deviations on average by Noordenbos and Serniclaes (2015)), suggesting a reproducible group difference albeit with considerable overlap between dyslexic and control groups.

It has been posited by several authors that individuals with dyslexia are specifically impaired at temporal processing, which has been argued to affect the processing of brief sounds (Gaab et al., 2007; Lehongre et al., 2011; Richardson et al., 2004; Tallal, 1980) and the processing of rapidly changing acoustic cues like formant transitions (Vandermosten et al., 2010, 2011).There is mixed experimental evidence from auditory psychophysics to support this temporal processing hypothesis: while there is little evidence for impairments in gap detection, significant group differences have been somewhat reliably found for amplitude modulation detection, rise time detection, and duration discrimination (for a meta-analysis, see Hämäläinen et al. (2013)). However, the details of these studies are often at odds with one another: while some studies have found differences in slow-rate ($<$ 5 Hz) amplitude modulation detection (Rocheron et al., 2002; Stuart et al., 2006), other studies have not (Amitay et al., 2002; Rocheron et al., 2002; Witton et al., 2002). There are similarly contradictory findings for faster modulation rates: while some studies report group differences in detection of modulation rates $>$ 100 Hz (Lorenzi et al., 2000; Menell et al., 1999; Rocheron et al., 2002), this is not a consensus finding (Amitay et al., 2002; Stuart et al., 2006). Furthermore, a meta-analysis of effect sizes associated with psychophysical studies of basic auditory processing in indvidiuals with dyslexia suggested that none of the temporal measures considered—amplitude modulation, duration discrimination, or rise time detection—was associated with more than 52% non-overlap between control and dyslexic groups (Hämäläinen et al., 2013). As such, the literature does not clearly support a specific temporal processing deficit in the majority of individuals with dyslexia.

In previous work, we investigated the possibility that the abnormal performance on phoneme labeling tasks seen in individuals with dyslexia was driven by the presence of dynamic speech cues (O'Brien et al., 2018). We replicated the study design of Vandermosten and colleagues (2010; 2011) comparing identification performance on two different stimulus continua: one featuring a steady-state spectral envelope cue and the otehr containing a dynamic formant transition cue. In the Vandermosten group's studies, a cohort of individuals with dyslexia produced shallower psychometric functions (on average) than a control group when asked to label sounds that varied along the formant transition continuum, but not when they labeled sounds on the steady-state continuum. In our study, by contrast, individuals with dyslexia showed reduced ability to categorically label speech sounds along both continua. One potentially important difference between the two studies is that in ours, the duration of the informative cues was equated (i.e., the formant transition and spectral envelope cue were both present for 100 ms), whereas in Vandermosten and colleagues' original study they were not (i.e., the formant transition lasted 100 ms and the spectral envelope cue was present for 350 ms).

Our results indicated that individuals with dyslexia are not specifically impaired at categorizing speech sounds on the basis of dynamic auditory information—when presented with brief fricatives that differed based only on spectral envelope, our participants with dyslexia showed a similar degree of impairment as when they were labeling stop consonants. However, the discrepancy between the results of our study and Vandermosten and colleagues' results for steady-state cues raised the possibility that many individuals with dyslexia have impaired categorization on the basis of any "brief" cue. This would be consistent with the hypothesis of Tallal and colleagues, who proposed that individuals with dyslexia perform poorly whenever rapid auditory processing is involved (Merzenich et al., 1996; Tallal et al., 1996). It would also be consistent with the hypothesis of Goswami and colleagues, who proposed that individuals with dyslexia irregularly sample auditory information (Goswami, 2011). If insufficient sampling were the primary problem, then allowing a listener to acquire more samples by providing a longer exposure to the cue should reduce the performance gap between individuals with dyslexia and typical readers. Agnostic of these two temporal processing hypotheses, it is also conceivable that some participants with dyslexia weight sensory information differently than their typically developing peers (Nittrouer, 1999), and require more evidence acquired over more "looks" at a signal to make consistent category judgments.

To investigate these possibilites, we sought to determine whether longer exposure to steady-state acoustic cues in the context of a phoneme identification task benefits children with dyslexia more than children with typical reading skills. Note that the definition of dylsexia varies throughout the literature, and the majority of studies on auditory processing in this population employ a group-level design (Calcus et al., 2016; Law et al., 2014; Ramus, 2003; Talcott et al., 2000). Typically, a sample of individuals with low standardized reading scores and/or a diagnosis of dyslexia are compared to a sample of individuals with average or better standardized reading scores, yielding two groups that are well separated in terms of reading skill. While we wish to maintain parity with existing studies for ease of comparison, we believe that treating reading skill as a continuous variable provides a more insightful analysis for several reasons. First, previous work (including our own) suggests that the cutoff between individuals with dyslexia and below-average readers is arbitrary (O'Brien et al., 2018; Shaywitz et al., 1992a). Second, several auditory measures, including measures of categorical labeling, have been shown to vary continuously across a range of reading skills (Goswami et al., 2002; O'Brien et al., 2018; Vandermosten et al., 2010).

As such, we present two kinds of analyses in the present study: first, we examine how measures of auditory perception are related to reading skill as a continuous variable, including readers who score below-average on literacy assessments but do not meet our criteria for inclusion in the dyslexia group. Second, for continuity with previous studies, we compare individuals with dyslexia (defined as reading scores more than 1 standard deviation below the population mean) with a well-separated control group exhibiting reading skills at or above the population mean.

As in our earlier study, we employed a fricative continuum that spans from [s] to [ʃ]. We tested two conditions: one in which the steady-state fricative is available to listeners for 100 ms (short cue duration), and one in which it is present for 300 ms (long cue duration). We chose these durations because our previous results indicated that participants with dyslexia had difficulty categorizing 100-ms-long fricatives, while Vandermosten and colleagues had shown that participants with dyslexia had no difficult categorizing 350-ms-long steady-state vowels. As before, we recruited children with a range of reading abilities to participate.

3.2 Methods

3.2.1 Participants

A total of 83 native English-speaking children ages 8-12 were recruited for the study. Children without known auditory disorders were recruited from a database of volunteers in the Seattle area (University of Washington Reading & Dyslexia Research Database). Parents and/or legal guardians of all participants provided written informed consent under a protocol that was approved by the University of Washington Institutional Review Board. All subjects demonstrated normal or corrected-to-normal vision. Participants were tested on a battery of cognitive and literacy assessments, including the Woodcock-Johnson IV (WJ-IV) Letter-Word Identification and Word Attack subtests (Schrank et al., 2014), the Test of Word Reading Efficiency (TOWRE 2; Torgesen et al. (2011)) and the Weschlerl Abbreviated Scale of Intelligence (WASI-II; Wechsler (2011)). All participants underwent a hearing screening to ensure pure tone detection at octave frequencies between 500 Hz and 8000 Hz in both ears at 25 dB HL or better.

One subject who was initially recruited did not pass the hearing screening and was not entered into the study, and seven others did not meet the inclusion criterion for IQ (performance no less than 2 standard deviations below the population mean on the age-normed WASI-II FS-2 and nonverbal IQ measures). One subject was unable to complete training for the experimental task and so did not participate further. Another two subjects yielded data that could not be well fit with a psychometric function and were excluded from further analysis on the basis of low confidence in the psychometric parameter estimates, thus yielding data from 72 children in total. The guidelines for exclusion are described in greater detail in the following sections.

3.2.2 Demographics

In order to understand the relationship between phoneme categorization ability and reading ability, we selected our cohort of participants to span the continuum from impaired to highly skilled readers. Although we treat reading ability as a continuus measure in our statistical analyses, for the purpose of recruitment and data visualization we defined three groups based on the composite Woodcock-Johnson Basic Reading Score (WJ-BRS) and the TOWRE Index. For many of our subjects, standardized reading scores from multiple recent visits (within the past 14 months) to the lab were available. None of these children had participated in reading interventions or specialized training programs, so in order to gain the most stable measures of reading skill we averaged over available scores. As both the WJ-BRS and the TOWRE index are scored on the same standardized scale, a composite reading skill measure was created by averaging the two metrics for each participant. Using a composite of both measures as our criterion improved the confidence of our group assignments since they are highly correlated measures ($r = 0.91, p < 0.001$). The "Dyslexic" group comprised participants whose reading score fell 1 standard deviation or more below the mean (standardized score of 100); Above Average readers were defined as those with scores above the population mean.

Several individuals fell between these categories, and we have labeled them as "Below Average" in our dataset. These category definitions are consistent with our previous work (O'Brien et al., 2018) and similar to the definitions used by others (Rimrodt et al., 2010; Shaywitz et al., 2002). While we have included the complete data for theese individuals in our online data release, we opted to focus our statistical group comparisons on the Dyslexic and Above Average groups for ease of comparison with the broader dyslexia literature, where it is typical to compare to well-separated groups. Of the 17 individuals who fell into the Below Average group, 9 had parental reports of a dyslexia diagnosis. Thus, this group largely consists of individuals who at one point may have had reading difficulties, but have since received sufficient remediation to reach an age-typical level.

There were 36 subjects in the Dyslexic group (17 male), 17 in the Below average group (8 male), and 19 in the Above Average group (9 male). There were no significant differences in age between groups (Kruskal-Wallis rank sum test, $H(2) = 2.28, p = 0.32$), nor was there a significant correlation between age and reading score ($r = -0.009, p = 0.57$). We did not exclude participants with ADHD diagnoses from the study because ADHD is highly comorbid with dyslexia (Germanò et al., 2010). We expect this inclusion leads to a more representative sample of children with dyslexia. However, we did account for the presence of ADHD diagnosis in our statistical models. Of our 72 participants, 20 had a formal diagnosis of ADHD: four in the Above Average group, five in the Below Average group, and eleven in the Dyslexic group. The difference in the prevalence of ADHD across groups was not significant (Kruskal-Wallis rank sum test, $H(2) = 0.71, p = 0.70$).

Table 3.1 shows group comparisons on measures of reading and cognitive skills. We observed that IQ, both full-scale (incorporating verbal IQ) and nonverbal measures, varied as a function of group. Importantly, by design no subjects had a full-scale or nonverbal IQ falling more than 2 standard deviations below the population mean as measured by the WASI-II. Therefore, although there was a significant difference in IQ scores across groups, we were not concerned that abnormally low cognitive ability would prevent any child from performing the experimental task. To be certain our results were not confounded by this difference, we also included nonverbal IQ as a covariate in our statistical analyses to confirm the specificity of the relationship with reading skills as opposed to IQ.

3.2.3 Stimuli

Two 7-step speech continua were created using Praat version 6.0.37(Boersma and Weenink, 2019). Both were /ʃa/∼/sa/ continua, with the sole difference being the duration of the initial fricative. The fricative duration was either 100 or 300 ms. This choice of continuum was motivated by our previous study (O'Brien et al., 2018), in which participants with dyslexia behaved less categorically than above average readers when labeling /ba/∼/da/ and /ʃa/∼/sa/ continua with 100-ms-long consonant cues.

Table 3.1: Summary statistics and group differences on various demographic and behavioral measures.

	Above Average	Below Average	Dyslexic	Significance: Above Avg. v. Dys.	Significance: Below Avg. v. Dys.
	n =19 9♂, 10♀	n =17 8♂, 9♀	n =36 17♂, 19♀		
WASI-II					
FS-2	121.3 (16.2)	107.8 (14.8)	95.4 (9.2)	0.003	< 0.001
Nonverbal IQ	60 (11.8)	52.1 (7.7)	46.4 (6.3)	0.023	< 0.001
Woodcock-Johnson IV					
Basic Reading Score	113.5 (9.4)	95.8 (3.4)	77.5 (10.6)	< 0.001	< 0.001
Nonword	112 (9.2)	97.6 (6.6)	83.2 (10.8)	< 0.001	< 0.001
Real Word	113.1 (10.5)	94.7 (5.3)	74.3 (12)	< 0.001	< 0.001
TOWRE 2					
TOWRE Index	108.6 (9.4)	92.2 (6.9)	68.8 (8.3)	< 0.001	< 0.001
Nonword	104.5 (7.8)	90.7 (5.8)	70.9 (7)	< 0.001	< 0.001
Real Word	111.8 (11.8)	94.5 (9.1)	69.9 (10.7)	< 0.001	< 0.001
CTOPP 2					
Phonological Awareness	102.2 (14.2)	91.4 (12.7)	84.2 (11.3)	0.035	< 0.001
Phonological Memory	99.6 (19.8)	91.2 (13.8)	83 (11.8)	0.063	0.001
Rapid Naming	97.5 (13.6)	92.3 (10.3)	78.6 (10.3)	< 0.001	< 0.001

In order to test the hypothesis that increased cue duration would assist our dyslexic participants in the labeling task (as predicted by the temporal processing deficit hypothesis), it was important to include at least one condition in which we expected dyslexic participants to behave less categorically than control participants, so we replicated the 100 ms /ʃa/∼/sa/ condition from our prior work, and introduced a 300 ms condition as well.

The /ʃa/∼/sa/ continua were created by splicing synthesized fricatives onto a natural /a/ token excised from a spoken /sa/ syllable. The duration of /a/ was scaled to 250 ms using Praat's implementation of the PSOLA algorithm. Synthesized fricatives contained three spectral peaks centered at 2500, 3500, and 6500 Hz. The bandwidths and amplitudes of the spectral peaks were linearly interpolated between continuum endpoints in seven steps, and the resulting spectra were used to filter white noise. To improve the naturalness of the synthesized fricatives, a gentle cosine on- and off-ramp was imposed on the fricative envelope. The on-ramp lasted 75 ms in the short fricative and 225 ms in the long fricative, and the off-ramp lasted 20 ms in the short fricative and 60 ms in the long fricative. As such, the duration of the ramps was scaled proportional to the entire duration of the fricative, with the rise occurring in the first 75% of the fricative and the fall in the last 20%. Aside from this onset/offset ramping (which was applied equally to all continuum steps), the contrastive cue (the amplitudes and bandwidths of the spectral peaks) was steady throughout the duration of each fricative.

3.2.4 Procedure

Stimulus presentation and participant response collection were managed with PsychToolbox for MATLAB (Brainard, 1997). Auditory stimuli were presented at 75 dB SPL via circumaural headphones (Sennheiser HD 600). Children were trained to associate sounds from the two speech continua with animal cartoon characters (pink and purple snakes) on the right and left sides of the screen, and to indicate their answers with right or left arrow keypresses. Throughout all blocks, each cartoon was always associated with the same stimulus endpoint. After every 35 stimulus presentations, a reminder was displayed illustrating the snake associated with each sound.

Practice rounds were administered before the first test block. In practice rounds, participants were asked to categorize only endpoint stimuli and were given feedback on every trial. Participants had to score at least 75% correct on the practice round to advance to the experiment and were allowed to repeat the practice blocks up to three times. As mentioned, one child from the initial recruitment (belonging to the Below Average group) did not meet this criterion and was not included in the study.

In each test block, participants heard a single syllable and decided which category it belonged to by selecting an animal (no text labels were used). Generally, this was not difficult for our participants. Three subjects (one Above Average and two Dyslexic) lost track of the animals associated with the endpoints after succeeding at the practice rounds. In these cases, the experimenter provided written labels taped to the computer monitor to assist. Each block contained 5 presentations of each step on the continuum, for a total of 35 randomly ordered trials.

Six blocks were administered in total. In each block, only stimuli with short (100 ms) or long (300 ms) fricatives were presented (i.e., the stimulus duration was alternated on each block). The order of presentation was counterbalanced across participants such that half began with the 100 ms fricative continuum and half began with the 300 ms fricative continuum.

3.2.5 Psychometric curve fitting

Modeling of response data was performed with Psignifit 4.0, a MATLAB toolbox that implements Bayesian inference to fit psychometric functions (Schütt et al., 2015). We fit a logistic curve with four parameters, modeling the upper and lower asymptotes, the width of the logistic function, and the threshold. The width of the logistic function was transformed to the slope at the threshold value to give a measure of psychometric function slope.

In the Bayesian framework, each of these four parameters is estimated based on experimental observations weighted by a prior distribution. A prior distribution, or prior, defines the range over which a parameter could potentially vary. Incorporating this type of prior expectation into the model fitting procedure yields more stable estimates of model parameters, particularly in the case where parameters can be correlated (i.e., slope and asymptote). However, in the case of fitting sigmoidal functions, it is important not to use overly broad or narrow priors for the two asymptotic parameters, as this may lead to biases in slope estimates (Schütt et al., 2015).

To determine the most appropriate priors for fitting the two asymptotic parameters, we utilized a previously published approach (O'Brien et al., 2018). Rather than search for the best choice of priors for each asymptote separately, we assumed that the two asymptotes are determined by a common lapse rate. A lapse rate of 10% means that the lower and upper asymptotic parameters will be 0.1 and 0.9, respectively. Four-parameter fits of each experimental block were performed with each of 7 possible priors: a prior that fixed the lapse rate at zero (in line with the approach in most previous studies of auditory processing in dyslexia), and six uniform distribution priors with lower bounds of zero and upper bounds ranging from 5% to 30% in steps of 5%. Next, the optimal prior was chosen using ten-fold cross-validation. For each psychometric function measured for each participant in each condition (long and short fricative), psychometric curves were fit to 90% of the data, and then the summed log likelihood of the held-out 10% of data points was calculated. This process was repeated ten times, once for each unique 10% of the data, for each of the possible prior widths.

The estimated likelihoods of the held-out data points were used as a goodness-of-fit metric and pooled across blocks and cross-validation runs to determine the median likelihood for each prior width. The goodness-of-fit metric was normalized for comparison across participants by subtracting out each individual's median likelihood with a two-parameter model. Using the standard error rule (selecting the most restrictive prior width at which the standard error does not encompass 0), the optimal prior was determined to be a maximum lapse rate of 15%. As expected, the psychometric fits had the poorest fit to the data when the asymptotes were fixed at 0 and 1 (as in the two-parameter model).

Once we determined the optimal prior for the asymptote parameters, models were re-fit on the full dataset for each combination of participant and cue continuum to obtain final estimates of the four parameters. Any cases where the best-fit threshold parameter was not in the range of the stimulus continuum steps (1 through 7) were excluded from further analysis. Additionally, one subject in the Dyslexic group performed the labeling consistently backwards and produced a sharp psychometric function in the reverse direction, despite repeated reminders about the sound associated with each animal throughout the experiment. The psychometric functions were particularly ill-fitting for this subject and were removed from further analysis. Once the psychometric functions had been fit, data from 72 subjects were available for statistical analysis.

3.2.6 Statistical analysis of parameter estimates

After we fit psychometric functions for each subject in each condition, we used a series of generalized linear mixed models to determine the relationship between reading ability, stimulus duration, and three dependent measures. The first dependent measure was the psychometric slope. The second dependent measure was the average offset of the upper and lower asymptotes of the function (i.e., their deviations from 0 or 1, respectively). For the third dependent measure, we were interested in a composite measure of psychometric function shape incorporating all four of its parameters (slope, threshold, upper asymptote, and lower asymptote). We derived this composite measure by performing principal components analysis on these parameters for each psychometric function collected in the study, transforming each psychometric function to a single variable according to the linear weights prescribed by the first principal component (PC1).

We were motivated to consider the PC1 measure of psychometric function shape because of a well-known challenge in interpreting psychometric function parameters: it is difficult (if not impossible) to fit both the slope and asymptotic parameters without incurring bias in one domain due to these parameters trading relationship in the optimization space of sigmoidal functions (Treutwein and Strasburger, 1999). Moreover, data that are best fit with non-zero asymptotes are often treated as indicative of lapses of attention, but in fact may (in part) reflect aspects of the deficit under study (O'Brien et al., 2018; Serniclaes et al., 2004). Therefore, ideally, we would model slope and asymptotes simultaneously using a single measure of the psychometric shape that takes all parameters into account namely, the PC1. We have previously shown this measure to be an effective predictor of reading ability and potentially a more sensitive probe than either slope or asymptotic parameters alone (O'Brien et al., 2018).

For each dependent measure (slope, asymptote, and PC1), fixed-effect predictors with sum coding were used for the continuum (short versus long fricative duration) variable (i.e., the categorical predictors were represented as 0.5 and -0.5). Reading ability (the average of the WJ-BRS and TOWRE Index) was entered as a continuous fixed-effect predictor except where otherwise stated. Additional predictors were added for presence/absence of ADHD diagnosis (treatment coding) and for non-verbal IQ (WASI-II Matrix Reasoning score; continuous predictor). A random intercept for participant was also included.

For all model analyses, we began with a fully specified model of reading score as a function of the parameters of interest (slope, asymptote, or PC1) plus the two covariates (ADHD and non-verbal IQ) and a random intercept for subject identity. All model fitting was done using the `lme4` library for R (Bates et al., 2016). The contributions of the covariates were first tested using parametric bootstrapping using the `pbkrtest` library for R (Halekoh and Højsgaard, 2014), which is robust to non-normally distributed residuals. Model terms were retained if the bootstrapped p-value of the coefficient being nonzero was less than 0.1. In all cases, the covariates failed this test and were dropped from the model. After testing the covariates, we tested the terms of interest for the study (duration, reading ability, and the interaction between them) using the same approach.

Table 3.2: Selected model of slope

	β	SE	p
(Intercept)	1.627	0.121	< 0.001
Reading skill	0.020	0.007	0.003
Duration	0.504	0.111	< 0.001

3.3 Results

3.3.1 Relationship between phoneme categorization and reading skill

We found an association between reading skill and several aspects of psychometric function shape, shown in Figure 3.1. Slope, asymptote, and individual PC1 measures are all shown with their correlations to reading ability. Two columns show how these psychometric function parameters are correlated with reading skill at each fricative duration, 100 ms or 300 ms. Upon visual inspection, it is clear that the relationship between task performance and reading skill is largely independent of fricative duration; we confirmed this with generalized linear mixed model analysis.

First, we examined the relationship between reading ability and psychometric slope. After model selection, the most parsimonious model of psychometric slope contained a continuous predictor for reading ability (average reading score on WJ-BRS and TOWRE), a main effect of stimulus duration, and a random intercept for each participant (Table 3.2). The interaction between reading score and stimulus duration was not significant and did not survive the model selection procedure. Reading score and stimulus duration were both significant predictors of psychometric slope.

We next considered asymptote as the dependent variable. The same initial model specification and simplification procedure used in the model predicting psychometric slope was also used for the model predicting the average asymptote parameter. The final model of asymptote contained a significant predictor for reading ability (Table 3.3); there was no

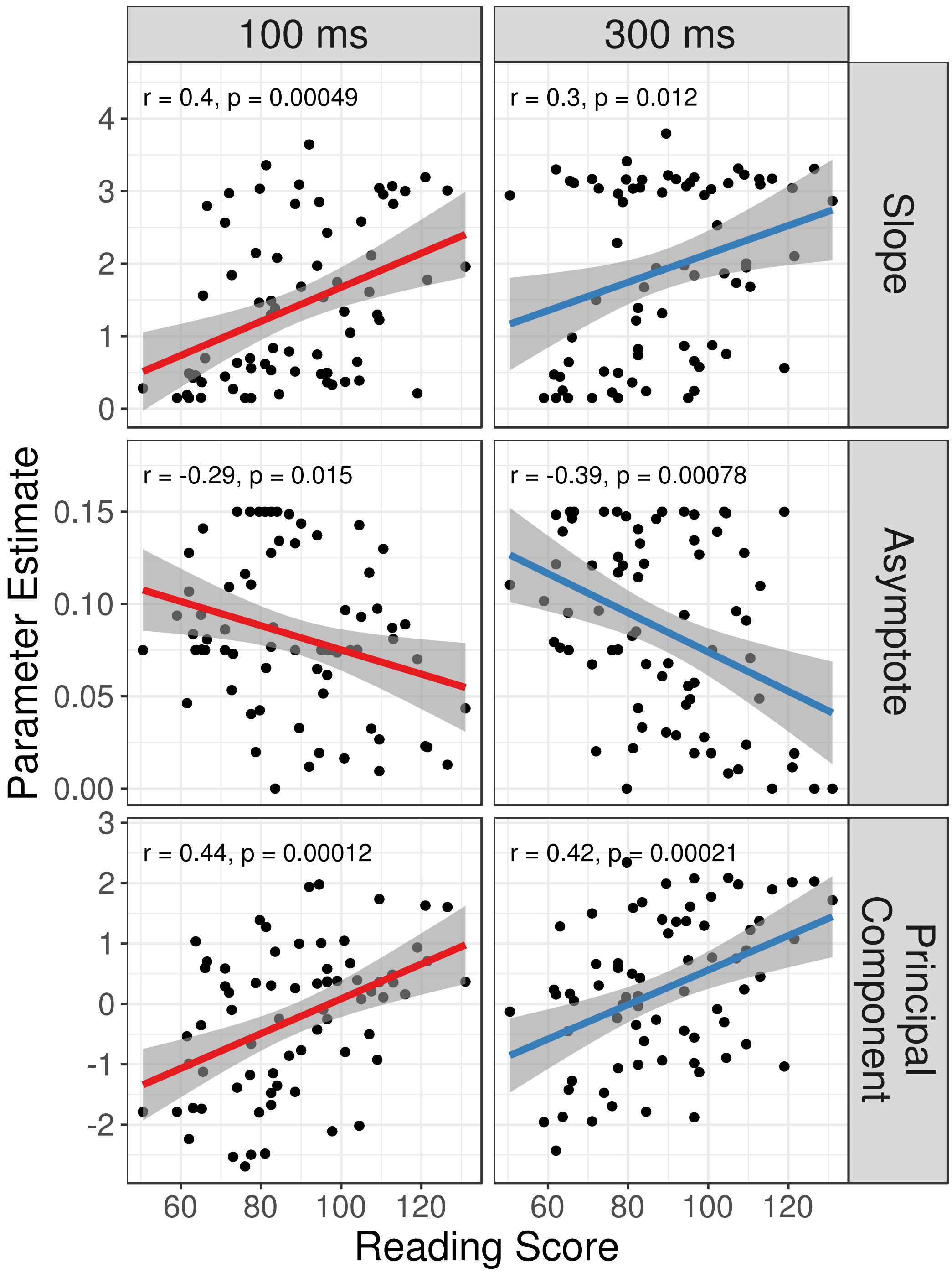


Figure 3.1: Plots of model parameter estimates versus reading score.

Each point corresponds to parameter estimates for one subject in one condition (100 ms or 300 ms fricative duration). Lines indicate the best fit regression line with 95% confidence intervals in shaded regions

Table 3.3: Selected model of lapse rate

	β	SE	p
(Intercept)	0.085	0.004	< 0.001
Reading skill	-0.001	0.000	0.002

Table 3.4: Selected model of PC1

	β	SE	p
(Intercept)	-0.016	0.110	0.88
Reading skill	0.027	0.006	< 0.001
Duration	0.509	0.157	0.002

evidence for a significant main effect of duration or an interaction between duration and reading ability. Higher reading ability was associated with smaller asymptotic parameters (about 0.08% smaller per 1-unit increase of reading score). Thus, compared to strong readers, poor readers showed shallower psychometric slopes and were also more likely to label stimuli inconsistently even near the continuum endpoints. On average, subjects showed a steeper psychometric function when provided a longer cue irrespective of their reading ability. There was no evidence to suggest that the degree of within-category consistency (asymptote) depended on the cue duration (short or long fricative).

Finally, we investigated how the PC1 composite measure of psychometric function shape varied with reading skill and stimulus duration. This component explained 37.8% of the variance in our psychometric function data, reflecting a linear combination of the slope, threshold, and both asymptotes (slope: 0.543; threshold: 0.542; lower asymptote: -0.636; upper asymptote: -0.086). The most parsimonious generalized linear mixed model of PC1 contained predictors for reading skill and fricative duration, but not the interaction between these two variables (Table 3.4). Thus, reading skill and fricative duration were significant predictors of psychometric function shape as summarized by PC1.

While our model selection procedure led us to drop nonverbal IQ from all our models so far, the fact that nonverbal IQ differed significantly among reading groups suggested a need for caution in discarding this covariate. To be sure that nonverbal IQ was not a meaningful predictor of task performance separate from reading skill, we modeled psychometric function shape as a function of nonverbal IQ, now with reading as the covariate. Here, we used PC1 as the outcome measure because PC1 yielded stronger correlations with reading skill than either slope or asymptote measures (see Figure 1). In this model, nonverbal IQ was not a significant predictor of PC1 ($\beta = 0.018$, SE $= 0.013$, $p = 0.177$), but reading skill was ($\beta = 0.023$, SE $= 0.008$, $p = 0.004$). This finding, combined with the outcome of our model selection process, suggests that the observed relationship between reading skill and phoneme categorization task performance could not be explained in terms of our nonverbal IQ measure.

3.3.2 Group differences in phoneme categorization

As discussed in the Introduction, it is conventional in dyslexia research to analyze group-level differences in outcome measures. Therefore, to supplement the findings relating reading score to phoneme categorization, we also conduct a series of group comparisons in order to compare effect sizes to the broader literature on phoneme labeling in people with dyslexia. Because we wish to provide a comparison of two groups that are clearly separated in terms of reading ability, these analyses were carried out using data from our subjects in the Dyslexic and Above Average Reader groups (omitting the Below Average group, see Table 3.1).

Using a linear mixed effects model with group as a categorical predictor (Above Average as the reference group), subject as a random factor, and following the same model selection procedure as before, a model containing main effects of group and duration was selected. In this reduced model, mixed-effect ANOVA tests indicated that both main effects were significant predictors of slope (group: $F(1, 50.79) = 7.26, p = 0.010$, degrees of freedom estimated via the Kenward-Rogers approximation; duration: $F(1, 65.85) = 14.96, p < 0.001$). The interaction between duration and group was eliminated by model selection. The estimated

Cohen's d for the separation of slope by group was 0.78, with a 95% confidence interval ranging from 0.19 to 1.38.

Next, we considered a model of average psychometric asymptote parameter. The most parsimonious model contained only a main effect of group, which was significant ($F(1, 51.04) = 6.54, p = 0.014$). Group separability by lapse rate was measured with Cohen's d, giving an effect size of 0.73 with a 95% confidence interval ranging from 0.14 to 1.32.

Lastly, we confirmed the relationship between group and psychometric PC1 ($F(1, 50.35) = 15.13, p < 0.001$). The selected model also included a significant main effect of duration ($F(1, 51.71) = 11.07, p = 0.001$). A candidate model including the interaction between duration and group indicated that this term was not significant ($F(1, 66.78) = 0.13, p = 0.725$). To assess the separability of the groups, we calculated Cohen's d for a group comparison where each individual's PC1 estimate is averaged across the two test conditions (combining long and short durations). There was a large effect size for the comparison between the Dyslexic and Above Average groups ($d = 1.1$, with the 95% confidence interval spanning from 0.50 to 1.72).

To maintain continuity with the broader dyslexia literature, we were primarily interested in comparing two well-separated groups of readers. However, for completeness, we confirmed that the main findings were not substantially altered by including the cohort of Below Average readers as a third comparison group. Mixed-effect ANOVAs showed that for slope, main effects of group and duration were significant (group: $F(2, 64.98) = 3.53, p = 0.035$; duration: $F(1, 77.32) = 20.65, p < 0.001$). For asymptote, there was a significant main effect only of group ($F(2, 65.23) = 3.73, p = 0.029$). For PC1, both group and duration were significant predictors (group: $F(2, 65.67) = 8.48, p < 0.001$; duration: $F(1, 66.53) = 10.38, p = 0.002$). All of these results are consistent with those of the two-group analyses.

3.4 Discussion

Our results replicate and extend previous findings about the nature of phoneme categorization deficits in struggling readers. We confirm that reading ability is moderately predictive of the extent to which a child applies consistent labels to repeated presentations of the stimuli. This effect occurs whether the child has relatively long or short exposure to the identifying phonetic cue, which was in this case the spectral envelope of a fricative consonant. Our findings rule out the possibility that the performance of children with dyslexia on the identification task is limited by a general difficulty processing brief auditory cues (Merzenich et al., 1996), as well as the more recently posed hypothesis that they do not properly sample or integrate auditory information at the phonetic-cue scale (Goswami, 2011).

Although we cannot make claims about temporal sampling or processing in all contexts, our results suggest that difficulty making categorical judgments about speech sounds is not primarily driven by temporal features of the stimulus. The benefit of increasing cue duration was roughly equal across the spectrum of reading abilities. If children with dyslexia were unable to form clear categories of speech sounds because they did not properly sample phonetic information (the spectral envelope), then we would have expected to find a significant interaction between stimulus duration and reading ability. This would occur because increasing the duration of phonetic cue availability would give children with dyslexia a greater chance of perceiving the cue, improving their performance on the labeling task relative to strong readers. Conversely, we might expect an interaction in the opposite direction if children with dyslexia were so profoundly impaired at processing the phonetic cue that these children derived little or no benefit from increased duration, while control participants did. In any case, the fact that we did not find evidence for such an interaction for any of the psychometric function parameters testedincluding slope, asymptote, and the first principal component of these measuresargues that the well-established difficulty with phoneme identification in children with dyslexia is unlikely to be a consequence of impaired temporal processing.

There are several limitations to our study that must be noted. First, the nature of the task confounds our ability to tease apart cognitive processes that are specifically auditory and those that are domain general. While we are surely testing some aspects of auditory processing, phoneme categorization tasks are well-known to be influenced by many non-sensory factors, including the manner in which the stimuli are presented and how participants are instructed (Cleary and Pisoni, 2001; Pisoni et al., 1982; Pisoni and Tash, 1974). As other researchers in the area of dyslexia have noted, typical sensory processing is not sufficient to ensure categorical behavior (Vandermosten et al., 2018), and while we have sought to make our task simple, we cannot wholly exclude the possibility that the cognitive load was greater for our participants with dyslexia. In particular, in order to avoid orthographic cues in the task, listeners were asked to choose between two differently-colored snakes without text labels. We cannot be sure that this kind of abstract categorization was not more difficult for children with dyslexia than associating sounds with text. While we are reassured that we did not see a strong relationship between nonverbal IQ and task performance after controlling for reading skill, our nonverbal IQ measure, performance on a matrix reasoning task, may not be closely related to the cognitive demands of the auditory task.

A second limitation to our study is that children with dyslexia are a remarkably heterogeneous group: there is moderate comorbidity of dyslexia with both ADHD and language impairments, including speech sound disorders and Specific Language Impairments (Germanò et al., 2010; Pennington, 2006; Stevenson et al., 2005). While none of our participants reported having been diagnosed with Specific Language Impairment, we cannot be sure that they would not have met the criteria for diagnosis at some point in their lives. It remains possible that the relationship between reading skill and task performance is influenced by the presence of other developmental language impairments that we do not explore here. Our study can only speak to the specific case of ADHD. Specifically, we tested the effects of ADHD and nonverbal IQ and found that they were not significant predictors of task performance once reading skill was accounted for.

Despite these limitations, several conclusions can be drawn from our results. First, our results are incompatible with the temporal sampling hypothesis, which contends that individuals with dyslexia are unable to efficiently extract acoustic cues at the phonetic timescale because they do not have proper neural entrainment to the speech envelope. In the temporal sampling framework, weak entrainment to the speech envelope means a listener will sample phonetic information at sub-optimal phases in the speech signal. If this were the case, then increasing the duration that phonetic information is available to children with dyslexia should improve their performance on the phoneme labeling task by affording them more chances to glimpse the spectral envelope cue in our stimuli, whereas typically-developing children would not be expected to improve much. However, our results suggest that children with dyslexia struggle to assign categorical labels regardless of how many samples of acoustic information may be available. Similarly, our results are at odds with the original formulation of the rapid temporal processing hypothesis (Tallal, 1980), which predicts that children with dyslexia will perform poorly on tasks with brief auditory cues but will perform more like typically developing children as auditory cue duration increases. We note that we did not test how perception changed with fricative durations exceeding 300 ms. While it is possible that perceptual effects could be observed by increasing the duration of phonetic cues into this range, both temporal theories of dyslexia (Goswami, 2011; Tallal et al., 1996) would still need to be amended to explain our current findings.

A second conclusion is that the relationship between reading skill and an index of categorical labelingthe shape of a psychometric functionis well-modeled by a linear function. In other words, there is no obvious discontinuity in task performance between children who meet the criteria for dyslexia and children who dont. We found a similar result in our previous study of categorical labeling in 44 school-aged children (O'Brien et al., 2018). This argues against interpreting children with dyslexia as a separate population from typical readers, at least with regards to the aspects of speech perception tested here. Although group-level statistics comparing a cohort of poor readers to unimpaired readers may reveal group differences, as we show here, the relationship between reading skill and our measure of categorical labeling is indeed continuous.

At this juncture, there have been nearly forty studies of phoneme identification with reading-disabled children (for a summary of the literature, see Noordenbos and Serniclaes (2015)). There is no clear consensus on stimulus features that induce or remedy struggling readers impairments with tasks that ostensibly probe categorical labeling: the deficit has been found for synthetic (Godfrey et al., 1981; Manis et al., 1997; Serniclaes et al., 2001) and naturalistic speech, and for spectral (Johnson et al., 2011; Nittrouer, 1999; O'Brien et al., 2018; Steffens et al., 1992), temporal (Bogliotti et al., 2008; Breier et al., 2001; Chiappe et al., 2001; Hazan et al., 2009), and spectrotemporal cues (Van Beinum et al., 2005; Blomert and Mitterer, 2004; Boets et al., 2011; Messaoud-Galusi et al., 2011; Mody et al., 1997). In the present study, we confirm that both brief and long stimulus presentations are difficult for children with dyslexia to categorize. Although some studies have shown group differences on certain continua but not others (Nittrouer, 1999; Vandermosten et al., 2010, 2011), no reliable pattern seems to emerge for a specific acoustic feature that explains difficulties with phoneme categorization in children with dyslexia.

To make sense of these disparate results, it is important to remember that categorical labeling is a general feature of sensory perception, and the task design we employed draws upon working memory, statistical learning capacity, and auditory attention (Ramus and Ahissar, 2012). There has been growing interest in non-sensory explanations for why children with dyslexia tend to perform unusually on many tests of auditory processing, including their ability to form categories (Gabay et al., 2015), act as ideal observers (Ahissar et al., 2006; Ziegler, 2008), and make use of statistics in sensory information (Gabay et al., 2015; Vandermosten et al., 2018). While we found no evidence that ADHD or nonverbal IQ explained task performance better than reading skill in our sample, these measures may not be effective probes of the cognitive processes involved in categorical learning and decision making.

A final reason that we are inclined to suggest further research into general features of sensory decision making in children with dyslexia, versus speech processing in particular, is that the relationship between the phoneme labeling task and speech perception in natural conditions may be complex. It is controversial whether categorical labeling is a necessary stage of analysis in speech processing (Cleary and Pisoni, 2001), and indeed there is evidence that non-categorical processing is part of word recognition in typical listeners (McMurray et al., 2003, 2009; Toscano et al., 2010). With few exceptions (e.g., McQueen et al. (1999)), there is little experimental evidence (that we know of) linking categorical perception at the scale of individual phonemes to speech understanding in naturalistic settings. Therefore, rather than treating phoneme identification tasks as probes of auditory acuity, or of speech processing in everyday environments, it may be more fruitful to turn our attention to the domain-general functions that are tapped by such laboratory tasks. For now, the evidence from our study (and many others) strongly suggests that differences in temporal processing cannot explain why some individuals with dyslexia behave differently in this experimental context.

Chapter 4

BRIDGING SENSORY AND LANGUAGE THEORIES OF DYSLEXIA: TOWARDS A MULTIFACTORIAL MODEL

The work below is currently in review at the journal Developmental Science as a manuscript with co-author Jason Yeatman.

Competing theories of dyslexia posit that reading disability arises from impaired sensory, phonological, or statistical learning mechanisms. Importantly, many theories posit that dyslexia reflects a cascade of impairments emanating from a core deficit. Here we collect a battery of psychophysical and language measures in 106 school-aged children to investigate whether dyslexia is best conceptualized under a core-deficit model, or as a disorder with heterogenous origins. Specifically, by capitalizing on the drift diffusion model to separate sensory encoding from task-related influences on performance in a visual motion discrimination experiment, we show that deficits in motion perception, decision making and phonological processing manifest largely independently. Based on statistical models of how variance in reading skill is parceled across measures of sensory encoding, phonological processing and decision-making, our results challenge the notion that a unifying deficit characterizes dyslexia. Instead, these findings indicate a model where reading skill is explained by several distinct, additive predictors, or risk factors, of reading (dis)ability.

4.1 Introduction

Recently, there has been growing adoption of the view that dyslexia, a reading disability, is probabilistic in nature: children with a family history of dyslexia are considered "at-risk", and compensatory skills such as strong oral language or executive functions may be "protective factors" (Haft et al., 2016; Hulme et al., 2015; Muter and Snowling, 2009; Pennington, 2006). In this multifactorial framework, most cases of dyslexia cannot be explained by a single cognitive deficit. Despite this heterogeneity, it is broadly accepted that phonological awareness (PA) and rapid automatized naming (RAN) are two of the strongest—if imperfect—predictors of reading development (Pennington et al., 2012; Wolf and Bowers, 2000).

In parallel, there is a broad literature characterizing dyslexia as the consequence of a fundamental deficit that supersedes phonological processing. There are many reports indicating that people with dyslexia also perform worse in experiments targeting various aspects of visual (Stuart et al., 2006; Talcott et al., 2002) and auditory processing (Hämäläinen et al., 2013; Noordenbos and Serniclaes, 2015), as well domain general mechanisms such as processing speed and statistical learning (Gabay et al., 2015; Vandermosten et al., 2018). These findings have spurred competing theories that explain dyslexia as the consequence of cascading effects from a fundamental sensory processing deficit. Generally, these cascading deficit theories contend that relatively low- or mid-level aspects of sensory processing disrupt phonological processing, and by this mechanism disrupt reading development.

Notably, these two branches of research remain largely distinct: while multifactorial models of reading disability are increasingly accepted among researchers studying high-level cognitive and linguistic functions, these models largely ignore lower-level deficits in sensory processing. In the sensory-processing literature, on the other hand, cascading deficit models continue to dominate and appeals to a core mechanism of dyslexia are still commonplace. Indeed, a PubMed search for the phrase core deficit of dyslexia turns up 118 results from 1986 to the present. Presently, hypotheses positing a core deficit with cascading effects are the focus of many neuroscientific and psychophysical studies of reading disability (Casini et al., 2018; Colling et al., 2017; Frey et al., 2019a,b; Gori et al., 2016; Krause, 2015; Lieder et al., 2019; Nicolson and Fawcett, 2018; Vidyasagar, 2019).

A core deficit model is inherently at odds with a multifactorial model; to accept both models implies that a deficit is not really core in the majority of individuals with dyslexia. Reconciling the many disparate theories of reading disability remains a formidable challenge.

To further compound the difficulty, there are several variants of the cascading deficit theory: one is the magnocellular deficit theory of dyslexia, in which a low-level impairment in the motion-sensitive magnocellular pathway of the visual system is said to disrupt reading skill development (Stein, 2001, 2018; Stein and Walsh, 1997). Proponents of this theory have argued that sensitivity to transient sensory information may not be restricted to vision, but could also affect auditory processing (Stein and Talcott, 1999; Van Ingelghem et al., 2001; Witton et al., 1998). Hypothetically, auditory insensitivity to rapid cues could diminish an individual's ability to learn the sounds of their language, and hence develop PA. A recent spin on this theory is the temporal processing hypothesis, which contends that, in fact, slow temporal mechanisms involved in entraining to the envelope of speech are the fundamental disorder (Casini et al., 2018; Goswami, 2015; Huss et al., 2011). Distinct from these sensory theories, proponents of the statistical-learning hypothesis argue that a domain-general deficit in sensory learning and perceptual decision-making could explain why people with dyslexia perform poorly on myriad psychophysical tasks (Ahissar, 2007; Nicolson and Fawcett, 2018; Ziegler, 2008). It also purports to explain why children with dyslexia struggle to learn the mapping between letters and sounds.

Today, the literature remains inconclusive for several reasons. First, the various cascading deficit models contradict one another as each posits distinct mechanisms for disrupting phonological processing. While a statistical learning model of dyslexia could potentially explain why so many struggling readers also perform poorly on visual psychophysics, it has not been established whether these two types of deficits occur in the same individuals. The widespread use of group-level statistics makes it challenging to interpret how many individuals show a given pattern of low-level deficits, and the few studies focusing on individual patterns across a battery of diverse tasks do not encourage much hope for a uniform profile (Amitay et al., 2002; Ho et al., 2002; Menghini et al., 2010; Ramus, 2003; White et al., 2006).

Perhaps more importantly, it remains challenging to understand what relationship predictors from psychophysical tasks have with phonological predictors in determining reading ability–in other words, whether the influence of low-level sensory processing mechanisms on reading skill is mediated by phonological processing. Perhaps Talcott et al. (2000) best addressed this question by administering auditory, visual, and phonological tasks to 32 readers, concluding that a measure of visual motion processing explained some additional variance in reading skill beyond a measure of PA. A follow-up study in more than 300 school-aged children replicated the finding that visual and auditory psychophysics explained variance in both phonological and literacy skills but did not clarify the fit of a cascading model (Talcott et al., 2000). Several others have observed evidence that psychophysical measures influence reading skill separate from the proposed phonological pathway (Snowling et al., 2019; Stein, 2001; White et al., 2006). Despite these findings, cascading deficit models remain at the forefront of the dyslexia debate, particularly for theories that hold a central role for sensory deficits (reviewed in Goswami, 2015).

There are several reasons why studies such as Talcott et al.'s are well-cited, but not broadly adopted as conclusive evidence about sensory processing in dyslexia. In the last two decades, there has been growing focus on non-sensory mechanisms that may affect how struggling readers perform on psychophysical tasks a confound that many studies may not have sufficiently accounted for (Banai and Ahissar, 2004; Ramus and Ahissar, 2012). Furthermore, in the multifactorial literature, it is increasingly accepted that at least two dissociable aspects of phonological processing (PA and RAN) contribute to reading skill (Pennington et al., 2012; Wolf and Bowers, 1999, 2000). Most literature explores the relationship of sensory measures to a single dimension of PA. As evidence mounts that PA alone is unlikely to explain many (or even most (Pennington et al., 2012)) cases of dyslexia, it remains worth considering how individual differences in visual motion processing, or perceptual decision making more generally, will fit into changing conceptions of reading disability.

Emanating from the rift in the literature, and the incompatibility of the myriad of core deficit models, this study investigates whether a cascading-deficit model, in which an underlying deficit in some other lower-level sensory or cognitive process disrupts phonological processing, is compatible with the pattern of behavioral and psychophysical results seen in a large sample of children with dyslexia. We also investigate to what extent phonological predictors alone (a phonological-core model) explain reading disability in our sample. The phonological-core model and cascading-deficit models all posit that there is essentially one factor that predicts reading ability. Under this model, other measures that correlate with reading skill would be proxies for that fundamental predictor. As such, these models predict a certain pattern of correlation between sensory and behavioral predictors: if visual motion processing is the fundamental impairment, for example, then non-sensory aspects of decision making should not explain additional variance in reading skill, and the relationship between psychophysical task performance and reading should be mediated by PA. If, on the other hand, those deficits were shown to be largely separate contributors to reading skill, that would imply that these deficits cannot be attributed to a single underlying impairment and provide evidence in favor of multiple deficit models of dyslexia.

In order to separate the contributions of sensory encoding of visual motion from non-sensory aspects of the decision-making process, we revisit a widely used measure of visual motion sensitivity (random dot motion discrimination) with a mathematical modeling approach. The drift diffusion model (DDM) estimates the generating function that corresponds to an individual's pattern of responses and reaction times on a task (Ratcliff and McKoon, 2008), and has been previously used to understand how cognitive mechanisms associated with aging (Ratcliff et al., 2004), ADHD (Huang-Pollock et al., 2017), and development (Ratcliff et al., 2012) manifest in psychophysical task performance. The model has been extensively used to describe decision-making on the motion discrimination task (Gold and Shadlen, 2007; Palmer et al., 2005; Shadlen et al., 2013), and many of its assumptions are validated by electrophysiological work in non-human primates (Shadlen and Newsome, 2001). As such, the DDM provides a rigorous way to explore the intersection of sensory integration and decision making in relation to reading skill.

Contrary to predictions of the myriad of core-deficit models, our data reveal a heterogeneity of deficits among children with dyslexia; no single factor, including measures of phonological processing, can reliably distinguish children with dyslexia from control subjects. The DDM reveals that sensory encoding and perceptual decision making are separable factors which predict unique variance in reading skill above and beyond phonological processing, and that there is not a consistent pattern of impairments among children with dyslexia. Furthermore, these sensory predictors are useful in addition to PA and RAN in characterizing an individual's level of reading (dis)abilities. As a whole, these data provide further evidence against core-deficit models of dyslexia and indicate that multiple-deficit models must consider the combined influence of sensory, cognitive and linguistic factors on the development of reading skills.

4.2 Methods

4.2.1 Participants

A total of 119 native English-speaking school-aged children ages 8-12 were recruited for the study. Children without histories of neurological or sensory disorders were recruited from a database of volunteers in the Seattle area (University of Washington Reading & Dyslexia Research Database). Parents and/or legal guardians of all participants provided written informed consent under a protocol that was approved by the University of Washington Institutional Review Board. All subjects demonstrated normal or corrected-to-normal vision. Participants were tested on a battery of cognitive and literacy assessments, including the Woodcock-Johnson IV (WJ-IV) Letter Word Identification and Word Attack subtests, the Test of Word Reading Efficiency (TOWRE-2), Comprehensive Test of Phonological Processing (CTOPP-2) and the Weschler Abbreviated Scale of Intelligence (WASI-II). All subjects had normal or corrected-to-normal vision.

Five subjects did not complete the psychophysics. An additional two subjects did not show evidence of performing above chance (greater than 60.5% accuracy at any of the four stimulus coherence levels) and were excluded from analysis. A further six subjects did not produce enough usable data to fit the DDM (no more than 15% responses outside of the acceptable response time window from 200 ms to 10 s). This left 106 subjects with usable data.

4.2.2 Demographics

We recruited participants whose reading abilities ranged from profoundly impaired to highly proficient. Since reading abilities fall on a continuum, we treat reading ability as a continuous measure in our main statistical analyses. For the purpose of comparison with other studies we include group-level analyses (Dyslexic versus Control) in our Supplementary Materials (Tables B.7, B.8 and B.9). Group labels were assigned on the basis of the composite Woodcock-Johnson Basic Reading Score (WJ-BRS) and TOWRE Index. As both the WJ-BRS and TOWRE Index are scored on the same standardized scale, a composite reading skill measure was created by averaging the two scores for each participant. The Dyslexic group comprised participants whose reading score fell 1 standard deviation or more below the population mean (reading score < 85); the Control group had reading skill measures above this cutoff and had never been diagnosed with a reading disability. There were 43 subjects in the Dyslexic group and 48 in the Control group. A remaining 15 subjects were not well-described by either label (e.g., reading score> 85 but an indication of a dyslexia diagnosis) so were not included in the group comparisons. As in previous work by our group and others 30,66 , we did not IQ-match these groups, but rather controlled for nonverbal-IQ explicitly in our statistical analyses. Additionally, ADHD diagnosis was not grounds for study exclusion because of the high comorbidity between ADHD and dyslexia. The presence of ADHD was entered into our linear modeling analyses as a covariate. Relationships between demographic characteristics, phonological, IQ measures and reading skill are presented in Supplementary Tables B.1 and B.2.

4.2.3 Healthy Brain Network dataset

The Healthy Brain Network dataset is provided to the public by the Child Mind Institute. At the time of writing, the released dataset included 1814 subjects. From this dataset, we identified 124 school-aged individuals (ages 5-17) in the urban New York City region who had been diagnosed with Specific Learning Disorder with Impairment in Reading by a panel of clinicians affiliated with the Child Mind Institute and had also been administered the CTOPP-2. We also identified 119 individuals who were similarly assessed and given no diagnosis of any kind. Due to the large number of participants available, we were able to create nonverbal-IQ matched control groups on the basis of the Wechsler Intelligence Scale for Children's Matrix Reasoning scaled score (Dyslexia: $n = 110$; Control: $n = 105$). These groups did not significantly differ in terms of nonverbal-IQ ($t(208.85) = -1.0668, p = 0.287$) or age ($t(212.65 = 1.041, p = 0.299$).

4.2.4 Psychophysics stimuli and apparatus

Stimuli for the motion discrimination experiment were created using MATLAB (The Mathworks Corporation, Natick, MA, USA) in conjunction with the Psychophysics Toolbox (Brainard, 1997). Stimuli were displayed on a LG liquid crystal display (1,920 × 1,080 resolution, 120 Hz refresh rate, subtending 51° horizontally). The subjects response was collected using keypresses. The viewing distance was 56 cm. We used random-dot motion stimuli (150 dots) that were displayed in a circular aperture (14° in diameter) centered around the fixation mark (1°) at the center of display. Light (271 cd/m^2) and dark (0 cd/m^2) dots (dot size = 0.15°) moved at the speed of 8°/s on a gray background (135 cd/m^2). Each dot was assigned a random lifetime from a uniform distribution between 0 and 200 ms (24 video frames). When a dots lifetime expired, it was randomly re-positioned within the aperture and assigned the maximum lifetime (200 ms). Motion coherence was defined as the percentage of dots moving together in the same direction compared to dots moving in random directions. The stimuli were equivalent to those used in Joo et al. (2017a) except (a) with fixed coherence levels, and (b) stimuli remained on the screen until the subject indicated a decision with a button press (as opposed to fixed duration).

4.2.5 Psychophysics procedure

Each session comprised 6 experimental blocks. For each subject, three blocks of fifty stimuli were tested with a brief break in between. This was followed by a longer break to collect reading, phonological and IQ measures, and followed by the final set of three blocks. At the beginning of the session, subjects completed 10 practice trials comprising high coherence motion (60—100%). Subjects were allowed to repeat the practice up to three times, until they got at least 70% correct. All participants were able to do this.

Stimuli were presented at five coherence levels: 6%, 12%, 24%, 48%, and 100%. However, early in the study we realized that many subjects (unrelated to reading ability) found 100% coherence difficult and reported varying visual percepts. Performance typically declined for 100% coherence stimuli compared to 48% coherence. Therefore, we analyzed only the range of stimulus coherence levels where performance was generally monotonic, from 6% to 48%. Each stimulus coherence level was presented 60 times for a total of 300 presentations, 240 of which were included in the analysis.

Each trial started with a fixation mark at the center of the display. After 500 ms, random-dot motion stimuli were displayed until the subject made a keypress (or until 10 seconds had elapsed). Subjects pressed right or left arrow keys on a standard keyboard to report motion direction. The fixation mark was turned off when the response was made, and visual and auditory feedback was given to indicate correct and incorrect responses. The experiment did not proceed until subjects reported the motion direction. The inter-trial interval was 1 s, and after this interval the fixation mark re-appeared at the center of the display to indicate the beginning of the next trial.

4.2.6 Drift diffusion model

Fundamentally, the DDM tries to maximize the likelihood of observing a distribution of reaction times according to the probability density function

$$f(x \mid v, a, z) = \frac{\pi}{a^2} \exp(-vaz - \frac{v^2 x}{2}) \times \sum_{k=1}^{\infty} k \exp(-\frac{k^2 \pi^2 x}{2a^2}) \sin(k\pi z) \tag{4.1}$$

where x is reaction time, v is drift rate (the average rate at which evidence is accumulated for a decision), a is the distance between the two decision boundaries, and z is a bias term that allows for an observer to prefer one alternative to the other (Wald, 1947). The parameter v is allowed to vary with stimulus level. Additionally, the DDM fits a parameter t, which corresponds to the non-decision time—in other words, the time taken by all sensory and motor processes besides accumulating evidence for a decision, such as planning and executing a motor response and converting incoming sensory inputs to units of evidence.

The above equation predicts perfectly symmetric distributions of correct and error response times, and so the DDM model has been extended to include three free parameters that allow it to approximate more realistic distributions. The first of these parameters is sz, the trial-to-trial variaiblity in the drift process starting point (centered around the halfway-point between the two decision bounds), which allows the model to predict fast errors. The other parameters are sv, the trial-to-trial variability in average drift rate, and st, the trial-to-trial variability in residual time t. When the DDM is fit with these three additional parameters, it is referred to as the *full* DDM.

Although employing additional parameters can make a model less parsimonious, we elected to use the full DDM for several reasons. First, we were directly interested in whether the parameter sz might be increased in individuals with dyslexia. As there has been a great deal of recent inquiry into whether people with dyslexia employ less-than-optimal decision-making criteria on psychophysical tasks, the parameter sz represents one candidate mechanism: the propensity to execute decisions without holding a constant criterion for the relative amount of evidence across trials. Second, we assessed the Akaike Information Criterion for all subjects in the study with the full and reduced model and found that, on average, AIC was lower for the full model: the average reduction in AIC going from the reduced to full model was -2.87 with a standard error of 1.05.

We assessed the split-half reliability for all parameter estimates in the full model by fitting the DDM to two randomly-divided portions of each subject's response data (Table B.3). These results were encouraging in that, although the estimates for the trial-to-trial variability parameters were less reliable than the standard parameters, most were not prohibitively unreliable. Lastly, our interest lay in correlating the DDM parameter estimates with reading skill, so if parameter estimates were unreliable due to an overabundance of flexibility, we would reduce the likelihood of finding significant correlations between model parameters and behavioral measures. Therefore,the decision to use the full model is actually more conservative from the perspective of identifying parameters that are associated with reading skill.

The DDM was fit using the Hierarchical Drift Diffusion Model toolkit for Python (Wiecki et al., 2013). Although this package includes utilities for Hierarchical Bayesian modeling of the DDM, which provide more stable fits on fewer trials by weighting parameter estimations according to group-level distributions, we opted to fit a standard DDM to each individual. As we are interested in individual differences in behavior, and due to the statistical power available from our relatively large number of participants, we prioritized somewhat less-reliable but less-biased estimations ofF each individual's DDM parameters.

The DDM was fit to each individual's behavioral responses and reaction times using the Maximum Likelihood fitting method (as recommended by Van Zandt, 2011). The optimization scheme was attempted five times per individual and parameter estimates from the best run were saved. In all cases, the optimization scheme terminated successfully. As recommended by the makers of the HDDM package (Wiecki et al., 2013), the DDM was fit with a mixture model that allowed up to 5% of responses to be assigned to a uniform lapse distribution. This has the effect of reducing bias in drift rate estimates due to occasional lapses. Because this mixture component was included, we employed only a coarse screen for outlier detection before DDM fitting: responses occurring before 200 ms (before a typical behavioral response can be executed) and after 10s (after the stimulus had concluded) were excluded. In most subjects, this led to minimal data exclusion: although we excluded two participants with $> 15\%$ data loss, the average participant in the remaining sample had 98.1% usable data.

4.2.7 Outlier detection

To determine the presence of highly unusual model fits, we computed the Mahalanobis distance for each individual with respect to the 9 parameters estimated by the DDM. The Mahalanobis distance for multiple dimensions follows a chi-squared distribution, and so we use this measure to detect outliers (Filzmoser, 2004). Specifically, individuals with a Mahalanobis distance corresponding to values beyond the $p < 0.001$ threshold were deemed to be outliers. Two such individuals were detected; both had been fit with extremely high a values ($a = 8.17$ and $a = 5.60$). One of these individuals had a composite reading score in the Dyslexic range, whereas the other would have been above our cutoff. These two points were excluded from further analysis as we have cause to doubt the quality of their DDM parameter estimates, but their results are included with the full dataset online.

4.2.8 Stepwise model selection procedure

In our analyses of the relationships between various parameter estimates from the DDM reading skill, we employed a stepwise model selection procedure. In all cases, we considered three covariates: age, nonverbal-IQ, and the presence of an ADHD diagnosis. Each model selection procedure began with a fully specified model of reading score as a function of the parameter(s) of interest plus the three covariates. Fitting was performed with the base R `lm()` function, except where mixed model usage is noted, in which case the `lme4` library was used (Bates et al., 2016). The contributions of the covariates were first tested using an anova test. Model terms were retained if the p-value associated with the more complex model was less than 0.1. Next, parameters of interest were tested similarly. Throughout the manuscript, wherever model selection is performed we report the selected (most parsimonious) model.

4.2.9 Mediation analysis

Mediation analysis was performed using the `mediation` package for R (Tingley et al., 2014) . In all mediation models, nonverbal IQ was entered as a covariate. 4000 bootstrap simulations were used to estimate the proportion of mediation of a variable of interest by phonological awareness (the CTOPP-2 Phonological Awareness composite score) in a linear model of reading skill. Bias-corrected and accelerated confidence intervals were estimated from the bootstrapped simulations.

4.3 Results

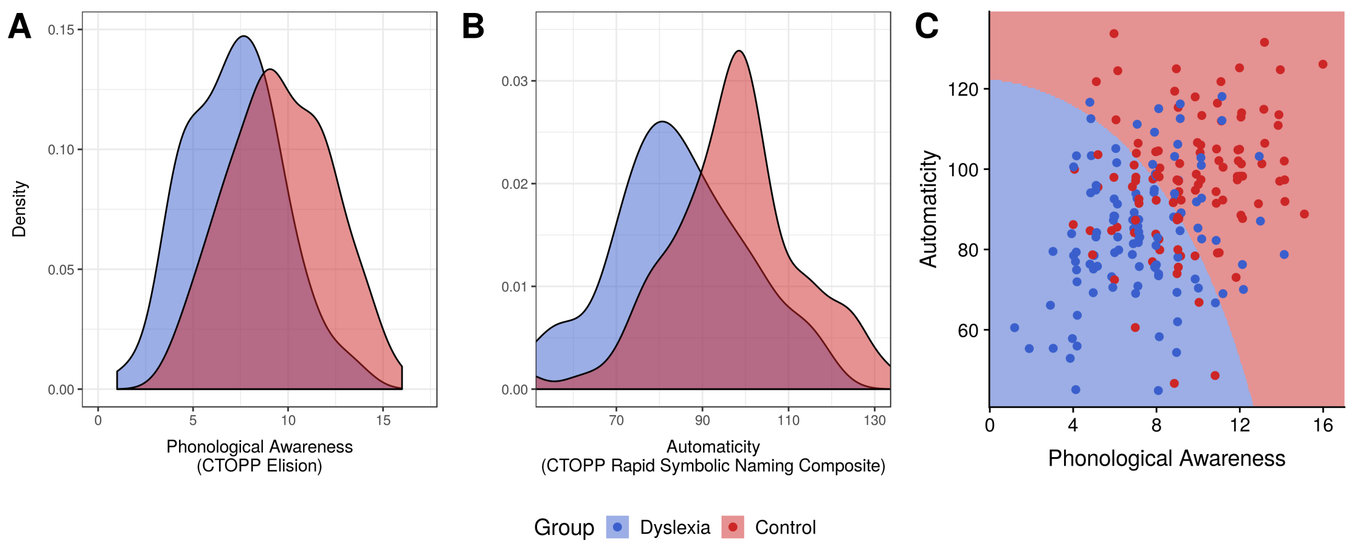
4.3.1 Predicting dyslexia from phonological measures

We first assessed the phonological core deficit model by quantifying the extent to which deficits in phonological awareness (PA), rapid automatized naming (RAN) or both differentiate individuals with dyslexia from control subjects with typical reading skills (Figure 4.1). A recently released public dataset, the Child Mind Institute's Healthy Brain Network (HBN) (Alexander et al., 2017), allows us to explore this question in a large sample of children ($N = 1814$, $n = 110$ children with dyslexia, $n = 105$ matched children with no neurological or psychiatric diagnosis). Using quadratic discriminant analysis (QDA) with age and nonverbal-IQ matched groups, we asked what proportion of children could be correctly classified on the basis of two predictors: the Comprehensive Test of Phonological Processing (CTOPP-2)'s Elision measure of PA and Rapid Symbol Naming Composite measure of RAN (both age-normed). A QDA classifier[1] trained with leave-one-out cross validation could correctly classify 67.4% of individuals with a specificity of 68.2% and a sensitivity of 66.7%.

This result is undoubtedly in alignment with the extensive literature on phonological processing: PA and RAN are both meaningful predictors of reading skill. Yet, these two measures alone fail to account for many cases of dyslexia. Furthermore, many individuals with apparently typical reading abilities would be predicted to be dyslexic on the basis of their PA and RAN scores alone.

[1]Note that a support vector machine achieved equal accuracy.

Figure 4.1: Phonological processing measures from a large public dataset.



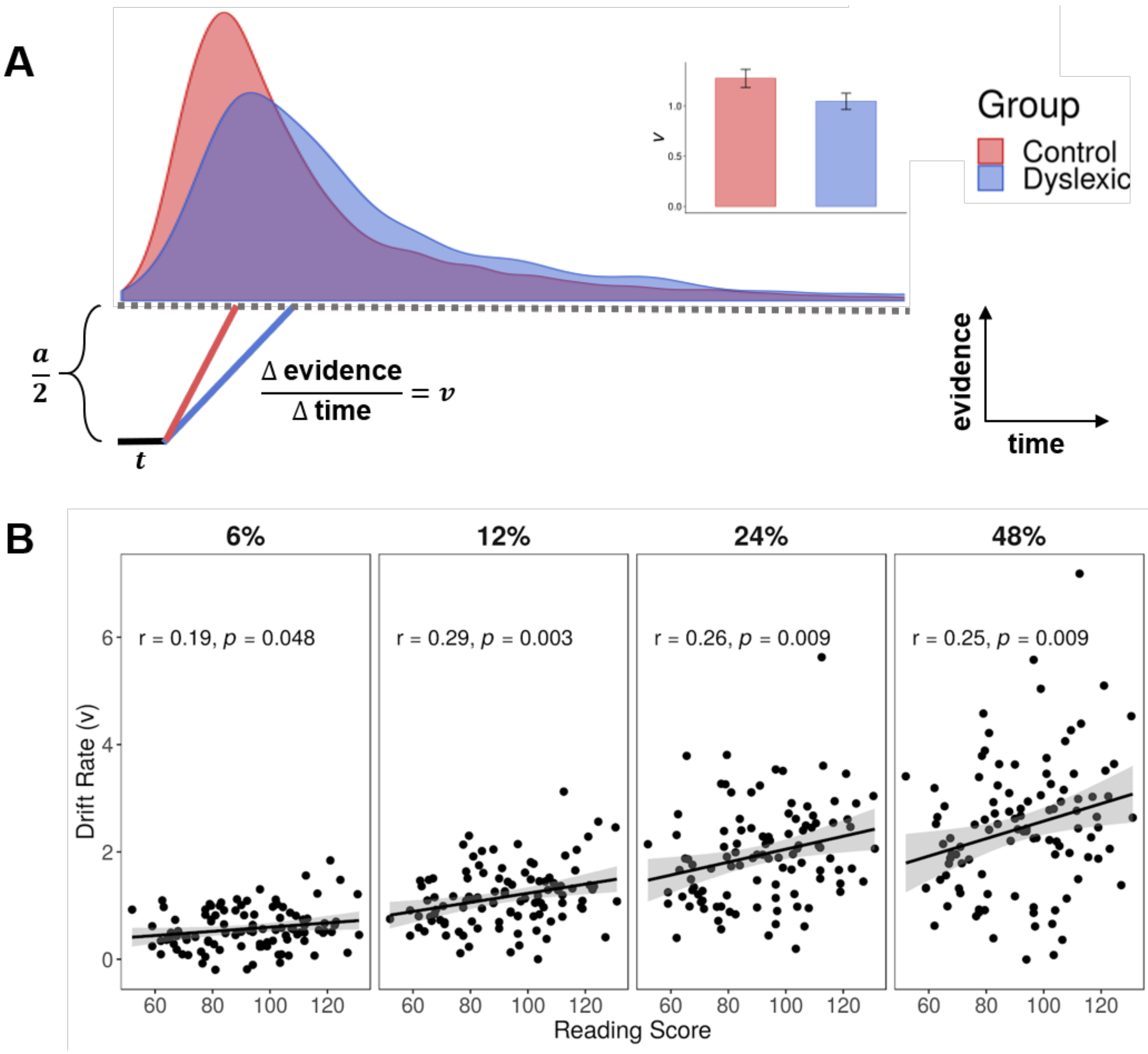
Panels A-B: Density plots for phonological awareness (PA, CTOPP Elision) and rapidautomatized naming (RAN, CTOPP Rapid Symbolic Naming Composite) in the Healthy Brain Network (HBN) dataset in two groups. The Dyslexia group (blue) consists of 110 school-aged children diagnosed with dyslexia by a panel of clinicians. The red density plot represents an age- and nonverbal-IQ-matched control group of 105 children identified as having no psychiatric or neurological diagnoses by the same panel. Panel C: The decision boundary of a quadratic discriminant analysis trained on the entire dataset is shown. Dots represent observations from the dataset with slight jitter added for visibility of overlapping points.

4.3.2 Differences in visual motion processing

Having demonstrated that phonological predictors alone are an insufficient to explain all cases of dyslexia, we next consider the contribution of visual motion processing to reading abilities: do visual motion processing difficulties typically coincide with phonological impairments, as would be expected in a cascading model of reading disability? Or are they a separable contributor to reading outcomes which explain cases of dyslexia that were not captured by the phonological core deficit model? Here we present the results of the motion discrimination experiment (conducted in the lab) in 106 school-aged children, including 42 individuals who meet our criteria for dyslexia. Accuracy and response times were collected for stimuli presented at four coherence levels: 6%, 12%, 24%, and 48%.

Before we model the respective contributions of sensory and decision processes to task performance, it is important to establish that task performance is related to reading skill. We confirmed that reading skill was related to reaction time: using model selection, we identified that the most parsimonious model of median reaction time included main effects of stimulus coherence ($\beta = -0.173$, SE = 0.009, $p < 0.001$), age ($\beta = -0.059$, SE = 0.0214, $p < 0.001$), and reading skill ($\beta = -0.006$, SE = 0.002, $p < 0.001$) with a random effect of subject (Figure B.1 and Table B.4). Accuracy was not significantly related to reading skill (Table B.5), likely reflecting the fact that the motion stimuli remained on the screen until the subject provided a response. Notably, we also observed that the ratio of correct to error median response times within each subject was significantly associated with reading skill ($\beta = -0.004$, SE = 0.002, $p = 0.0497$), with poor readers showing an increased tendency to make fast errors relative to correct response times (Figure B.2 and Table B.6). The presence of fast errors is notable because this phenomenon is typically associated with non-sensory mechanisms, including a tendency to initiate guesses before an optimal amount of evidence is considered (Smith and Ratcliff, 2004). Thus, raw reaction time data indicated that children with dyslexia were not only less efficient than control subjects, but also showed a qualitatively different pattern of responses.

Figure 4.2: Performance on the visual motion task.



Panel (A): A schematic of the drift diffusion model (DDM) with reaction time distributions (at 12% coherence) from the control and dyslexic groups imposed above. The red and blue lines in the schematic show how differences in drift rater predict differences in the reaction time distributions. The DDM model was fit separately to each individual's data and the average drift rate parameter for the dyslexic and control groups is shown in the bar plot in panel A (+/- 1 standard error). (B) The relationship between estimated drift rate and reading skill at four different stimulus coherence levels. Lines are best fit regression lines and shaded regions are confidence intervals.

4.3.3 Less efficient visual motion processing in dyslexia

To decouple sensory encoding of visual motion from the process of forming and executing a binary decision, we fit the drift diffusion model (DDM) to each subject's distribution of behavioral responses and reaction times. In the DDM for a two-alternative forced-choice judgment, it is assumed that an observer samples sensory input at discrete moments in time, and that these samples are accumulated in a noisy decision variable that represents the integrated evidence over the course of the trial (plus internal noise). When this decision variable reaches a threshold, the observer initiates a decision (Figure 4.2). The DDM therefore separates the encoding and evaluation of sensory information (which drives changes in the decision variable) from non-sensory processes, such as the magnitude of the threshold for triggering a decision and the trial-to-trial variability in the decision process (for a detailed review of the DDM, see Ratcliff et al. (2004); Ratcliff and McKoon (2008)).

After fitting the DDM to each subject's behavioral responses, we investigated whether there was a relationship between the drift rate parameter, v, and reading skill. Drift rate models the efficiency with which information is extracted and integrated from incoming sensory signals. For example, drift rate monotonically increases with stimulus coherence level ($beta = 0.719$, SE $= 0.0249$,$p < 0.001$) indicating the visual system can more efficiently extract motion information when stimulus noise is low. If individuals with dyslexia do not have any difficulties with sensory encoding, as predicted by the statistical learning hypothesis, we would expect drift rate to be uncorrelated with reading skill once covariates like IQ, age, and ADHD diagnosis are controlled for. Note that in our analyses, we treat reading as a continuous measure, but we also provide analyses where reading disability is treated as a categorical variable in the Supplement (see Tables B.7, B.8 and B.9).

Individual estimates of drift rate are shown in Figure 4.2). Drift rate was best modeled by a main effect of reading skill, a main effect of stimulus coherence, a main effect of age, and the interaction of reading skill and stimulus coherence (Table 4.1). Our results therefore indicate that drift rate increases with stimulus coherence, as expected, as well as age and

Table 4.1: Selected model of drift rate

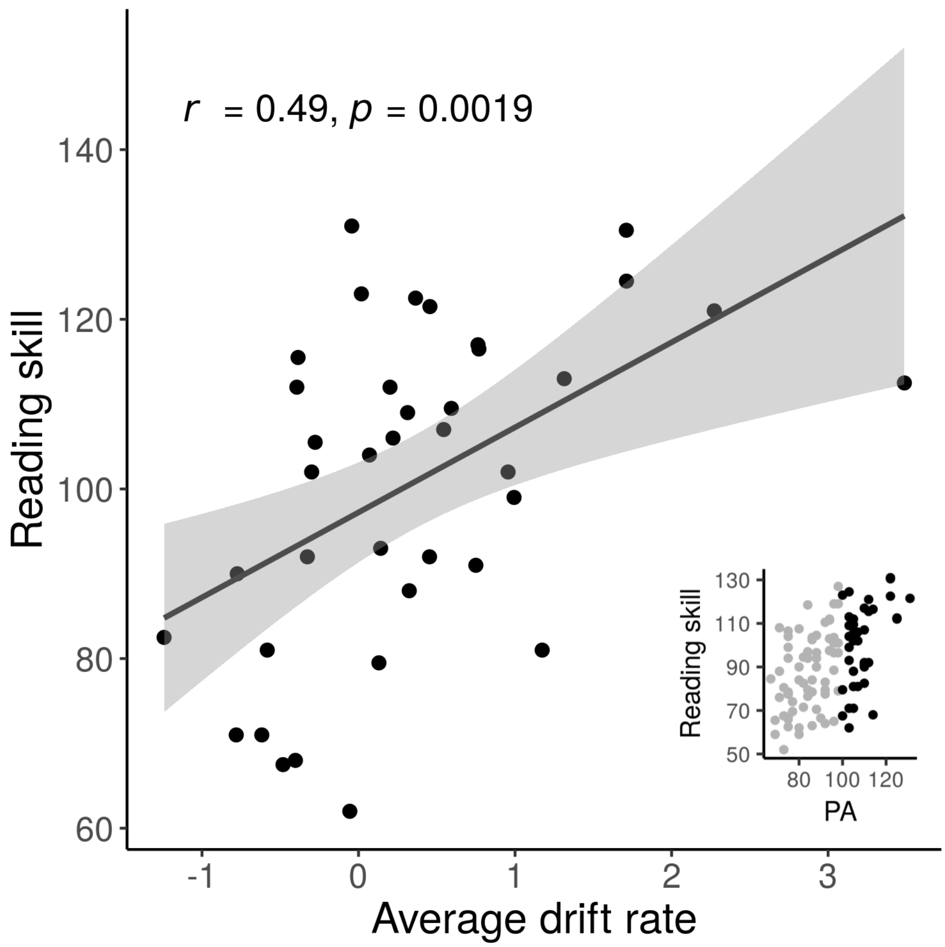
	β	SE	p
(Intercept)	1.534	0.062	<0.001
Stimulus coherence	0.719	0.025	< 0.001
Age	0.268	0.062	< 0.001
Reading skill	0.173	0.062	0.007
Stimulus coherence:Reading skill	0.087	0.025	< 0.001

reading skill. Furthermore, there is a stronger relationship between reading skill and drift rate at high stimulus coherence levels, which is likely a consequence of the fact that estimates of drift rate are more reliable at higher coherence levels.

The DDM also estimates a parameter modeling the trial-to-trial variability in drift rate, sv. This parameter is known to be correlated with drift rate under certain conditions, with higher average drift rates being associated with greater trial-to-trial variability (Wagenmakers et al., 2005; Wagenmakers and Brown, 2007). Unsurprisingly, we found that sv was correlated with drift rates at every stimulus level. It was positively related to reading skill, but this effect did not reach significance ($r = 0.13$, $p = 0.095$).

As to the question of whether drift rate explains additional variance in reading skill beyond phonological processing, consider the subset of readers in our sample with above average phonological awareness (PA scores ≥ 100). Within this subgroup of 38 participants, 9 children (23.7%) met our criteria for dyslexia despite having high phonological awareness and reading skill was significantly correlated with mean drift rate ($r = 0.49, p = 0.002$). For these individuals, knowing drift rate explains 24% of variance in reading skill (Figure 4.3. In readers with average-or-better phonological awareness, it appears that individual differences in motion encoding and sensory integration distinguish between struggling and expert readers.

Figure 4.3: Drift rate versus reading skill in a subset of individuals with good phonological awareness.



Average drift rate is calculated by averaging each individual's z-scored drift rate estimates at each stimulus coherence level. Inset: a scatter plot indicating in black which subset of the study sample is included in the good phonological awareness group.

4.3.4 Decision making parameters are related to reading skill and independent of sensory processing

We next consider the predictions of the non-sensory hypothesis by analyzing the relationship between non-sensory parameters of the DDM and reading skill (Figure 4.4A-D). If poor readers struggled with the task only because of differences in sensory encoding, we would expect no parameters besides drift rate (and sv) to be correlated with reading skill.

To the contrary, the parameter sz was correlated with reading skill and, after model selection, the best model of sz contained only a main effect of reading skill ($\beta = -0.084$, SE =0.028, $p = 0.003$). The parameter sz represents the trial-to-trial variability in the relative amount of evidence required to initiate a judgment; individuals with high sz values are prone to making fast errors. Indeed, we confirmed that the ratio of median correct response times to error response times within a subject was correlated with the DDM estimation of sz ($r = 0.452, p < 0.001$).

Similarly, we observed that the parameter representing the threshold of evidence required to initiate a decision, a, had a modest but significant correlation with reading skill ($\beta = -0.136$, SE = 0.063, $p = 0.033$), indicating that worse reading skill is associated with employing a more conservative criterion for initiating a perceptual decision. No covariates (age, nonverbal IQ or ADHD diagnoses) were retained by model selection.

Lastly, we examined parameters that represent the lumped contributions of all non-decision processes to reaction time, including the time necessary to encode a sensory stimulus and execute a motor response. Because some individuals with dyslexia are known to have slower processing speed, we might expect this time to be longer in children with worse reading skills. Indeed, the parameter t representing an individual's average non-decision time showed an overall negative relationship with reading skill. However, the magnitude of the effect was not nearly large enough to attain statistical significance, and after model selection, only age was retained as a predictor of t ($\beta = -0.050$, SE = 0.0163, $p = 0.003$). As such, maturation is associated with reduced non-decision time. Interestingly, a parameter modeling trial-to-trial variability in non-decision time, st, was best modeled by main effects of reading skill ($\beta = -0.081$, SE = 0.028, $p = 0.004$) and age ($\beta = -0.085$, SE = 0.028, $p = 0.003$).

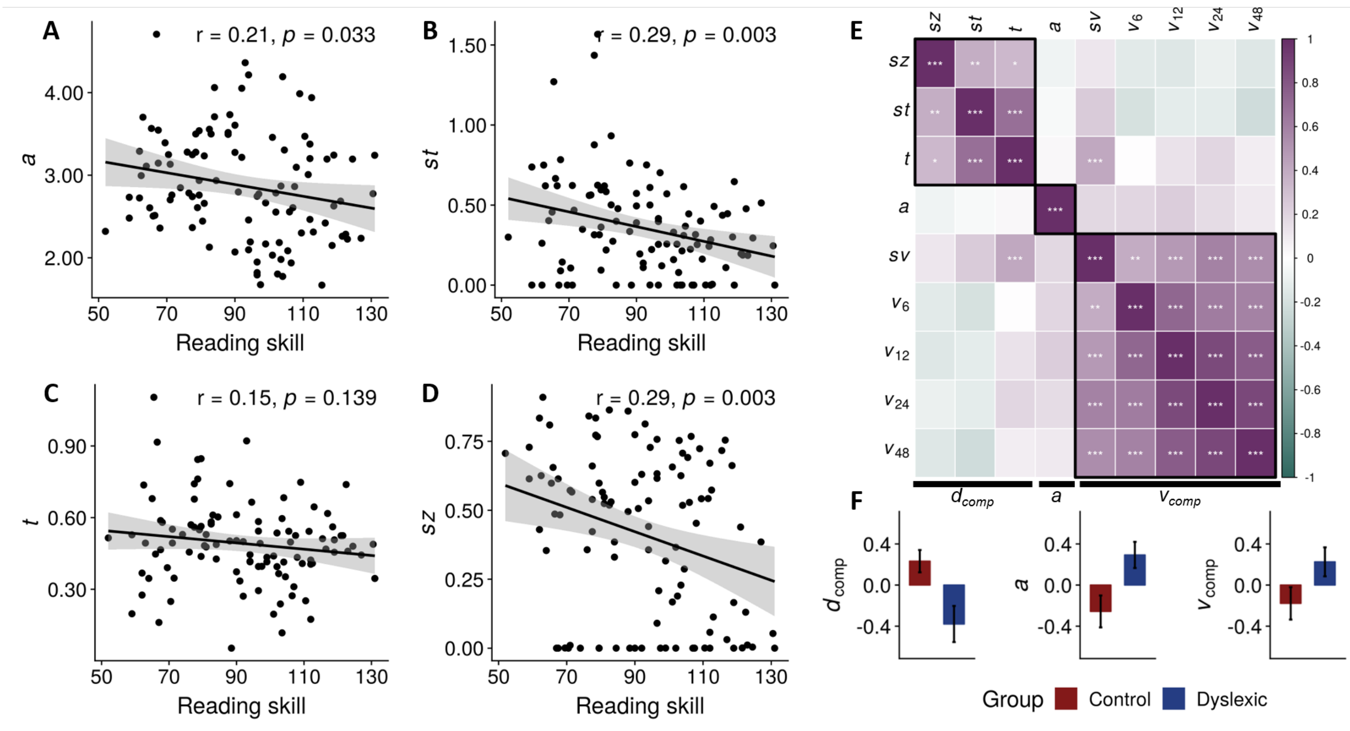
We have so far identified several parameters of the DDM, both sensory and non-sensory, that show associations with reading skill. We next considered the extent to which these parameters were correlated with one another (Figure 4.4E). As expected, we noted strong correlations between the four drift rate parameters. None of the drift rate parameters were significantly correlated with any non-sensory parameters after correction for multiple comparisons. There were moderate correlations between three non-sensory parameters, st, t and sz (st and t: $r = 0.685$, $p < 0.001$; t and sz: $r = 0.335$, $p = 0.0005$; sz and st: $r = 0.386$, $p < 0.001$) These three parameters largely contribute to modeling the leading edge of the reaction time distribution: sz allows for the presence of relatively fast errors, t shifts the response time distribution along the time axis, and st allows for responses before an individual's average response time. Finally, we noted that the parameter a was uncorrelated with any of the other parameters.

Hierarchical clustering with Ward's method (Ward, 1963) indicated that the correlation matrix was consistent with three clusters of parameters: a cluster consisting only of a, another consisting of the st, t, and sz, and a final cluster including all four drift rates and sv. This suggests that the DDM captures several independent mechanisms underlying sensory encoding and perceptual decision making.

4.3.5 Sensory and non-sensory predictors both explain reading outcomes

So far in our analysis, there seem to be several separate profiles of performance on the motion discrimination task that are associated with low reading skill: reduced ability to encode and integrate sensory information, setting a more conservative decision criterion, and generally more variability in terms of the time taken to gather evidence and/or execute a decision. The lack of correlations between many of the DDM parameter estimates indicates that individuals who display a deficit in terms of one process (e.g., sensory encoding), are not necessarily the same individuals who perform abnormally in terms of another process (e.g., decision-making)

Figure 4.4: Summary of fitted drift diffusion model parameters.



*Panels A-D: The relationship between reading score and four non-sensory parameters of the DDM. (A) decision threshold a, (B) variability in drift process starting point sz, (C) non-decision time t, and (D) variability in non-decision time st. Panel E: correlations between parameters of the DDM. Boxes indicate hierarchical clustering results (Ward's method) and stars indicate significant correlations after Holmes-Sidak correction for multiple comparisons: $p < 0.05$ is noted with *, $p < 0.01$ with **, and $p < 0.001$ with ***. Panel F: group comparisons for the three composite measures based on hierarchical clustering of the DDM parameters: d_{comp}; composite of sz, st, and t, the a parameter, and v_{comp} : composite of the four drift rate parameters and sv. Note that all three composite parameters are z-scored. Error bars represent one standard error of the mean.*

Table 4.2: Selected model of reading skill from DDM parameters

	β	SE	p
(Intercept)	0.972	0.351	<0.001
v_{comp}	-0.274	0.078	< 0.001
a	-0.339	0.119	0.005
d_{comp}	0.291	0.076	< 0.001
Nonverbal IQ	0.045	0.077	< 0.001

and that profiles of performance are variable across subjects. Therefore, we might expect that each parameter contributes separately to explaining variance in reading outcomes.

To test whether each dimension of task performance is indeed a unique contributor to a model of reading skill, we employed a linear model. To simplify the number of parameters, we introduce several composite measures based on the correlation matrix of DDM parameters and our clustering analysis (Figure 4.4F). Drift rate is summarized as a composite measure, v_{comp}, by taking the first principal component of the four drift rates and sv. A second composite measure d_{comp} was derived from the first principal component st, t, and sz, which we expect represents aspects of variability in the decision-making process.

The dyslexic and control groups differed in terms of each of these three mechanisms (Figure 4.4F). We performed model selection, starting with the full model with reading score as the dependent measure and all hypothesized DDM parameters and the three covariates (v_{comp}, d_{comp} , a, nonverbal IQ, ADHD diagnosis and age) as predictors. The selected model retained all three predictors from the DDM and nonverbal IQ (Table 4.2).

This result confirms that non-sensory mechanisms explain additional variance in reading skill once the quality of sensory evidence encoding is accounted for. As such, even within this single psychophysical task, there are multiple non-correlated dimensions of variance contributing to the pattern of responses observed in individual's with dyslexia: the ability to extract evidence from sensory information, choice of decision threshold, and trial-to-trial variability in behavior.

Table 4.3: Selected model of reading skill including phonological predictors

	β	SE	p
(Intercept)	-0.554	0.256	0.0331
v_{comp}	-0.120	0.0601	0.0491
a	-0.193	0.0871	0.0293
d_{comp}	0.142	0.0570	0.0140
Nonverbal IQ	0.335	0.0602	< 0.001
CTOPP PA	0.172	0.0653	0.0097
CTOPP RAN	0.521	0.0582	< 0.001

4.3.6 Psychophysical measures are not proxies for phonological awareness

To address the question of whether performance on the motion discrimination task is related to reading skill by way of phonological processing, or in addition to it, we explore a series of models. We first test the hypothesis that predictors from the psychophysical task do not explain additional variance in reading skill once phonological processing is accounted for. We again modeled reading skill as a function of our parameters of interest from the DDM— v_{comp}, d_{comp}, and a—as well as two phonological processing measures, PA and RAN, and the three covariates. Model selection retained all predictors except ADHD diagnosis and age (Table 4.3). Correspondingly, an ANOVA F-test comparing the selected model to a reduced model with only PA, RAN and nonverbal IQ confirmed that adding predictors from the DDM explained variance in reading skill above and beyond the reduced model ($F(100, 97) = 4.044, p = 0.009$). The reduced model also had a higher AIC (selected model AIC = 794.4, reduced model AIC = 800.7) and BIC (selected model BIC = 813.9, reduced model BIC = 815.6). From this analysis, we can confirm that all three predictors from the DDM are useful for explaining differences in reading skill in addition to traditional measures of phonological awarenes.

Because ordinary least squares models may not be suitable for obtaining reliable parameter estimates in the presence of multicollinearity, we also applied lasso regression with 10-fold crossvalidation (Friedman et al., 2010) to search for the sparsest model that accurately predicts reading score. Lasso regression seeks to reduce the number of parameters in a model by penalizing the sum of the absolute values of all regression coefficients, while also minimizing squared residuals of the model.

Lasso regression starting with the full model indicated that the only parameter that should be dropped is age (Supplemental Figures B.3 and B.4 show the effect of regularization on model accuracy and number of parameters). While lasso regression is not a tool for hypothesis testing, it confirms that the selected predictors—v_{comp}, d_{comp}, a, nonverbal IQ, PA, RAN, and ADHD diagnosis—are all sufficiently useful for explaining variance in held-out observations to justify their cost (Table B.10). In other words, removing any of these parameters would reduce the model's accuracy in predicting held- out observations.

4.3.7 Do sensory deficits have cascading effects?

It has been argued that deficits in sensory processing or decision-making could affect reading skill because they disrupt the typical development of phonological awareness (Lieder et al., 2019; Manis et al., 1997; Richardson et al., 2004). We therefore explored whether this hypothesis is borne out in our data by employing a mediation analysis. While a was not significantly correlated with PA, v_{comp} and d_{comp} showed modest correlations (v_{comp} and PA: $r = 0.324, p < 0.001$; d_{comp} and PA: $r = 0.182, p = 0.036$).

We first tested a model with PA mediating the relationship between v_{comp} and reading skill and found a significant, partial mediation effect (42.3%, $p = 0.005$). Similarly, the d_{comp} -reading skill relationship is partially mediated by PA (22.2% mediation, $p = 0.022$)), but there was also still a significant direct relationship ($\beta = 4.293$, SE = 1.501, $p = 0.005$). As such, our results provide some support for the idea that in certain poor readers, low PA could be a consequence of a more fundamental impairment in either sensory or non-sensory mechanisms. However, our data suggest a partial mediation, indicating that many individuals would not be well described by this cascading model and that there are also direct links between the model parameters and reading skill.

4.3.8 Multiple dimensions of skilled and disabled reading

Contrary to theories that seek to discover a unified deficit that characterizes children with dyslexia, we have established that sensory processing of visual motion is separable from non-sensory aspects of perceptual decision making, and both factors account for independent variance in reading skill. To speak to the question of how many separable underlying factors predict reading skill, we next apply exploratory factor analysis (EFA). EFA is an unsupervised learning approach for identifying the number, and characteristics, of latent factors that explain the correlation structure of a multi-dimensional data set (Ferguson and Cox, 1993; Costello and Osborne, 2005; Kline, 2013). We applied EFA to characterize the space of the DDM parameters, nonverbal-IQ, and the six subtests of the CTOPP (measure of reading skill were not included in the EFA). An analysis of the eigenvalues of the correlation matrix indicated that four latent factors were warranted (i.e., the first four eigenvalues > 1, see scree plot in Figure B.5). This was confirmed by parallel analysis (Hayton et al., 2004) (i.e., in a simulation of 1000 random correlation matrices, the first four resulting eigenvalues were lower than the corresponding eigenvalues from our data's correlation matrix 95% of the time). The four factors are shown in Figure 4.5 with orthogonal varimax rotation. The total proportion of explained common variance by the four-factor model was 55.8% (Factor 1: 20.3%, Factor 2: 14.2%, Factor 3: 10.7%, Factor 4: 10.6%).

The loadings of the first factor are dominated by the four drift rate parameters, whereas the second factor is loaded most heavily by nonverbal-IQ and four of the CTOPP subtests. The remaining two subtests, Rapid Digits and Rapid Letters, load onto their own factor (in line with the double-deficit hypothesis (Wolf and Bowers, 2000)). An additional factor appears to reflect non-decision time and variability parameters of the DDM st, sz, and t. Notably, the evidence threshold parameter, a, is not particularly associated with any factor; 87% of variance in a is unexplained by this model.

This factor analysis largely conforms to the intuitions we have built so far through linear models: drift rate, although correlated with phonological processing and perhaps partially mediated by it, is identified as a separate factor. Drift rate and the non-sensory parameters of the DDM are modeled as observations from two distinct factors. It is likely that a is representative of an additional factor, consistent with its lack of correlations with any other parameter of the DDM (note that without multiple estimates of a, EFA cannot estimate measurement noise and consequently does not assign it to a new factor). Critically, each of these four factors was significantly related to reading skill demonstrating that, rather than representing a single underlying construct, there are multiple, independent cognitive and sensory dimensions characterizing individual differences in reading skill (Figure 4.5). A linear model of reading skill as a function of scores on the four factors indicated that all effects were significant (see coefficients in Figure 4.5). Furthermore, the full model also had a lower AIC (full model AIC = 798.8, single factor model AIC = 869.9) and BIC (full model BIC = 814.6, single factor model BIC = 877.8).

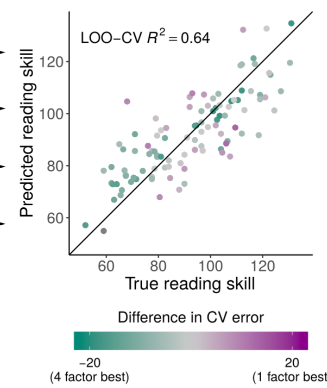
In addition to standard model selection, we compared the accuracy of the four-factor model on predicting held-out observations to the accuracy of a single-factor model. Using leave-one-out cross validation to control for overfitting, the four-factor model explained 63.9% of variance in reading skill for the held-out points. The single factor model used only Factor 2, which is largely a composite of the CTOPP measures of phonological awareness, phonological memory, and nonverbal IQ. This model was only able to explain 27.4% of variance in reading skill for held-out observations (Figure B.6), indicating the necessity of considering multiple underlying dimensions (at least 4) in order to accurately predict individual differences in reading ability.

4.4 Discussion

Our results demonstrate that (1) a core phonological deficit model is insufficient to account for many cases of developmental dyslexia, (2) abnormal performance on the motion discrimination experiment in children with dyslexia cannot be ascribed to a uniform profile of either sensory or non-sensory deficits, (3) both sensory and non-sensory mechanisms explain variance in reading skill above and beyond phonological processing, (4) the correlational

Figure 4.5: Results of exploratory factor analysis

	a	t	sz	st	sv	v_6	v_{12}	v_{24}	v_{48}	Nonverbal IQ	CTOPP Rapid Letters	CTOPP Rapid Digits	CTOPP Phoneme Isolation	CTOPP Nonword Repetition	CTOPP Memory for Digits	CTOPP Elision	CTOPP Blending Words	
Rapid automatized naming	−0.18	−0.04	−0.04	−0.16	0.16	0.04	0.08	0.09	0.06	0.06	0.93	0.82	0.05	0.08	0.03	0.39	−0.06	β = 10.6 (1.1), $p < 1 \times 10^{-15}$
Decision variability	−0.04	0.79	0.38	0.89	0.39	−0.09	0.01	0.03	−0.08	0.11	−0.08	−0.14	−0.16	−0.09	0	−0.09	−0.12	β = -3.0 (1.1), p = 0.01
Phonological processing	−0.09	0.03	−0.25	−0.12	0.06	0.1	0.22	0.21	0.2	0.65	0.03	0.13	0.48	0.62	0.61	0.71	0.51	β = 9.9 (1.1), $p < 1 \times 10^{-15}$
Motion processing	0.22	0.16	−0.08	−0.13	0.58	0.68	0.86	0.94	0.88	−0.06	0.07	0.15	0.13	0.06	0.15	0.23	0.12	β = 2.4 (1.1), p = 0.03



Factor loadings for the orthogonal four-factor model are shown in the table; shading corresponds to absolute value of the loading. The scatterplot shows the correspondence between true (measured) and predicted reading skill using a linear model with all four factors as predictors. Each point was predicted using leave-one-out cross-validation (LOO-CV). Color indicates whether that point was more accurately predicted by the single-factor model or the full model with all four factors. Green points had a lower squared error when predicted by the four- factor model, and purple points had a lower squared error when predicted by the single-factor. Gray points had similar prediction accuracy for both models.

structure of cognitive, linguistic and sensory measures explored here is consistent with, at minimum, four underlying factors, (5) each of these four factors accounts for unique variance in children's reading abilities. In sum, our results are not consistent with models of dyslexia that only consider phonological processing or models in which impairments in sensory encoding or decision making primarily affect reading development via a disruption of phonological processing. Instead, dyslexia should be conceptualized as a disorder that may arise from several distinct loci.

Our work is consistent with that of the Pennington and colleagues, which has capitalized on large samples to demonstrate that individuals with dyslexia have a heterogeneous profile of cognitive and linguistic impairments (Pennington, 2006; Pennington et al., 2012; Peterson and Pennington, 2015a). The present work extends this perspective to address the role of sensory processing and perceptual decision-making deficits in dyslexia.

Several preceding studies have attempted to investigate multiple candidate mechanisms of dyslexia, including auditory, visual, and motor processes. Our work generally conforms to the finding of at least four such studies (Ho et al., 2002; Menghini et al., 2010; Ramus, 2003; White et al., 2006) that show a heterogenous pattern of deficits present in struggling readers. In a study with related methodology, Talcott et al. (2002) collected several psychophysical measures in 350 school aged children and found that each uniquely explained a small percentage of variance in literacy skill. Our study similarly finds that several distinct mechanisms each explain a small, but unique, proportion of variance in reading outcomes.

To our knowledge, the present study is the first use of the DDM to model motion discrimination in children with dyslexia. Our results serve as a partial validation of two seemingly contradictory theories: some poor readers show a pattern of performance consistent with reduced ability to extract information from incoming sensory signals, while others are better described as having normal sensory processing but altered decision-making characteristics (including, as the propensity to make fast errors reveals, more trial-to-trial variability in the relative amount of evidence needed to initiate a decision). Neither the statistical learning hypothesis, which would argue that sensory deficits are not meaningful, nor the magnocellular deficit hypothesis, which would fail to predict the non-sensory parameters of the DDM relate to reading skill, entirely match our results. Yet we see evidence for both sensory- and non-sensory profiles of impairment in our sample. In line with these findings, we propose that each mechanism should be reconceptualized as a dimension of risk, as opposed to a single cause, of reading difficulties.

As a correlational study, our results cannot validate any particular causal mechanism. It is possible that each factor represent clusters of symptoms that indicate underlying impairment in a processing system, but are not a direct cause of dyslexia themselves. For example, the fact that differences in visual motion processing predict unique variance in reading skill does not necessarily mean that, for those individuals, poor perception of visual motion is the cause of their reading difficulty. Instead, measurements of task performance may be a proxy for the fidelity with which the visual system constructs a sensory representation of a noisy stimulus (Sperling et al., 2005), or the efficiency of information transfer between visual regions (Yeatman et al., 2012, 2013), or the integration of sensory signals over time (Joo et al., 2017a). Skilled reading requires rapid communication among a distributed network of visual, auditory and language processing systems and an impairment in any one of these systems, or the connections between them, could cause difficulties learning a complex skill like reading (Wandell and Yeatman, 2013).

Our main conclusion is a lack of concordance with either a single deficit or cascading model. As such, our results contradict claims that a single mechanism, either phonological or sensory, can be considered the fundamental or core deficit of dyslexia. In particular, our work opposes the recent claim that the majority of individuals with dyslexia have a magnocellular processing deficit (Stein, 2018); if the DDM is accepted as a reasonable model of behavior on the motion discrimination task—a starting point with considerable basis (Huang-Pollock et al., 2017; Palmer et al., 2005; Ratcliff and McKoon, 2008)—then we conclude that a minority of children with dyslexia are best modeled as having a motion encoding deficit.

Furthermore, while we do not directly test auditory theories of dyslexia here, our results still speak to this research. For example, the influential temporal sampling hypothesis holds that the core deficit of dyslexia is abnormal processing of syllable-scale acoustic features, which in turn disrupts PA development and manifests as sampling problems in the visual domain (Casini et al., 2018; Goswami, 2011). Our results indicate that, even if we could establish that abnormal auditory processing impairs PA, many cases of dyslexia would still be unaccounted for based on the effectiveness of a phonological-core model. Furthermore, the idea that difficulties sampling incoming stimuli largely explains poor performance on the motion discrimination task is specious, as we have demonstrated that there are several reasons (some non-sensory) why individuals with dyslexia may perform differently on this task than typical readers. While the idea of a centralized deficit in some aspect of temporal processing has an elegant appeal, our data are simply not consistent with such a simple model.

The clinical implications of this multifactorial model are a target for future research. Whether or not different risk profiles predict outcomes for children enrolled in competing intervention programs is an empirical question that cannot be readily inferred from correlational data. For example, in a previous intervention study we demonstrated that individual differences in visual motion sensitivity have no prognostic value for predicting a child's response to intervention (Joo et al., 2017a).

Moving forward, we propose an additive risk factor mode of dyslexia in which multiple dimensions of sensory, cognitive and linguistic processes contribute distinct risk for reading difficulties. Our results are agnostic to whether poor performance on any given task indicates deficits in the specific targeted function (e.g., motion processing) or indexes processing capacities of a broader system (e.g., constructing a high-fidelity representation of a noisy visual signal). There are many proposed neurobiological mechanisms that could, in theory, be compatible with our findings (e.g., heterogeneous profiles of abnormal cortical migration (Hancock et al., 2017)).

In sum, our results demonstrate that an additive model outperforms cascading deficit models or models that only consider measures of phonological processing without considering the role of sensory processing and perceptual decision making. Thus, rather than continuing to seek an underlying cause of dyslexia, the field should systematically build towards a more complete model of the factors that add risk (or protection) for reading difficulties. Our data and model necessitate a shift towards theories that explain skilled and disabled reading as emerging from a high-dimensional space determined by several distinct processing systems.

Chapter 5

CONTEXT EFFECTS ON CATEGORICAL LABELING IN CHILDREN WITH DYSLEXIA

The work below is currently in review at the journal Scientific Studies of Reading as a manuscript with co-authors Liesbeth Gijbels and Jason Yeatman.

Research shows that on average, children with dyslexia behave less categorically in phoneme labeling tasks. This study investigates three subtle ways struggling readers may perform differently than their typically developing peers in this experimental context: sensitivity to the frequency distribution from which speech tokens are drawn, bias induced by previous stimulus presentations, and fatigue during the course of the task. We replicate findings that reading skill is related to categorical labeling, but we do not find evidence that sensitivity to the stimulus frequency distribution, the influence of previous stimulus presentations, and a measure of task engagement differs in children with dyslexia. It is therefore unlikely that the reliable relationship between reading skill and categorical labeling is attributable to artifacts of the task design, abnormal neural encoding, or executive function. Rather, categorical labeling may index a general feature of linguistic development whose causal relationship to literacy remains to be ascertained.

5.1 Introduction

It is well established that reading skill is correlated with performance on phoneme categorization tasks, in which listeners are asked to categorize spoken syllables based on a single contrastive phoneme (Noordenbos and Serniclaes, 2015; Vandermosten et al., 2010; O'Brien et al., 2018, 2019; Goswami et al., 2002). However, the mechanism underlying the link between impaired processing of phonemes and developmental dyslexia remains unclear.

While phonological awareness, the ability to identify and manipulate phonemes, is one of the strongest predictors of dyslexia, there are several reasons to question that phonological processing is the "core deficit" that explains why children with dyslexia struggle learning to read. Some researchers have criticized the "core phonological deficit theory" on the grounds that not enough children could be accurately diagnosed on the basis of phonological awareness alone (Wolf and Bowers, 2000; Pennington et al., 2012). Perhaps the most popular line of reasoning, though, is that children with dyslexia perform (on average) poorly on many measures of auditory, visual, working memory, and automaticity that cannot be explained by a phonological deficit alone. These observations have motivated a new wave of research searching for a more fundamental mechanism that might explain the myriad of deficits (including phonological awareness) that are associated with reading (dis)ability.

While some researchers have taken the perspective that individuals with dyslexia have fundamentally impaired auditory or visual processing (Stein, 2018; Goswami, 2011; Tallal et al., 1996), the psychophysical literature on the whole is currently inconsistent with a uniform pattern of sensory impairment (Hämäläinen et al., 2013; Amitay et al., 2002; Stuart et al., 2006; Rosen, 2003). Noting this, some researchers have argued that individuals with dyslexia are constrained not by sensory processing at a basic level, but by the demands of psychophysical tasks (Ramus and Ahissar, 2012; Ahissar, 2007).

While this hypothesis could potentially explain the heterogeneity observed in the sensory processing literature, a consensus is yet to be reached regarding which particular aspects of psychophysical tasks are the "bottleneck" in performance. One candidate is attention and task engagement; dyslexia is comorbid with ADHD (Light et al., 1995; Stevenson et al., 2005; Germanò et al., 2010). In accords with this hypothesis, one previous study has shown that performance on "catch trials" tends to degrade faster over the course of a task in poor readers than controls (Messaoud-Galusi et al. (2011); see also Roach et al. (2004)). Another candidate is statistical learning. Statistical learning was originally defined as "a way of extracting statistical regularities from the environment" (Saffran et al., 1996). It has been proposed that individuals with dyslexia are less able than their typically developing peers to take advantage of regularities in their environment (Banai and Ahissar, 2006; Ahissar et al., 2006; Lieder et al., 2019; Gabay et al., 2015). The statistical learning hypothesis is especially appealing because the process of literacy training involves forming connections between phonological and orthographic representations, which requires a learner to extract regularities from visual and auditory sequences (Ziegler and Goswami, 2005), and also to learn and automate the probabilistic relationship between a given letter and the phonemes it represents. Thus, there is growing interest in the possibility that individual differences in domain-general mechanisms such as statistical learning could explain degraded performance across an array of experiments as well as a difficulty with learning to read.

The phoneme categorization task is, in some ways, an ideal ground to explore how task performance may be differentially affected by task demands in struggling readers. Because it has been so extensively used, experimenters can have reasonable confidence that the key effect—shallower psychometric functions in struggling readers—is replicable and not specific to a particular acoustic cue (Noordenbos and Serniclaes, 2015; O'Brien et al., 2018, 2019).

Previously, we showed that reduced categorization in struggling readers is somewhat influenced by the working memory demands of the task, but could not be entirely explained by task difficulty: irrespective of the task-difficulty we found a correlation between reading skill and task performance (O'Brien et al., 2018). We now investigate several other aspects of task performance to clarify the extent to which the categorization-reading relationship depends on specific experimental conditions. We first consider the effects of varying the frequency distribution from which speech tokens are drawn. Typically in categorization experiments, stimuli are drawn from a uniform distribution. Distributional learning—sensitivity to the distribution from which stimuli are drawn—is a key part of language acquisition in development (Maye et al., 2002). Neurotypical adults show sensitivity to the stimulus distribution in the categorization task (Clayards et al., 2008): the task elicits more categorical behavior when stimuli are drawn from narrow bimodal distributions than broad distributions. Recently, work by Vandermosten et al. (2018) suggested that children with dyslexia were on average less able to utilize distributional cues to learn a non-native speech contrast. In the present study, we examine how children ages 8-12 performed on a categorical phoneme labeling task with two conditions: a bimodal and uniform distribution of native speech tokens.

Second, we explore how the immediate context of recently presented speech tokens affects judgments about the identity of the current stimulus. Considering how recent stimulus presentations influence performance is of interest for several reasons: it addresses longstanding claims that individuals with dyslexia struggle when stimuli are presented sequentially (Tallal, 1980) or that they show abnormal stimulus adaptation and faster implicit memory decay (Jaffe-Dax et al., 2017; Perrachione et al., 2016; Ahissar et al., 2006). Finally, we look for hallmarks of fatigue and disengagement in our participant's responses by examining changes in task performance over the duration of the experiment. In line with our previous work we find that there is a moderate relationship between phoneme categorization and reading skill; this relationship cannot be attributed to (1) the stimulus distribution, (2) stimulus recency effects or (3) task disengagement. Thus, we conclude that some people with dyslexia have difficulties categorizing speech sounds and that this deficit, though likely not universal, is not an artifact of experimental conditions such as the distribution, order and duration of the experiment.

5.2 Methods

5.2.1 Participants

A total of 62 native English-speaking children ages 8-12 were recruited for the study. Children without known auditory disorders were recruited from a database of volunteers in the Seattle area (University of Washington Reading & Dyslexia Research Database). Parents and/or legal guardians of all participants provided written informed consent under the University of Washington Institutional Review Board protocol. All subjects demonstrated normal or corrected-to-normal vision. Participants were tested on a battery of cognitive and literacy assessments, including the Woodcock-Johnson IV (WJ-IV) Letter-Word Identification and Word Attack subtests, the Test of Word Reading Efficiency (TOWRE), and the Weschler

Abbreviated Scale of Intelligence. All participants underwent a hearing screening to ensure pure tone detection at octave frequencies between 500 Hz and 8000 Hz in both ears at 25 dB HL or better.

5.2.2 Demographics

Here, we present analyses of task performance where reading skill is treated as either a continuous or discrete group variable. We have previously argued that reading skill is best modeled as a continuous variable, in agreement with Shaywitz et al. that there is no clear demarcation between readers who are below-average and readers who are dyslexic (Shaywitz et al., 1992b). Our results on phoneme categorization published thus far (O'Brien et al., 2018, 2019) confirm an entirely continuous relationship between task performance and reading skill. However, for completeness and ease of comparison with existing literature on dyslexia, we also provide group-level analyses.

Reading skill was summarised in a composite variable: as both the WJ-BRS and the TOWRE index are scored on the same standardized scale, a composite reading skill measure was created by averaging the two metrics for each participant. Using a composite of both measures as our criterion improved the confidence of our group assignments since they are highly correlated measures in our sample ($r = 0.877, p < 2e-16$). Participants were assigned to the Dyslexia group if their composite reading score was at least 1 standard deviation below the population mean (i.e., < 85). No participant in the Control group had any parental report of dyslexia or a history of reading disability. All subjects were required to have nonverbal IQ and full-scale IQ (WASI Matrix Reasoning and FS-2 scores respectively) no less than 1 standard deviation below the population mean (as in O'Brien et al., 2018, 2019); 3 subjects were below this cutoff and excluded from further analysis.

There were 24 subjects in the Dyslexic group (13 male) and 35 subjects in the Control group (16 male). There was not a significant difference in age between groups (Kruskal-Wallis rank sum test, $H(1) = 2.194, p = 0.139$), although we noted that there was a small correlation between age and reading skill ($r = 0.253$, $p = 0.053$). Importantly, we included age as a

Table 5.1: Summary statistics and group differences on demographic and behavioral measures.

	Control	Dyslexic	Significance
	n =35 16 male, 19 female	n =24 13 male, 11 female	
WASI-II			
FS-2	121 (12.7)	101.8 (13.1)	< 0.001
Nonverbal IQ	59.6 (8.7)	50.3 (7.8)	< 0.001
Woodcock-Johnson IV			
Basic Reading Score	111.6 (12.1)	79.8 (7.2)	< 0.001
Nonword	111.3 (13.5)	87.8 (10)	< 0.001
Real Word	110.4 (11.6)	74.8 (10.3)	< 0.001
TOWRE 2			
TOWRE Index	101.7 (12.8)	69.5 (7)	< 0.001
Nonword	100.5 (14.8)	72.9 (8.2)	< 0.001
Real Word	102.8 (12.3)	69.2 (9.3)	< 0.001
CTOPP 2			
Phonological Awareness	101 (13.8)	91.6 (11.4)	0.021
Phonological Memory	99.1 (13.7)	88.2 (12.1)	0.003
Rapid Naming	98.7 (11.2)	83.7 (9.8)	< 0.001

covariate in our analyses. We did not exclude participants with ADHD diagnoses from the study because ADHD is highly comorbid with dyslexia [Germano et al 2010]. Instead, we accounted for ADHD diagnosis as a covariate in our statistical analyses. Of 59 participants, 13 had a formal diagnosis of ADHD: 7 in the Dyslexic group and 6 in the Control group. The difference in prevalence of ADHD across groups was not significant ($H(1) = 1.178$, $p = 0.278$).

Table 5.1 shows group comparisons on measures of reading and cognitive skills. As in our previous findings, IQ (either measured as full-scale or nonverbal) differed by group. While we were not concerned that low IQ prevented any subject from understanding the task because low IQ was an exclusion criterion, we included nonverbal IQ as a covariate in our statistical analyses.

5.2.3 Stimuli

A 7-step /ba/∼/da/ speech continua were created using Praat v. 6.0.37. In the /ba/∼/da/ continuum, the starting frequency of the second vowel formant (F2) transition was varied. Note that this is the same continuum used in (O'Brien et al., 2019), to which we refer the reader for greater detail on the synthesis procedure. In brief, the seven speech tokens were identical except for their F2 formant contour. The starting frequency of F2 was varied in seven linearly-spaced steps from 1085 Hz (/ba/) to 1460 Hz (/da/). F2 rose in a linear ramp to a terminal value of 1225 Hz over the course of 100 ms, at which point the steady-state portion of the vowel was maintained for 250 ms.

5.2.4 Procedure

Stimulus presentation and participant response collection was managed with PsychToolbox for MATLAB (Brainard, 1997). Auditory stimuli were presented at 75 dB SPL via circum-aural headphones (Sennheiser HD 600). Children were trained to associate sounds from the two speech continua with animal cartoons on the left and right sides of the screen and to indicate their answers with right or left arrow keypresses. Large text labels were provided over each animal cartoon ("Ba" on the left side and "Da" on the right), so participants did not have to memorize the animal associated with each sound. Throughout all blocks, each cartoon was always associated with the same stimulus endpoint.

Participants first completed a practice round consisting of ten presentations, five of each continuum endpoint, with feedback on each trial. Participants were allowed to repeat the practice round up to three times, until they had achieved at least 75% accuracy. All participants were able to meet this minimum standard.

The main task was presented in two parts, one in which the stimuli were drawn from a uniform frequency distribution and another in which they were drawn from a bimodal distribution. In the unimodal condition, all stimuli were presented 15 times. In the bimodal condition, the presentation frequency was greatest at the continuum endpoints and least in

Table 5.2: Stimulus presentation frequency in bimodal condition.

Stimulus	1	2	3	4	5	6	7
Frequency	52	34	14	10	14	34	52

the center of the continuum (see Table 5.2).

Because we were interested in exploring the effects of recently presented stimuli on judgments about the current stimulus, we used a "random but frozen" list of stimuli. This means that we randomly generated the order in which stimuli would be presented in each condition, but every participant was tested with this fixed stimulus order. This reduced one source of variability across subjects so that we could perform more targeted investigations about how recent stimulus presentations differentially affect strong and poor readers.

In each condition (uniform or bimodal frequency distribution), participants heard a total of 210 speech sounds. After every 35 stimulus presentations, a quick optional break was presented. Between the two test conditions, reading assessments were performed. If a participant did not already have an IQ measure on file from a previous lab meeting, the WASI-II was also administered.

Seven participants completed the uniform condition first and 45 completed the bimodal condition first. The reason for this discrepancy is that during data collection for the first 15 subjects, we alternated which condition was presented first. After data was collected for these subjects, we were surprised to see little evidence that participants behaved differently in either condition. We therefore changed to a policy of always providing the bimodal condition first, wary that initial exposure to the uniform condition could affect category learning in subsequent conditions. We considered simply discarding the data of the seven subjects who performed the uniform condition first; however, we were unable to detect any significant differences between task performance in these individuals and the remainder of the cohort. Psychometric function slope did not significantly differ by group ($\beta = -0.140$, $\mathrm{SE} = 0.331$, $p = 0.674$), nor was there a significant interaction between the order conditions were presented and slope in each condition ($\beta = -0.0715$, SE = 0.309, $p = 0.818$). On this evidence, we retained these seven subjects in the data set, although they are clearly marked so any analyses can be re-run with or without them present.

5.2.5 Psychometric curve fitting

As in all previous chapters, we used the MATLAB toolbox Psignifit 4.0 to fit psychometric functions. The fitting routine optimized the fit of a logistic curve function with four parameters modeling the upper and lower asymptotes, width of the logistic function, and the threshold. The width of the logistic function was transformed to the slope at the threshold value to give a standardized measure of psychometric function slope.

Psignifit uses a Bayesian framework to optimize parameter estimates not only according to likelihood of generating a given set of behavioral responses, but also with regard to prior distributions of each parameter. In the case of psychometric function fitting, where the number of presentations of each stimulus is often relatively low, inappropriate priors can have an outsized influence on parameter estimates- particularly, as we and many others have summarized elsewhere, when it comes to estimates of the slope and asymptotes. We therefore used priors validated in our previous work on a similar task, with a similar participant demographic (O'Brien et al., 2019): the asymptotic priors were modeled as a uniform distribution on the range [0, .10]. In other words, the lower and upper asymptotic parameters could vary freely in the range [0, .10], to give a lower asymptote between 0-10% and an upper asymptote between 90-100%. This prior width was chosen on the basis of ten-fold cross-validation over the dataset (see O'Brien et al. (2018, 2019) for further details).

To ensure the validity of psychometric function parameter estimates, as in our previous work (O'Brien et al., 2018, 2019) we excluded any psychometric functions that could not be fit with a threshold between continuum steps 1 through 7. Only one psychometric function (produced by a subject in the Dyslexic group presented with a uniform stimulus distribution) was excluded on this ground.

5.2.6 Statistical analysis of parameter estimates

After we fit psychometric functions for each subject in each condition, we used a series of generalized linear mixed models to determine the relationship between reading ability, the frequency distribution from which stimuli are drawn, and three dependent measures. As in (O'Brien et al., 2018, 2019), these dependent measures were estimates of task performance based on behavioral responses: (1) psychometric function slope, (2) lapse rate, and (3) a composite measure of psychometric function shape.

Lapse rate was determined by averaging the upper and lower asymptote estimates of a given function (i.e., their deviations from 0 and 1 respectively). The composite measure was constructed from a principal components analysis on the four parameters of each psychometric function collected in the study. The first principal component captured 45.5% of variance in the four parameters, and was defined by the following linear weights: threshold: -0.385, slope: 0.486, upper asymptote: -0.498, lower asymptote: -0.606.

Model selection was performed using the `lme4` library for R with parametric bootstrapping tests of model fit performed with the `pbkrtest` library. For each dependent measure, fixed-effect predictors with sum coding were used for the condition (uniform or bimodal) variable. Reading ability was entered as a continuous fixed-effect predictor except where otherwise stated. Additional predictors were added for presence/absence of ADHD diagnosis (treatment coding), age (continuous predictor) and for non-verbal IQ (WASI-II Matrix Reasoning score; continuous predictor). A random intercept for participant was also included. For all model analyses, we began with a fully specified model of reading score as a function of the parameters of interest (slope, asymptote, or PC1) plus the covariates (ADHD, age and non-verbal IQ) and a random intercept for subject identity. The contributions of the covariates were first tested using parametric bootstrapping, which is robust to non-normally distributed residuals. Model terms were retained if the bootstrapped p-value of the coefficient being nonzero was less than 0.1. After testing the three covariates, we tested terms of interest for the study (condition, reading ability, and the interaction between them) using the same procedure.

5.3 Results

As expected on the basis of previous studies in our lab and others, we found relationships between reading skill and psychometric function shape (O'Brien et al., 2018, 2019; Vandermosten et al., 2010; Noordenbos and Serniclaes, 2015). In Figure 5.1, we can see that psychometric slope, asymptote and PC1 (a general measure of psychometric function shape) were all correlated with reading ability.

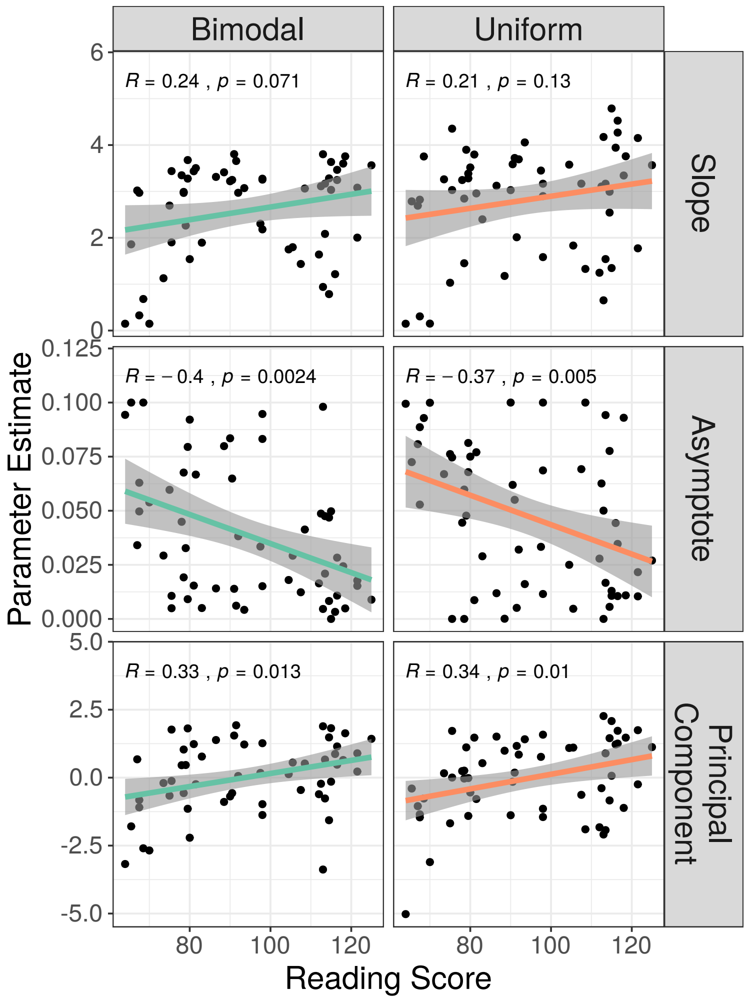
We confirmed this with generalized linear mixed model analysis, first with regard to the relationship between reading ability and psychometric slope. After model selection, the most parsimonious model of psychometric slope contained a continuous predictor for reading ability only (Table 5.3).

Note that in addition to the covariates (age, ADHD diagnosis, and nonverbal IQ), the main effect of condition ($\beta = 0.265$, SE = 0.162, $p = 0.107$) and the interaction between condition and reading ability ($\beta = -0.061$, SE = 0.163, $p = 0.708$) were dropped by model selection. Thus, we did not detect a significant relationship between psychometric slope and the frequency distribution from which stimuli were drawn. Moreover, we did not find evidence supporting the hypothesized interaction between reading skill and experimental condition (bimodal or uniform). In other words, we replicate the equivalent relationship between phoneme categoriztion and reading skill irrespective of the distribution from which the stimuli were drawn.

Similarly, the selected model of lapse contained only main effects of reading skill and the covariate, age (Table 5.3). Lastly, considering PC1 as the dependent variable, the selected model contained only main effects of reading skill and age (Table 5.3). Taken as a whole, no aspect of psychometric shape appears to depend on the task condition for strong or struggling readers.

For completeness, we also tested the interaction of reading skill and stimulus distribution with reading skill treated as a categorical variable (Dyslexic vs. Control). A mixed effects

Figure 5.1: Relationship between reading skill and phoneme categorization.



Plots of model psychometric function parameter estimates versus reading score. Each point corresponds to parameter estimates for one subject in one condition (bimodal or uniform). Lines indicate the best fit regression line with 95% confidence intervals in shaded regions.

Table 5.3: Selected models of psychometric function parameters

		β	SE	p
Slope				
	(Intercept)	2.723	0.12	< 0.001
	Reading skill	0.223	0.12	0.069
Lapse				
	(Intercept)	0.102	0.023	< 0.001
	Reading skill	-0.010	0.003	0.006
	Age	-0.006	0.002	0.013
PC1	(Intercept)	-2.777	1.111	0.016
	Reading skill	0.342	0.162	0.039
	Age	0.279	0.111	0.015

ANOVA with a random effect of subject was used to evaluate the interaction term, using the Kenwards-Rogers estimation of degrees of freedom. The interaction was not significant in a model of slope ($F(1, 55.6) = 0.860$, $p = 0.357$), lapse ($F(1, 61.7) = 0.390$, $p = 0.535$), or PC1 ($F(1, 53.3) = 0.390$, $p = 0.861$). The estimated Cohen's d for the separation of slope by group was 0.48, with a 95% confidence interval ranging from -0.09 to 1.04; this range is consistent with results in two of our previous studies and the average effect size in a meta-analysis of categorical labeling studies (O'Brien et al., 2018, 2019; Noordenbos and Serniclaes, 2015). For separation of lapse by group, $d = 0.75$ (CI = $[0.17, 1.32]$), and for separation of PC1 by group, $d = 0.64$ (CI = $[0.07, 1.20]$).

Altogether, our analyses indicate that from the behavioral responses alone, there was no evidence for sensitivity to the stimulus distribution in our sample of participants. This makes it challenging to interpret our data as evidence for or against a statistical learning deficit in struggling readers. What our data do suggest is that the apparent relationship between reading skill and performance on the categorical phoneme labeling task is not likely to be a simple artifact of the stimulus frequency distribution chosen by the experimenter.

5.3.1 *Effects of recent stimulus presentations on phoneme labeling*

Because we collected 420 responses per individual, our dataset may provide sufficient power to examine stimulus recency effects. To explore this possibility, we employed the modeling approach of Lieder et al., which uses generalized linear models to investigate how recent stimulus presentations affect the judgment of the current stimulus' identity (Lieder et al., 2019).

For every stimulus presentation in the data set, we determined the identity of the preceding four stimuli. As in Lieder et al. (2019), we adopt the following notation:

Let d_0 be the stimulus step (1-7) of the current stimulus presentation, t. Then d_1 is the difference in steps between d_0 and the stimulus presented at trial $t-1$. Similarly, d_2 is the difference in steps between d_0 and the stimulus presented at trial $t-2$, and so on for values d_3 and d_4.

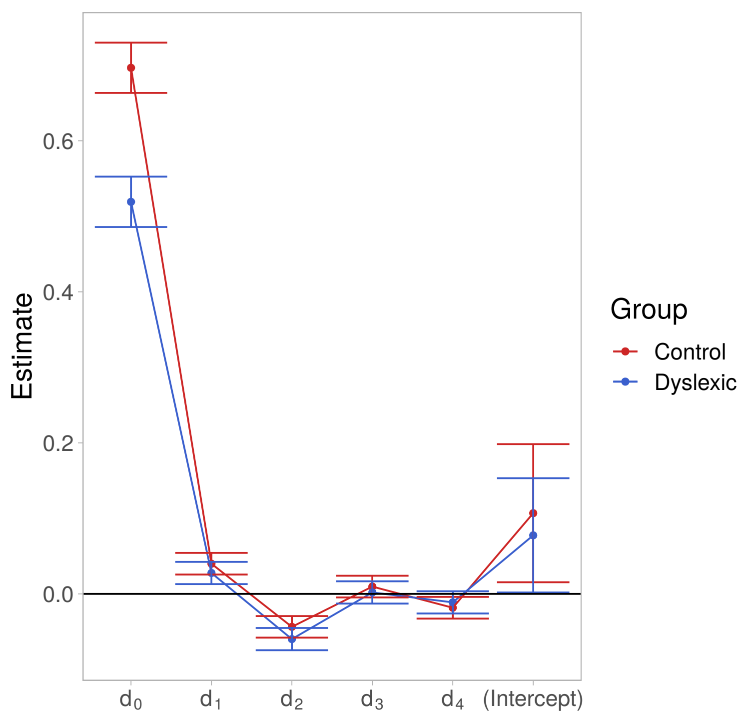
The mixed effects GLM specifying the relationship between the label assigned to the current stimulus presentation and the recent presentations is as follows:

$$response = f(\beta_0 + \beta_1 d_0 + \beta_2 d_1 + \beta_3 d_2 + \beta_4 d_3 + \beta_5 d_4 + (1|SubjectID)) \qquad (5.1)$$

where f is the probit link function, β coefficients are linear weights to be estimated, d_0 represents the continuum step of the current stimulus presentation, and $(1|SubjectID)$ is a random intercept for subject. The probit was chosen as the link function because the dependent variable, *response*, is binomially distributed–i.e., participants decided whether a sound was "da" or not. We note that the probit function contains only two parameters that variously adjust the slope and threshold of a sigmoid, and, therefore, attentional lapses would have an effect the estimated slope.

For the purpose of illustrating this approach clearly, we begin with an exploration of group differences and then move on to a model where reading skill is treated as a continuous variable. We first fit a mixed effects GLM to the responses of each group (Control and Dyslexic). The estimated coefficients are compared in Figure 5.2.

Figure 5.2: Estimates of recency effects in Control and Dyslexic groups.



Estimates of coefficients from a mixed effects generalized linear model (GLM) fit to group behavioral data. Bars indicate the 95% confidence interval surrounding a given estimate.

Table 5.4: Hypothesized model of behavioral response

	β	SE	p
(Intercept)	0.097	0.034	0.005
d_{inf}	0.634	0.010	< 0.001
Reading skill	0.002	0.034	0.95
d_1	0.034	0.005	< 0.001
d_2	-0.049	0.005	< 0.001
d_0*Reading skill	0.133	0.006	< 0.001

We can immediately see that stimulus recency effects exist and (d_1 through d_4) are quite similar between groups. The coefficient that differs by group is the weighting of d_0—in other words, the mixed effects GLM estimates that probit slope is lower in the Dyslexic group even when recent stimulus presentations are accounted for. The intercept term, which quantifies the threshold of the probit function, is non-zero (i.e., the threshold is not exactly at the center of the continuum) but appears not to differ meaningfully by group.

Having visualized the group-level differences, we follow up with a treatment of reading skill as a continuous measure in the GLM. On the basis of our initial exploration, we drop the terms d_3 and d_4 from the model, as the standard errors for these point estimates suggests we are under-powered to detect recency effects at this timescale.

First, we consider the mixed effects GLM containing main effects of d_1, d_2, d_0 and reading skill, plus an interaction of d_0 and reading skill. We hypothesized a significant interaction between reading skill and d_0 on the basis of the previous group-level model. Indeed, the interaction as well as the stimulus recency terms were all highly significant (Table 5.4).

We also tested augmenting our hypothesized model to include an interaction between reading skill and stimulus recency terms d_1 and d_2. Adding a d_1*Reading skill interaction increased the AIC from 15604.6 to 15604.8 and the BIC from 15660.7 to 15668.8, and the new interaction term was not significant (β = 0.007, SE = 0.005, p = 0.184). Considering a d_2*Reading skill interaction term fared no better: the interaction was not significant

($\beta = 0.0007$, SE $= 0.005$, $p = 0.894$), and AIC and BIC increased (to 15606.5 and 15670.7 respectively) compared to the simpler model.

From this investigation of stimulus recency effects, we find that we are able to detect significant effects of the last two stimulus presentations on judgments of the current stimulus' identity. Our models do not provide evidence for an interaction of reading skill and the influence of previous stimulus presentations. Rather, the model upholds the interpretation that psychometric functions are steeper in stronger readers, regardless of the context each stimulus presentation occurs in.

Lastly, we performed an analysis to characterize the mechanism by which recently presented stimuli influence judgments of the current stimulus. We hypothesized that if the current stimulus was ambiguous—i.e., it was drawn from the center of the /ba/∼/da/ continuum—then the influence of the previous stimulus would be greatest. In other words, listeners might make greater use of the contrast between the current and previous stimuli when the current stimulus is ambiguous than when thee current stimulus is a clear category exemplar. We tested this hypothesis with another mixed effects GLM. First, we created a new binary feature that distinguishes stimuli drawn from the center of the continuum versus the endpoints:

$$ambiguous = \begin{cases} 0 & \text{if } d_0 \in [\text{Steps 1,2,6,7}], \\ 1 & \text{otherwise.} \end{cases}$$

We then tested the model

$$response = f(\beta_0 + \beta_1 d_0 + \beta_2 d_1 * ambiguous + (1|SubjectID)) \tag{5.2}$$

where f is a probit function as in Eqn. 5.1. We were specifically interested in the interaction of d_1 and *ambiguous*: a significant interaction indicates that magnitude of the difference between the current and past stimulus depends on whether the current stimulus is a category exemplar or not. As expected, the interaction of d_1 and *ambiguous* was significant

($\beta = 0.085$, SE $= 0.007, p < 0.001$). This analysis upholds the intuition that previous trials influence the present judgment by providing a contrast to judge ambiguous stimuli by. If this is indeed the primary mechanism by which stimulus recency effects influence performance on the categorization task, then our study is not alone in finding a lack of interaction between reading skill and such contextual effects: Blomert and Mitterer (2004) also found no evidence for context effects at several linguistic scales in a phoneme labeling task like ours.

5.3.2 Quantifying fatigue during the task

The relatively large number of trials collected per subject allows us to revisit an analysis proposed by Messaoud-Galusi et al. (2011), to determine whether poor readers show precipitous declines in task performance as the study goes on. If poor readers become fatigued or distracted at a faster rate than strong readers, that could explain overall differences in task performance.

To this end, we modeled the probability of correctly labeling a clear "ba" or "da" exemplar as a function of trial number and reading skill. Note that clear category exemplars are stimuli drawn from the two ends of the continuum (steps 1 and 7). Having already established that poor readers produce shallower psychometric functions overall, we should expect that the probability of correctly labeling these tokens will be lower overall in poor readers. If an interaction of trial number and reading skill is found to be significant, that would suggest task fatigue occurs differentially across the spectrum of reading skill.

Once again, we used a mixed effects GLM with subject as a random effect (as each subject participated in two test conditions, each with 210 trials). The dependent variable was accuracy on labeling an endpoint of the continuum, which was coded as 1 or 0. The model included a main effect of trial number, a main effect of reading skill, and an interaction of the two. Trial number and reading skill were scaled and centered prior to modeling. The results of this analysis are provided in Table 5.5, and a visualization of group trends is shown in Figure 5.3.

While there was a significant interaction of reading skill and trial number, the direction

Table 5.5: Model of accuracy labeling continuum endpoints

	β	SE	p
(Intercept)	1.672	0.067	< 0.001
Trial	-0.103	0.022	< 0.001
Reading skill	0.264	0.067	< 0.001
Trial*Reading skill	-0.044	0.021	0.037

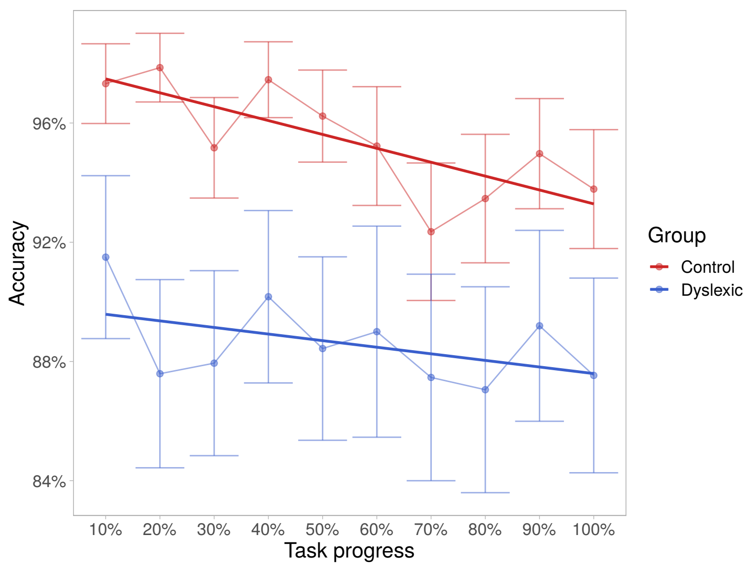
of this effect is actually opposite what we might have predicted—greater reading skill is associated with a more deleterious effect of trial number on accuracy. Inspection of our data reveals that this trend is strongly influenced by one particular subject in the Dyslexic group, who began with nearly chance accuracy and became more accurate over the course of the task. When we removed this subject from the model, the magnitude of the interaction effect more than halved (β went from -0.044 to -0.018) and the interaction was no longer significant ($p = 0.397$).

As such, our results do not corroborate the idea that poor readers are especially prone to becoming disengaged, distracted, or tired during the task—at least by this proposed measure of engagement.

5.4 Discussion

We considered how the frequency distribution from which stimuli are drawn during a standard phoneme labeling task might deferentially affect task-performance in children with dyslexia versus typical readers. Indeed, some authors have posited that differences in psychophysical task performance may, in some situations, actually reflect a difference in an individual's sensitivity to task conditions (i.e., the distribution of the stimuli). Because our task did not appear to induce measurable statistical learning effects in participants of any reading ability, we are unable to comment on how such effects may differ in children with dyslexia. However, our results replicate many previous findings in our lab and others about the relationship between reading skill and categorical labeling, and suggest that poor readers' behavior on

Figure 5.3: Accuracy throughout the task in Control and Dyslexic groups.



Each point represents the average accuracy within a group in a certain interval of the task. Error bars mark the 95% confidence interval of the mean. The x-axis marks progress through the task: 10% marks the first 21 of 210 trials, 20% marks trials 22-42, and so on. Dark lines indicate the best fit regression line.

the task is not strongly sensitive to the stimulus frequency distributions.

There are several reasons why we may not have seen clear effects of stimulus distribution, whereas other experimenters have in similar contexts (Vandermosten et al., 2018; Clayards et al., 2008): unlike in Vandermosten et al.'s study of distributional learning in children with and without reading disability, we used a native-language contrast that may have been over-learned by our participants prior to our study. If this explanation were true, it would imply that children with dyslexia are entirely equipped to leverage statistical learning to establish phonetic categories from their natural environment (albeit, perhaps to a lesser degree than their typically developing peers). Another explanation is that our measurements may not have been sufficiently precise: Clayards et al. detected stimulus distribution effects on psychometric function shape for a native-language contrast, but used eye-tracking to recover a time-series measure of looks to a closed set of choices displayed on a screen.

Our data set also allowed us to apply the modeling approach of Lieder et al. (2019) to investigate how previous stimulus presentations affect judgments about the current stimulus. We were able to detect effects of the previous two stimulus presentations, but critically, did not find that these effects interacted with reading skill. Our results are broadly consistent with Lieder et al.'s findings, which showed that stimulus recency effects were similar in adults with and without dyslexia (albeit in a task involving the judgment of tone frequency differences). For stimulus recency effects to be intact in children with dylsexia, it seems necessary to also have intact sensory encoding—if stimuli were not encoded with sufficiently high fidelity, it is difficult to imagine that children would be sensitive to differences between previous and current presentations. If our inference is valid, then our results may provide further evidence against the position that less categorical behavior on the phoneme labeling task is a consequence of impaired or noisier sensory processing in poor readers (Goswami, 2011; Hancock et al., 2017; Casini et al., 2018). Our results also seem to contradict claims that adaptation or anchoring to recent stimuli is somehow different in children with dyslexia (Jaffe-Dax et al., 2017; Ahissar et al., 2006; Perrachione et al., 2016; Krause, 2015; Nicolson and Fawcett, 2018).

Finally, we tested whether children with dyslexia showed signs of increased fatigue during the task by analyzing their performance on relatively easy trials throughout the course of the task. Although individuals with dyslexia showed a tendency to make more errors on "easy" trials throughout the experiment, we did not find evidence to suggest that they were merely becoming less attentive over time, as Messaoud-Galusi et al. (2011) did in a similar task. It is possible that our results differ because we allowed children a brief break (typically less than one minute) every 35 trials, whereas participants in Messaoud-Galusi and colleague's study adhered to a different schedule. Additionally, our participant demographics may have differed, as many of our children are well-accustomed to computer games at home and at school. While we are therefore cautious to generalize our results broadly, we can conclude that task fatigue is unlikely to explain the patterns of categorical labeling we present here. This may be reassuring with regard to the large literature on categorical labeling in individuals with dyslexia: while experimenters must remain vigilant of ways that overall decreased accuracy can bias measures of task performance (Wichmann and Hill, 2001a; Roach et al., 2004), our results suggest that a simple explanation of task engagement alone is unlikely to account for the entire relationship between reading and categorical labeling.

Considering our results and the current state of the field, we believe researchers are at an intriguing moment: there is compelling evidence that in certain experiments, apparent deficits in groups of participants with dyslexia are well-explained by non-linguistic and non-sensory mechanisms (Gabay et al., 2015; Banai and Ahissar, 2004, 2006), and this framework has considerably more power to explain the diversity of deficits associated with reading disability than purely sensory or phonological models. Still, there are considerable gaps in this explanation: not only are there are experimental contexts where individuals with dyslexia appear to have no statistical learning deficit (Gabay and Holt, 2018; Staels and Van den Broeck, 2015; Gould and Glencross, 1990; Inácio et al., 2018; Jiménez-Fernández et al., 2011; Samara and Caravolas, 2017; Du and Kelly, 2013)—perhaps reflecting ongoing vagueness in what "statistical learning" encompasses—but effect sizes in studies that do detect group differences are too small to accurately separate most cases of dyslexia from typical reading (Lieder et al., 2019; Vandermosten et al., 2018).

Even if we take the view that reduced categorical labeling in struggling readers is entirely the consequence of impaired sensitivity to phonetic categories over the course of many years of language exposure, group separability would remain quite modest: the average effect size in categorical labeling studies is 0.66 (Noordenbos and Serniclaes, 2015), meaning that only 9.7% of individuals with dyslexia would fall below the 95% confidence interval of the control population. A further problem for the statistical learning hypothesis is that it often relies on an assumption that a very subtle impairment can cascade to have drastic effects on literacy by disrupting the development of typical phonological processing. However, our previous studies (O'Brien et al., 2018, 2019) and several others (Robertson et al., 2009; Snowling et al., 2019; Talcott et al., 2000; Calcus et al., 2018) suggest that a cascading model is inadequate: performance on psychophysical tasks can relate to reading skill separately from the proposed phonological processing mediation pathway. It may be that the categorical labeling task is an index of something far broader than phonological awareness, such as overall developmental status or linguistic experience. Sharpening of category boundaries may be partially a result of reading experience itself.

In sum, our results are consistent with the perspective that multiple causal routes relate performance on various behavioral and psychophysical measures to reading skill (Pennington et al., 2012; Ziegler et al., 2019). Under this model, deficits in learning category boundaries from speech sounds may be one of many factors that contribute to difficulties with reading. In light of this, we are most optimistic towards future research that explores how constellations of risk factors, including but not limited to reduced category learning, phonological processing, and sensitivity to the statistics of non-linguistic stimuli, act in concert to determine a child's reading skill.

Chapter 6

CONCLUDING REMARKS

In the course of four experiments, my coauthors and I have investigated several contemporary theories of dyslexia using psychophysical and statistical methods. Considering our results in the context of the literature as a whole, we arrive at several conclusions that we hope will prove useful to the field.

First, psychometric function shape on the phoneme categorization task is a robust and replicable predictor of reading skill in school-aged children. This relationship is not strongly dependent on the type of phonetic cue (purely spectral or spectrotemporal), cue duration, frequency distribution of stimulus presentations (uniform or bimodal), or the precise format of the psychophysical experiment design (3-interval or single interval). The result is also not easily accounted for by considering the influence of attentional lapses, recent stimulus presentations, or engagement during the task.

Our results are in some ways supportive of historical interpretations of what phoneme categorization means about language and reading, but in other ways pose challenges. On the one hand, our results encourage the view that a great deal of literature on phoneme categorization in struggling readers reflects a very real linguistic difference (Mody et al., 1997; Serniclaes et al., 2004; Noordenbos and Serniclaes, 2015) that is very unlikely to be an artifact of attention or task-specific challenges. Because this trend can be seen in so many experimental contexts—including synthetic and natural stimuli, purely spectral and purely temporal contrasts, vowels and consonants, in children and adults, with tonal and atonal language speakers—we believe it is no longer tenable to suggest the origin is a particular sensory deficit, as in any flavor of the temporal processing hypotheses of dyslexia. While our results alone do not rule out that sensory deficits of any kind could be present in some individuals with dyslexia, we conclude that the relationship between phoneme categorization and reading skill is not attributable to such a deficit in most cases.

At the same time, this work does not strongly support the classical thinking that phoneme categorization is causally related to phonological awareness: throughout the literature, it is common to see phoneme categorization equated with the ability to segment words into phonemes or blend phonemes into words. In other words, it is assumed that the relationship between phoneme categorization and reading skill is mediated by phonological awareness. Yet we find only limited empirical evidence for this model, as do several other recent studies in relatively large samples (Robertson et al., 2009; Hakvoort et al., 2016; Snowling et al., 2019). Our work adds to accumulating evidence that phoneme categorization is a direct predictor of reading skill, acting to an extent *in addition* to phonological awareness. This is perhaps unsurprising considering that only limited empirical evidence supports a view of speech perception in which context-invariant phonemes are extracted from speech in natural listening conditions. To quote Holt and Lotto (2010, p. 1223):

> "Everyday speech perception in the wild is likely to tap into a broader set of processes than those captured in individual laboratory tasks. It is important to note that this is not to suggest that adult (or even infant or animal) listeners cannot categorize speech; there is abundant evidence that they can. Rather, these data suggest that the cognitive and perceptual processes involved in speech categorization and those in online perception of fluent speech may not be one and the same."

Indeed, it is perhaps more important for development that children learn to be appropriately *un*-categorical as listeners: the task of speech perception involves adapting to many different talkers with many different dialects. The theoretical foundation linking phoneme categorization to phoneme awareness in dyslexia seems to presuppose that speech perception proceeds through a necessary downsampling stage from rich acoustic cues to abstract phonemes, which are then mapped to graphemes during literacy training—an idea that's been substantially challenged in the decades since the study of phoneme categorization in dyslexia began (Cleary and Pisoni, 2001; Port and Leary, 2005; Port, 2007; McMurray et al., 2009; Toscano et al., 2010). Without disregarding the immense contributions of these classical ideas to promoting many thoughtful avenues of research, it may now be time to reconsider the theoretical framework connecting speech to literacy.

This relates to a second conclusion of the dissertation: a purely phonological account of dyslexia is insufficient to account for most cases of disabled and typical reading. Not only is phonological awareness itself inadequate as a predictor, we find evidence for several other predictors that do not appear to be mediated by phonological awareness. Automaticity, phonological awareness, visual motion processing and non-sensory aspects of decision making appear to each confer separate risk for reading disability (though it is important to note that our results do not imply that each risk factor has a causal relationship with reading skill, as the predictors we identified may only be correlates of an unknown causal mechanism). Our results add to the emerging picture that a single-deficit model, or even double-deficit model, of reading disability is over-simplified (Pennington et al., 2012; Peterson and Pennington, 2015b; Ziegler et al., 2019; Booth et al., 2000; Schatschneider et al., 2016; Spencer et al., 2014; Erbeli et al., 2018). While we do not contest that phonological awareness remains an important predictor of reading ability, our findings are at odds with the notion of a characteristic "profile" of cognitive impairments in dyslexia.

6.1 The future of dyslexia research

In light of our results and the broader literature, we can make several proposals for future research. The shift we hope to see in dyslexia research is less emphasis on "core deficits" and more on "risk factors"—i.e., a transition from a deterministic to stochastic view acknowledging that dyslexia, and likely other learning disabilities, may not have a unifying causal mechanism that explains the majority of cases. Taking this view will require less reliance on small-scale psychophysical studies comparing groups of "control" and "dyslexic" participants and increased focus on studies with hundreds, or even thousands, of participants where natural variability in reading skill can be exploited. There are both theoretical and practical ramifications for understanding and predicting reading disability, which are briefly detailed here.

6.1.1 Designing studies with multifactorial models in mind

Because our results and others point to a multi-dimensional model of reading skill without clear discontinuities or clusters of readers, studies that aim to clarify risk factors of reading disability must prioritize diverse recruitment and rigorous statistical modeling over traditional comparisons between typical readers and "pure dyslexic" individuals without any other known diagnoses. From the standpoint of any of the measures presented in this dissertation, there is no validity in discretizing participants into "dyslexic" and "control" groups; statistics should almost always be carried out with continuous measures except in rare cases when a hypothesis warrants it. Studies of reading disability will be helped by including not only children diagnosed with dyslexia, but other developmental disorders that are co-morbid with it, such as ADHD and dyscalculia (in agreement with Peters and Ansari (2019)).

Because a multifactorial model implies many small effects (i.e., risk factors) working in concert to explain variance in reading skill, researchers must overcome the momentum of relying on small sample sizes and invest in designing studies that can scale. Presently, many visual and auditory psychophysical tasks can be delivered over the internet to vastly increase the number and diversity of participants. To our knowledge, this technology is still largely underutilized in the field.

With regard to experimental methods, little more can be learned by assessing group-level differences in performance on a single psychophysical task—as we and others have shown, it is nearly trivial to find tasks where struggling readers perform differently than typical readers. The key will be in understanding which measures appear to tap the same latent variables and which can be disassociated from one another. To reach this goal quickly, the research community must adopt a number of cultural changes: datasets and experimental software must be made publicly available, and measurement reliability must be reported whenever a new psychophysical or behavioral test is used. In tandem with this recommendation, we also urge a shift in how correlations and group-level differences are discussed: the presence of a modest effect that does not explain the majority of variance in reading skill does not imply that measurement error explains the rest. This kind of thinking may partially underlie the continued popularity of auditory explanations of dyslexia, even though effect sizes for these measures are typically underwhelming.

6.1.2 Asking new questions about the role of speech perception in reading skill

While categorical phoneme labeling is a predictor of reading skill, its relationship to everyday speech perception remains unclear. More studies of phoneme categorization will reveal little about naturalistic speech perception in children with dyslexia, and should only be undertaken with the specific intention of studying speech sound categorization. Researchers would benefit from taking advantage of methods such as eyetracking and pupillometry that yield nuanced time-series measures of listening effort and on-line prediction in response to natural speech. This would allow researchers to directly test many hypotheses about dyslexia, including whether struggling readers engage in less prediction of upcoming words (by the statistical learning hypothesis) or exert more effort than their peers in difficult listening conditions (sensory theories).

Furthermore, children with known hearing impairments and/or cochlear implants provide another important avenue for understanding the role of the auditory system in reading. In children with cochlear implants, experimenters have a rigorous understanding of the maximal information that can be transmitted via the device—thus, specific hypotheses about the effects of reduced access to speech acoustics can be tested in a population with known, rather than assumed, auditory impairments. In particular, children with cochlear implants have reduced ability to understand speech in noise and lack access to the fine temporal information that aids typical listeners in discerning speech sounds and talkers (and, in tonal languages, words). It is especially worth noting that many children with cochlear implants still achieve age-appropriate reading skills (Dillon et al., 2012). Understanding the protective factors that allow these children to excel may provide useful insights for dyslexia, if indeed there is any causal relationship between the kind of subtle variance in auditory acuity that occurs in the non-hearing impaired population and reading disability.

As a final note on auditory processing, some researchers have speculated that becoming literate drives top-down changes in speech perception (Huettig and Pickering, 2019), and vocabulary growth (often a side-effect of literacy) may drive changes in sensitivity to phonetic cues (Cleary and Pisoni, 2001). Intervention studies in which children receive literacy training and are tested on speech perception or auditory measures before and after can directly test these ideas. Studies of this nature will help settle debates about causes of reading disability versus consequences of it.

6.1.3 Embracing a stochastic model of learning disability.

In a deterministic model of reading skill, a child's ultimate reading skill is a function of their particular cognitive profile (e.g., the strength of their phonological ability) and external circumstances. In a probabilistic framework, weak phonological skill may predispose a child to struggle at reading, but this outcome is not inevitable. There are both theoretical and practical ramifications of this shift: in theory, it implies that typically, no one cognitive mechanism in isolation determines reading skill. As such, devoting research to understanding predictors of phonological processing (as much of auditory research has done so far) will have only limited relevance to understanding dyslexia.

If there are several distinct risk factors for dyslexia, it is natural to wonder if the ideal treatment may differ across individuals based on their pattern of risk factors. Presently, the literature on intervention broadly supports curricula that include a strong phonics component, but also includes text exposure and training in vocabulary, morphology, and semantics (Wolfgang et al., 2000; Roth and Schneider, 2001; Piasta and Wagner, 2010; Wolf and Katzir-Cohen, 2001); approaches targeted to auditory or visual training are not backed by evidence. Currently, there is little research to suggest that the efficacy of standard evidence-based intervention protocols differs systematically across children—for example, children of both high-and low- socioeconomic status appear to benefit equally (Torgesen, 2000). To my knowledge, there is thus no high-quality evidence, positive or negative, that suggests struggling readers can be matched to an ideal intervention program based on their cognitive profile.

Perhaps the simplest way to begin addressing this question rigorously would be to measure rapid automatized naming and phonological awareness in a cohort of reading disabled children (two dimensions of reading skill predictors that are correlated, but separable). Then children would be randomly assigned to one of two interventions, one focusing on automatized naming and the other on phonological awareness. Researchers would test whether children who were assigned to an intervention that matched their deficit received greater benefit than their peers who were mismatched. This sort of study design would begin to address, at a very coarse level, whether there is promise in "optimizing" curricula for particular learners.

There is also the matter of early prediction of reading disability, in the hopes that providing early intervention will prevent future academic struggles. The exact accuracy that a predictive model must attain to be socially and economically viable is a question that can only be answered through a cost-benefit analysis (weighing the risks of over- and under-diagnosis); for now, even the best model presented here would still be too inaccurate to be clinically useful. In practice, a stochastic framework may limit how successfully researchers can screen young children for future learning disabilities.

Although it is possible that identifying more indicators of risk will allow for more precise estimation of a child's future reading skill, it is worth considering that no number of measures may ever yield a useful predictive model because the nature of disability may be governed by noisy random processes. Recently, this was highlighted in a presentation by Arvind Narayanan, part of Princeton's Fragile Families project (Narayanan, 2019). In a data science challenge, the Fragile Families team shared a dataset of nearly 13,000 measures on children (ages 9 and under) from 4,242 families. Data scientists were asked to predict children's eventual GPA at age 15, measures which were held-out during the competition. In the winning model, only 19% of variance in GPA at age 15 could be explained. Most importantly, this model—containing almost 13,000 predictors— performed only marginally better than a model with 4 hand-chosen predictors. Narayan labeled the task of predicting scholastic success in individuals—like predicting job success, terrorism, or recidivism—as "fundamentally dubious". Although predicting future reading disability is not exactly the same as predicting future GPA, this example illustrates how tenuous the notion of individual educational forecasting is today.

Finally, there is the question of what we intend to predict and why. Even if we could develop a model that perfectly predicted a child's eventual decoding skills, decoding is only relevant to public health to the extent that it determines a child's social and economic future. Considering the modest correlation between socioeconomic status and decoding skill (it is estimated that 10% of variance in decoding is explained by socioeconomic status (Peterson and Pennington, 2015a)) and the strong correlation between socioeconomic status and academic achievement (estimated around 70% of variance by Reardon (2016)), it seems clear that the majority of American children are not limited in their academic achievement by any neuroscientific mechanism but by structural inequality. It is therefore my opinion that studying the cognitive mechanisms that support strong reading skill is a valuable basic science question, but ultimately, predicting who will fall in the tail of the normal distribution of reading skill must be a low public health priority while widespread social, economic, and racial disparities still exist in American schools.

6.1.4 Conclusion

There is still much to be done to explain the variability in reading skill that occurs in the population, and more importantly, to provide timely and effective services to all learners who require them. As we learn to better characterize the prevalence of various risk factors, it may be possible—with institutional support and funding—to identify at-risk children early and make preventative literacy training a standard feature of public education. We should also be prepared for the possibility that neither "big data" nor individualized curricula will show material benefits over simply providing a universally high standard of education to all children from a reasonably early age. Further research is required to answer this empirical question. To assist the majority of American children, though, I believe no more psychophysics or intervention studies are needed: educational reform is a matter of resources and priorities, and will occur when we as a society commit to it.

BIBLIOGRAPHY

Adlard, A. and Hazan, V. (1998). Speech Perception in Children With Specific Reading Difficulties (Dyslexia). *The Quarterly Journal of Experimental Psychology Section A*, 51(1):153–177.

Agus, T. R., Carrión-Castillo, A., Pressnitzer, D., and Ramus, F. (2013). Perceptual Learning of Acoustic Noise by Dyslexic Individuals. *Journal of speech, language, and hearing research : JSLHR*, pages 1092–4388.

Ahissar, M. (2007). Dyslexia and the anchoring-deficit hypothesis. *Trends in Cognitive Sciences*, 11(11):458–465.

Ahissar, M., Lubin, Y., Putter-Katz, H., and Banai, K. (2006). Dyslexia and the failure to form a perceptual anchor. *Nature Neuroscience*, 9(12):1558–1564.

Alexander, L. M., Escalera, J., Ai, L., Andreotti, C., Febre, K., Mangone, A., Vega-Potler, N., Langer, N., Alexander, A., Kovacs, M., Litke, S., O'Hagan, B., Andersen, J., Bronstein, B., Bui, A., Bushey, M., Butler, H., Castagna, V., Camacho, N., Chan, E., Citera, D., Clucas, J., Cohen, S., Dufek, S., Eaves, M., Fradera, B., Gardner, J., Grant-Villegas, N., Green, G., Gregory, C., Hart, E., Harris, S., Horton, M., Kahn, D., Kabotyanski, K., Karmel, B., Kelly, S. P., Kleinman, K., Koo, B., Kramer, E., Lennon, E., Lord, C., Mantello, G., Margolis, A., Merikangas, K. R., Milham, J., Minniti, G., Neuhaus, R., Levine, A., Osman, Y., Parra, L. C., Pugh, K. R., Racanello, A., Restrepo, A., Saltzman, T., Septimus, B., Tobe, R., Waltz, R., Williams, A., Yeo, A., Castellanos, F. X., Klein, A., Paus, T., Leventhal, B. L., Craddock, R. C., Koplewicz, H. S., and Milham, M. P. (2017). Data Descriptor: An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Scientific Data*.

Amitay, S., Ben-Yehudah, G., Banai, K., and Ahissar, M. (2002). Disabled readers suffer from visual and auditory impairments but not from a specific magnocellular deficit. *Brain : a journal of neurology*, 125(Pt 10):2272–2285.

Andruski, J. E., Blumstein, S. E., and Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition*, 52(3):163–187.

Banai, K. and Ahissar, M. (2004). Poor frequency discrimination probes dyslexics with particularly impaired working memory. *Audiology and Neuro-Otology*, 9(6):328–340.

Banai, K. and Ahissar, M. (2006). Auditory processing deficits in dyslexia: Task or stimulus related? *Cerebral Cortex*, 16(12):1718–1728.

Baron, R. and Kenny, D. (1986). The moderator-mediator variable distinction in social psychological research. *Journal of personality and social psychology*, 51(6):1173–1182.

Bates, D., Maechler Martin, and Walker, S. (2016). Package 'lme4'.

Blomert, L. and Mitterer, H. (2004). The fragile nature of the speech-perception deficit in dyslexia: Natural vs. synthetic speech. *Brain and Language*, 89(1):21–26.

Boersma, P. and Weenink, D. (2019). Praat: doing phonetics by computer.

Boets, B., Beeck, H. P. O. D., Vandermosten, M., Scott, S. K., Gillebert, C. R., Mantini, D., Bulthé, J., and Sunaert, S. (2013). Intact But Less Accessible Phonetic Representations in Adults with Dyslexia. *Science*, 342(December):1251–1255.

Boets, B., Smedt, B., Cleuren, L., Vandewalle, E., Wouters, J., and Ghesquière, P. (2010). Towards a further characterization of phonological and literacy problems in Dutch-speaking children with dyslexia. *British Journal of Developmental Psychology*, 28(1):5–31.

Boets, B., Vandermosten, M., Poelmans, H., Luts, H., Wouters, J., and Ghesquière, P. (2011). Preschool impairments in auditory processing and speech perception uniquely predict future reading problems. *Research in Developmental Disabilities*, 32(2):560–570.

Boets, B., Wouters, J., van Wieringen, A., and Ghesquière, P. (2007). Auditory processing, speech perception and phonological ability in pre-school children at high-risk for dyslexia: A longitudinal study of the auditory temporal processing theory. *Neuropsychologia*, 45(8):1608–1620.

Bogliotti, C., Serniclaes, W., Messaoud-Galusi, S., and Sprenger-Charolles, L. (2008). Discrimination of speech sounds by children with dyslexia: Comparisons with chronological age and reading level controls. *Journal of Experimental Child Psychology*, 101(2):137–155.

Booth, J. R., Perfetti, C. A., MacWhinney, B., and Hunt, S. B. (2000). The Association of Rapid Temporal Perception With Orthographic and Phonological Processing in Children and Adults With Reading Impairment. *Scientific Studies of Reading*, 4(2):101–132.

Bradley, L. and Bryant, P. E. (1983). Categorizing sounds and learning to read - A causal connection. *Nature*, 301(5899):419–421.

Bradlow, A. R., Kraus, N., Nicol, T. G., McGee, T. J., Cunningham, J., Zecker, S. G., and Carrell, T. D. (1999). Effects of lengthened formant transition duration on discrimination and neural representation of synthetic CV syllables by normal and learning-disabled children. *The Journal of the Acoustical Society of America*, 106(4):2086.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4):433–436.

Brandt, J. and Rosen, J. J. (1980). Auditory phonemic perception in dyslexia: Categorical identification and discrimination of stop consonants. *Brain and Language*, 9(2):324–337.

Breier, J. I., Gray, L., Fletcher, J. M., Diehl, R. L., Klaas, P., Foorman, B. R., and Molis, M. R. (2001). Perception of Voice and Tone Onset Time Continua in Children with Dyslexia with and without Attention Deficit/Hyperactivity Disorder. *Journal of Experimental Child Psychology*, 80(3):245–270.

Bus, A. G. and van IJzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology*, 91(3):403–414.

Calcus, A., Deltenre, P., Colin, C., and Kolinsky, R. (2018). Peripheral and central contribution to the difficulty of speech in noise perception in dyslexic children. *Developmental Science*, 21(3).

Calcus, A., Lorenzi, C., Collet, G., Colin, C., and Kolinsky, R. (2016). Is There a Relationship Between Speech Identification in Noise and Categorical Perception in Children With Dyslexia? *Journal of Speech Language and Hearing Research*, 59(4):835.

Casini, L., Pech-Georgel, C., and Ziegler, J. C. (2018). It's about time: revisiting temporal processing deficits in dyslexia. *Developmental Science*, 21(2):e12530.

Cheung, H., Chung, K. K. H., Wong, S. W. L., McBride-Chang, C., Penney, T. B., and Ho, C. S. H. (2009). Perception of tone and aspiration contrasts in Chinese children with dyslexia. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 50(6):726–733.

Chiappe, P., Chiappe, D. L., and Siegel, L. S. (2001). Speech Perception, Lexicality, and Reading Skill. *Journal of Experimental Child Psychology*, 80(1):58–74.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., and Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, 108(3):804–809.

Cleary, M. and Pisoni, D. B. (2001). Speech perception and spoken word recognition: Research and theory. In Goldstein, E. B., Humphreys, G. W., Shiffrar, M., and Yost, W., editors, *Blackwell handbook of perception*, chapter 16, pages 499–534. Wiley-Blackwell.

Collet, G., Colin, C., Serniclaes, W., Hoonhorst, I., Markessis, E., Deltenre, P., and Leybaert, J. (2012). Effect of phonological training in French children with SLI: Perspectives on voicing identification, discrimination and categorical perception. *Research in Developmental Disabilities*, 33(6):1805–1818.

Colling, L. J., Noble, H. L., and Goswami, U. (2017). Neural entrainment and sensorimotor synchronization to the beat in children with developmental dyslexia: An EEG study. *Frontiers in Neuroscience*, 11:360.

Connine, C. M. (2004). It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin and Review*, 11(6):1084–1089.

Costello, A. B. and Osborne, J. W. (2005). Best Practices in Exploratory Factor Analysis: Four Recommendtions for Getting the Most From Your Analysis. *Practical Assessment, Research & Evaluation*.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., and Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, 16(5-6):507–534.

Dawes, P., Sirimanna, T., Burton, M., Vanniasegaram, I., Tweedy, F., and Bishop, D. V. M. (2009). Temporal auditory and visual motion processing of children diagnosed with auditory processing disorder and dyslexia. *Ear and Hearing*, 30(6):675–686.

de Gelder, B. and Vroomen, J. (1998). Impaired speech perception in poor readers: evidence from hearing and speech reading. *Brain and language*, 64(3):269–281.

Dillon, C. M., de Jong, K., and Pisoni, D. B. (2012). Phonological awareness, reading skills, and vocabulary knowledge in children who use cochlear implants. *Journal of Deaf Studies and Deaf Education*, 17(2):205–226.

Dole, M., Hoen, M., and Meunier, F. (2012). Speech-in-noise perception deficit in adults with dyslexia: Effects of background type and listening configuration. *Neuropsychologia*, 50(7):1543–1552.

Du, W. and Kelly, S. W. (2013). Implicit sequence learning in dyslexia: A within-sequence comparison of first- and higher-order information. *Annals of Dyslexia*, 63(2):154–170.

Erbeli, F., Hart, S. A., Wagner, R. K., and Taylor, J. (2018). Examining the Etiology of Reading Disability as Conceptualized by the Hybrid Model. *Scientific Studies of Reading*, 22(2):167–180.

Farmer, M. E. and Klein, R. M. (1995). The evidence for a temporal processing deficit linked to dyslexia: A review. *Psychonomic Bulletin & Review*, 2(4):460–493.

Ferguson, E. and Cox, T. (1993). Exploratory Factor Analysis: A UsersGuide. *International Journal of Selection and Assessment*.

Filzmoser, P. (2004). A multivariate outlier detection method. Technical report.

Frey, A., François, C., Chobert, J., Besson, M., and Ziegler, J. C. (2019a). Behavioral and electrophysiological investigation of speech perception deficits in silence, noise and envelope conditions in developmental dyslexia. *Neuropsychologia*, 130:3–12.

Frey, A., François, C., Chobert, J., Velay, J. L., Habib, M., and Besson, M. (2019b). Music training positively influences the preattentive perception of voice onset time in children with dyslexia: A longitudinal study. *Brain Sciences*, 9(4):91.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of statistical software*.

Fry, D. B. (1975). Simple Reaction-Times to Speech and Non-Speech Stimuli. *Cortex*, 11(4):355–360.

Gaab, N., Gabrieli, J. D. E., Deutsch, G. K., Tallal, P., and Temple, E. (2007). Neural correlates of rapid auditory processing are disrupted in children with developmental dyslexia and ameliorated with training: An fMRI study. *Restorative Neurology and Neuroscience*, 25:295–310.

Gabay, Y. and Holt, L. L. (2018). Short-term adaptation to sound statistics is unimpaired in developmental dyslexia. *PLoS ONE*, 13(6).

Gabay, Y., Thiessen, E. D., and Holt, L. L. (2015). Impaired Statistical Learning in Developmental Dyslexia. *Journal of Speech Language and Hearing Research*, 58(3):934.

Garnica, O. K. (1973). The development of phonemic speech perception. In *Cognitive development and the acquisition of language*, pages 215–222. Academic Press.

Germanò, E., Gagliano, A., and Curatolo, P. (2010). Comorbidity of ADHD and dyslexia. *Developmental Neuropsychology*, 35(5):475–493.

Gibson, L. Y., Hogben, J. H., and Fletcher, J. (2006). Visual and auditory processing and component reading skills in developmental dyslexia. *Cognitive Neuropsychology*, 23(4):621–642.

Godfrey, J. J., Syrdal-Lasky, a. K., Millay, K. K., and Knox, C. M. (1981). Performance of dyslexic children on speech perception tests. *Journal of experimental child psychology*, 32(3):401–424.

Gold, J. I. and Shadlen, M. N. (2007). The neural basis of decision making. *Annual review of neuroscience*, 30.

Gori, S., Seitz, A. R., Ronconi, L., Franceschini, S., and Facoetti, A. (2016). Multiple Causal Links between Magnocellular-Dorsal Pathway Deficit and Developmental Dyslexia. *Cerebral Cortex*, 26(11):4356–4369.

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences*, 15(1):3–10.

Goswami, U. (2015). Sensory theories of developmental dyslexia: Three challenges for research. *Nature Reviews Neuroscience*, 16(1):43–54.

Goswami, U., Thomson, J. M., Richardson, U., Stainthorp, R., Hughes, D., Rosen, S., and Scott, S. K. (2002). Amplitude envelope onsets and developmental dyslexia: A new hypothesis. *Proceedings of the National Academy of Sciences*, 99(16):10911–10916.

Gould, J. H. and Glencross, D. J. (1990). Do children with a specific reading disability have a general serial-ordering deficit? *Neuropsychologia*, 28(3):271–278.

Haft, S. L., Myers, C. A., and Hoeft, F. (2016). Socio-emotional and cognitive resilience in children with reading disabilities. *Current Opinion in Behavioral Sciences*, 10:133–141.

Hakvoort, B., de Bree, E., van Der Leij, A., Maassen, B., van Setten, E., Maurits, N., and van Zuijen, T. (2016). The role of categorical speech perception and phonological processing in familial risk children with and without dyslexia. *Journal of Speech, Language, and Hearing Research*, 59(6):1448–1460.

Halekoh, U. and Højsgaard, S. (2014). A Kenward-Roger Approximation and Parametric Bootstrap Methods for Tests in Linear Mixed Models - The R Package pbkrtest. *Journal of Statistical Software*, 59(9).

Hämäläinen, J. A., Salminen, H. K., and Leppänen, P. H. T. (2013). Basic Auditory Processing Deficits in Dyslexia. *Journal of Learning Disabilities*, 46(5):413–427.

Hancock, R., Pugh, K. R., and Hoeft, F. (2017). Neural Noise Hypothesis of Developmental Dyslexia. *Trends in Cognitive Sciences*, 21(6):434–448.

Hayton, J. C., Allen, D. G., and Scarpello, V. (2004). Factor Retention Decisions in Exploratory Factor Analysis: A Tutorial on Parallel Analysis. *Organizational Research Methods*, 7(2):191–205.

Hazan, V., Messaoud-Galusi, S., Rosen, S., Nouwens, S., and Shakespeare, B. (2009). Speech Perception Abilities of Adults With Dyslexia: Is There Any Evidence for a True Deficit? *Journal of Speech, Language, and Hearing Research*, 52(6):1510–1529.

Ho, C. S. H., Chan, D. W. O., Tsang, S. M., and Lee, S. H. (2002). The cognitive profile and multiple-deficit hypothesis in Chinese developmental dyslexia. *Developmental psychology*, 38(4):543.

Holt, L. L. and Lotto, A. J. (2010). Speech perception as categorization. *Attention, Perception, and Psychophysics*, 72(5):1218–1227.

Huang-Pollock, C., Ratcliff, R., McKoon, G., Shapiro, Z., Weigard, A., and Galloway-Long, H. (2017). Using the Diffusion Model to Explain Cognitive Deficits in Attention Deficit Hyperactivity Disorder. *Journal of Abnormal Child Psychology*, 45(1):57–68.

Huettig, F. and Pickering, M. J. (2019). Literacy Advantages Beyond Reading: Prediction of Spoken Language. *Trends in Cognitive Sciences*, 23(6):464–475.

Hulme, C., Hatcher, P. J., Nation, K., Brown, A., Adams, J., and Stuart, G. (2002). Phoneme awareness is a better predictor of early reading skill than onset-rime awareness. *Journal of Experimental Child Psychology*, 82(1):2–28.

Hulme, C., Nash, H. M., Gooch, D., Lervåg, A., and Snowling, M. J. (2015). The Foundations of Literacy Development in Children at Familial Risk of Dyslexia. *Psychological Science*, 26(12):1877–1886.

Huss, M., Verney, J. P., Fosker, T., Mead, N., and Goswami, U. (2011). Music, rhythm, rise time perception and developmental dyslexia: Perception of musical meter predicts reading and phonology. *Cortex*, 47(6):674–689.

Inácio, F., Faísca, L., Forkstam, C., Araújo, S., Bramão, I., Reis, A., and Petersson, K. M. (2018). Implicit sequence learning is preserved in dyslexic children. *Annals of Dyslexia*, 68(1):1–14.

Jaffe-Dax, S., Frenkel, O., and Ahissar, M. (2017). Dyslexics faster decay of implicit memory for sounds and words is manifested in their shorter neural adaptation. *eLife*, 6.

Jaffe-Dax, S., Lieder, I., Biron, T., and Ahissar, M. (2016). Dyslexics' usage of visual priors is impaired. *Journal of Vision*, 16(9):10.

Jiménez-Fernández, G., Vaquero, J. M., Jiménez, L., and Defior, S. (2011). Dyslexic children show deficits in implicit sequence learning, but not in explicit sequence learning or contextual cueing. *Annals of Dyslexia*, 61(1):85–110.

Joanisse, M. F., Manis, F. R., Keating, P., and Seidenberg, M. S. (2000). Language Deficits in Dyslexic Children: Speech Perception, Phonology, and Morphology. *Journal of Experimental Child Psychology*, 77(1):30–60.

Johnson, E. P., Pennington, B. F., Lowenstein, J. H., and Nittrouer, S. (2011). Sensitivity to structure in the speech signal by children with speech sound disorder and reading disability. *Journal of Communication Disorders*, 44(3):294–314.

Joo, S. J., Donnelly, P. M., and Yeatman, J. D. (2017a). The causal relationship between dyslexia and motion perception reconsidered. *Scientific Reports*, 7(1).

Joo, S. J., White, A. L., Strodtman, D. J., and Yeatman, J. D. (2017b). Optimizing text for an individual s visual system : The contribution of crowding to reading difficulties. *CORTEX*, 103(April):1–27.

Kline, R. (2013). Exploratory and confirmatory factor analysis. In Y. Petscher and C. Schatsschneider, editors, *Applied Quantitative Analysis in Education and the Social Sciences*. Routledge, New York.

Krause, M. B. (2015). Pay Attention!: Sluggish Multisensory Attentional Shifting as a Core Deficit in Developmental Dyslexia. *Dyslexia*, 21(4):285–303.

Law, J. M., Vandermosten, M., Ghesquière, P., and Wouters, J. (2014). The relationship of phonological ability, speech perception, and auditory perception in adults with dyslexia. *Frontiers in Human Neuroscience*, 8.

Lehongre, K., Ramus, F., Villiermet, N., Schwartz, D., and Giraud, A. L. (2011). Altered low-gamma sampling in auditory cortex accounts for the three main facets of dyslexia. *Neuron*, 72(6):1080–1090.

Liberman, A. M. (1970). Some characteristics of perception in the speech mode. *Research publications - Association for Research in Nervous and Mental Disease.*

Liberman, A. M. and Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21(1):1–36.

Lieder, I., Adam, V., Frenkel, O., Jaffe-Dax, S., Sahani, M., and Ahissar, M. (2019). Perceptual bias reveals slow-updating in autism and fast-forgetting in dyslexia. *Nature Neuroscience*, 22(2):256.

Light, J. G., Pennington, B. F., Gilger, J. W., and DeFries, . J. C. (1995). Reading disability and hyperactivity disorder : Evidence for a common genetic etiology. *Developmental Neuropsychology*, 11(3):323–335.

Lorenzi, C., Dumont, A., and Füllgrabe, C. (2000). Use of temporal envelope cues by children with developmental dyslexia. *Journal of Speech, Language, and Hearing Research*, 43(6):1367–1379.

Lyon, R. G., Shaywitz, S. E., and Shaywitz, B. A. (2003). A definition of dyslexia. *Annals of Dyslexia*, 53(1):1–14.

Maassen, B., Groenen, P., Crul, T., Assman-Hulsmans, C., and Gabreëls, F. (2001). Identification and discrimination of voicing and place-of-articulation in developmental dyslexia. *Clinical Linguistics and Phonetics*, 15(4):319–339.

Manis, F. R., McBride-Chang, C., Seidenberg, M. S., Keating, P., Doi, L. M., Munson, B., and Petersen, A. (1997). Are Speech Perception Deficits Associated with Developmental Dyslexia? *Journal of Experimental Child Psychology*, 66(2):211–235.

Marshall, C. M., Snowling, M. J., and Bailey, P. J. (2001). Rapid Auditory Processing and Phonological Ability in Normal Readers and Readers With Dyslexia. *Journal of Speech, Language, and Hearing Research*, 44(4):925–940.

Marslen-Wilson, W. and Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, 101(4):653–675.

Maye, J., Werker, J. F., and Gerken, L. A. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82(3):B101–B111.

McAnally, K. I. and Stein, J. F. (1997). Scalp potentials evoked by amplitude-modulated tones in dyslexia. *Journal of speech, language, and hearing research : JSLHR*, 40:939–945.

McLennan, C., Luce, P., and Charles-Luce, J. (2003). Representation of lexical form. *J Exp Psychol Learn Mem Cogn*, 29(4):539–553.

McMurray, B., Tanenhaus, M. K., and Aslin, R. N. (2009). Within-category VOT affects recovery from “lexical” garden-paths: Evidence against phoneme-level inhibition. *Journal of Memory and Language*, 60(1):65–91.

McMurray, B., Tanenhaus, M. K., Aslin, R. N., and Spivey, M. J. (2003). Probabilistic constraint satisfaction at the lexical/phonetic interface: Evidence for gradient effects of within-category VOT on lexical access. In *Journal of Psycholinguistic Research*, volume 32, pages 77–97.

McQueen, J. M., Norris, D., and Cutler, A. (1999). Lexical influence in phonetic decision making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance*, 25(5):1363–1389.

Menell, P., McAnally, K. I., and Stein, J. F. (1999). Psychophysical sensitivity and physiological response to amplitude modulation in adult dyslexic listeners. *Journal of Speech, Language, and Hearing*, 42(4):797–803.

Menghini, D., Carlesimo, G. A., Marotta, L., Finzi, A., and Vicari, S. (2010). Developmental dyslexia and explicit long-term memory. *Dyslexia (Chichester, England)*, 16(3):213–225.

Merzenich, M. M., Jenkins, W. M., Johnston, P., Schreiner, C., Miller, S. L., and Tallal, P. (1996). Temporal processing deficits of language-learning impaired children ameliorated by training. *Science*, 271(5245):77–81.

Messaoud-Galusi, S., Hazan, V., and Rosen, S. (2011). Investigating Speech Perception in Children With Dyslexia: Is There Evidence of a Consistent Deficit in Individuals? *Journal of Speech Language and Hearing Research*, 54(6):1682.

Mody, M., Studdert-Kennedy, M., and Brady, S. (1997). Speech perception deficits in poor readers: auditory processing or phonological coding? *Journal of experimental child psychology*, 64(2):199–231.

Moulines, . and Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, 9(5-6):453–467.

Muter, V. and Snowling, M. J. (2009). Children at familial risk of dyslexia: Practical implications from an at-risk study. *Child and Adolescent Mental Health*, 14(1):37–41.

Narayanan, A. (2019). How to recognize AI snake oil.

Nicolson, R. I. and Fawcett, A. J. (2018). Procedural learning, dyslexia and delayed neural commitment. In Lachmann, T. and Weis, T., editors, *Reading and Dyslexia*, pages 235–269. Springer, Cham.

Nittrouer, S. (1992). Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. *Journal of Phonetics*, 20:351–382.

Nittrouer, S. (1999). Do Temporal Processing Deficits Cause Phonological Processing Problems? *Journal of Speech Language and Hearing Research*, 42(4):925.

Noordenbos, M. W., Segers, E., Serniclaes, W., and Verhoeven, L. (2013). Neural evidence of the allophonic mode of speech perception in adults with dyslexia. *Clinical Neurophysiology*, 124(6):1151–1162.

Noordenbos, M. W. and Serniclaes, W. (2015). The Categorical Perception Deficit in Dyslexia: A Meta-Analysis. *Scientific Studies of Reading*, 19(5):340–359.

O'Brien, G. E., McCloy, D. R., Kubota, E. C., and Yeatman, J. D. (2018). Reading ability and phoneme categorization. *Nature Scientific Reports*, 8(1):16842.

O'Brien, G. E., McCloy, D. R., and Yeatman, J. D. (2019). Categorical phoneme labeling in children with dyslexia does not depend on stimulus duration. *The Journal of the Acoustical Society of America*, 146(1):245–255.

Palmer, J., Huk, A. C., and Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision*, 5(5):1–1.

Paul, I., Bott, C., Heim, S., Wienbruch, C., and Elbert, T. R. (2006). Phonological but not auditory discrimination is impaired in dyslexia. *European Journal of Neuroscience*, 24(10):2945–2953.

Pennington, B. F. (2006). From single to multiple deficit models of developmental disorders. *Cognition*, 101(2):385–413.

Pennington, B. F., Santerre-Lemmon, L., Rosenberg, J., MacDonald, B., Boada, R., Friend, A., Leopold, D. R., Samuelsson, S., Byrne, B., Willcutt, E. G., and Olson, R. K. (2012). Individual prediction of dyslexia by single versus multiple deficit models. *Journal of abnormal psychology*, 121(1):212.

Perrachione, T. K., Del Tufo, S. N., Winter, R., Murtagh, J., Cyr, A., Chang, P., Halverson, K., Ghosh, S. S., Christodoulou, J. A., and Gabrieli, J. D. (2016). Dysfunction of Rapid Neural Adaptation in Dyslexia. *Neuron*.

Peters, L. and Ansari, D. (2019). Are specific learning disorders truly specific, and are they disorders? *Trends in Neuroscience and Education*, 17:100115.

Peterson, R. L. and Pennington, B. F. (2015a). Developmental Dyslexia. *Annual Review of Clinical Psychology*, 11(1):283–307.

Peterson, R. L. and Pennington, B. F. (2015b). Developmental dyslexia. *Annual review of clinical psychology*, 11:283–307.

Piasta, S. B. and Wagner, R. K. (2010). Developing Early Literacy Skills: A Meta-Analysis of Alphabet Learning and Instruction. *Reading Research Quarterly*, 45(1):8–38.

Pisoni, D. B., Aslin, R. N., Perey, A. J., and Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts. *Journal of Experimental Psychology*, 8(2):297–314.

Pisoni, D. B. and Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, 15(2):285–290.

Poelmans, H., Luts, H., Vandermosten, M., Boets, B., Ghesquière, P., and Wouters, J. (2011). Reduced sensitivity to slow-rate dynamic auditory information in children with dyslexia. *Research in Developmental Disabilities*, 32(6):2810–2819.

Port, R. (2007). How are words stored in memory? Beyond phones and phonemes. *New Ideas in Psychology*, 25(2):145–172.

Port, R. F. and Leary, A. P. (2005). Against Formal Phonology. *Language*, 81(4):927–964.

Ramus, F. (2003). Developmental dyslexia: Specific phonological deficit or general sensorimotor dysfunction? *Current Opinion in Neurobiology*, 13(2):212–218.

Ramus, F. and Ahissar, M. (2012). Developmental dyslexia: The difficulties of interpreting poor performance, and the importance of normal performance. *Cognitive Neuropsychology*, 29(1-2):104–122.

Ramus, F. and Szenkovits, G. (2008). What phonological deficit? In *Quarterly Journal of Experimental Psychology*, volume 61, pages 129–141.

Ratcliff, R., Love, J., Thompson, C. A., and Opfer, J. E. (2012). Children are not like older adults: A diffusion model analysis of developmental changes in speeded responses. *Child Development*, 83(1):367–381.

Ratcliff, R. and McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, 204(4):873–922.

Ratcliff, R., McKoon, G., and Gomez, P. (2004). A Diffusion Model Account of the Lexical Decision Task. *Psychological Review*, 111(1):159.

Reardon, S. F. (2016). School district socioeconomic status, race, and academic achievement. Technical report, Stanford Center for Educational Policy Analysis, Stanford.

Reed, M. A. (1989). Speech perception and the discrimination of brief auditory cues in reading disabled children. *Journal of Experimental Child Psychology*, 48(2):270–292.

Repp, B. H. (1981). Perceptual equivalence of two kinds of ambiguous speech stimuli. *Bulletin of the Psychonomic Society*, 18(1):12–14.

Richardson, U., Thomson, J. M., Scott, S. K., and Goswami, U. (2004). Auditory processing skills and phonological representation in dyslexic children. *Dyslexia*, 10(3):215–233.

Rimrodt, S. L., Peterson, D. J., Denckla, M. B., Kaufmann, W. E., and Cutting, L. E. (2010). White matter microstructural differences linked to left perisylvian language network in children with dyslexia. *Cortex*, 46(6):739–749.

Roach, N. W., Edwards, V. T., and Hogben, J. H. (2004). The tale is in the tail: An alternative hypothesis for psychophysical performance variability in dyslexia. *Perception*, 33(7):817–830.

Robertson, E. K., Joanisse, M. F., Desroches, A. S., and Ng, S. (2009). Categorical speech perception deficits distinguish language and reading impairments in children. *Developmental Science*, 12(5):753–767.

Rocheron, I., Lorenzi, C., Füllgrabe, C., and Dumont, A. (2002). Temporal envelope perception in dyslexic children. *Neuroreport*, 13(13):1683–7.

Rosen, S. (2003). Auditory processing in dyslexia and specific language impairment: Is there a deficit? What is its nature? Does it explain anything? *Journal of Phonetics*, 31(3-4):509–527.

Rosen, S. and Mangari, E. (2001). Speech and Nonspeech Auditory Processing in Children With Dyslexia ? *Journal of speech, language and hearing research*, 44(August):720–736.

Roth, E. and Schneider, W. (2001). The effectiveness of kindergarten programs which aim at preventing reading and spelling problems in school: A comparison of three different approaches. *Psychology: The Journal of the Hellenic Psychological Society*, 8(3):313–392.

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274(5294):1926–198.

Samara, A. and Caravolas, M. (2017). Artificial Grammar Learning in Dyslexic and Nondyslexic Adults: Implications for Orthographic Learning. *Scientific Studies of Reading*, 21(1):76–97.

Schatschneider, C., Wagner, R. K., Hart, S. A., and Tighe, E. L. (2016). Using simulations to investigate the longitudinal stability of alternative schemes for classifying and identifying children with reading disabilities. *Scientific Studies of Reading*, 201(1):34–48.

Schrank, F. A., McGrew, K. S., Mather, N., Wendling, B. J., and LaForte, E. M. (2014). Woodcock-Johnson IV Tests of Achievement.

Schulte-Körne, G., Deimel, W., Bartling, J., and Remschmidt, H. (1998). Role of auditory temporal processing for reading and spelling disability. *Perceptual and motor skills*, 86(3 Pt 1):1043–1047.

Schütt, H., Harmeling, S., Macke, J. H., and Wichmann, F. A. (2015). Psignifit 4: Pain-free Bayesian Inference for Psychometric Functions. *Journal of vision*, 15(12):474.

Serniclaes, W., Sprenger-charolles, L., Carré, R., and Demonet, J.-F. (2001). Perceptual Discrimination of. *Journal of Speech Language and Hearing Research*, 44(April 2001):384–399.

Serniclaes, W., Van Heghe, S., Mousty, P., Carré, R., and Sprenger-Charolles, L. (2004). Allophonic mode of speech perception in dyslexia. *Journal of Experimental Child Psychology*, 87(4):336–361.

Serniclaes, W., Ventura, P., Morais, J., and Kolinsky, R. (2005). Categorical perception of speech in illiterate adults. *Cognition*, 98(2):B35–B44.

Shadlen, M. N., Hanks, T. D., Churchland, A. K., Kiani, R., and Yang, T. (2013). The Speed and Accuracy of a Simple Perceptual Decision: A Mathematical Primer. In Doya Kenji, Ishii Shin, Pouget Alexandre, and Rao P. N. Rajesh, editors, *Bayesian Brain*, chapter 10, pages 207–236. MIT Press, Cambridge, Massachusetts.

Shadlen, M. N. and Newsome, W. T. (2001). Neural Basis of a Perceptual Decision in the Parietal Cortex (Area LIP) of the Rhesus Monkey. *Journal of Neurophysiology*, 86:1916–1936.

Shaywitz, B. A., Shaywitz, S. E., Pugh, K. R., Mencl, W., Fulbright, R. K., Skudlarski, P., Constable, R., Marchione, K. E., Fletcher, J. M., Lyon, G. R., and Gore, J. C. (2002). Disruption of posterior brain systems for reading in children with developmental dyslexia. *Biological Psychiatry*, 52(2):101–110.

Shaywitz, S. E. (1998). Dyslexia. *The New England journal of medicine*, 338(5):307–312.

Shaywitz, S. E., Escobar, M. D., Shaywitz, B. A., Fletcher, J. M., and Makuch, R. (1992a). Evidence That Dyslexia May Represent the Lower Tail of a Normal Distribution of Reading Ability. *New England Journal of Medicine*, 326(3):145–150.

Shaywitz, S. E., Shaywitz, B. A., Fletcher, J. M., and Makuch, R. (1992b). Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability. *New England Journal of Medicine*.

Siegel, L. S. and Ryan, E. B. (1989). Subtypes of developmental dyslexia: The influence of definitional variables. *Reading and Writing*, 1(3):257–287.

Smith, P. L. and Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends in Neurosciences*, 27(3):161–168.

Snowling, M. J. (1998). Dyslexia as a Phonological Deficit: Evidence and Implications. *Child Psychology and Psychiatry Review*, 3(1):4–11.

Snowling, M. J. (2000). *Dyslexia*. Wiley-Blackwell, Hoboken, New Jersey, 2 edition.

Snowling, M. J., Lervåg, A., Nash, H. M., and Hulme, C. (2019). Longitudinal relationships between speech perception, phonological skills and reading in children at high-risk of dyslexia. *Developmental Science*.

Spencer, M., Wagner, R. K., Schatschneider, C., Quinn, J. M., Lopez, D., and Petscher, Y. (2014). Incorporating RTI in a hybrid model of reading disability. *Learning Disability Quarterly*.

Sperling, A. J., Lu, Z. L., Manis, F. R., and Seidenberg, M. S. (2005). Deficits in perceptual noise exclusion in developmental dyslexia. *Nature Neuroscience*.

Staels, E. and Van den Broeck, W. (2015). No solid empirical evidence for the SOLID (serial order learning impairment) hypothesis of dyslexia. *Journal of Experimental Psychology: Learning Memory and Cognition*.

Steffens, M. L., Eilers, R. E., Gross-Glenn, K., and Jallad, B. (1992). Speech perception in adult subjects with familial dyslexia. *Journal of speech and hearing research*, 35(1):192–200.

Stein, J. (2001). The Magnocellular Theory of Developmental Dyslexia. *Dyslexia*, 7(1):12–36.

Stein, J. (2018). The current status of the magnocellular theory of developmental dyslexia. *Neuropsychologia*, 130:66–77.

Stein, J. and Talcott, J. (1999). Impaired neuronal timing in developmental dyslexia - The magnocellular hypothesis. *Dyslexia*, 5(2):59–77.

Stein, J. and Walsh, V. (1997). To see but not to read; the magnocellular theory of dyslexia. *Trends in Neurosciences*, 204(4):147–152.

Steinbrink, C., Zimmer, K., Lachmann, T., Dirichs, M., and Kammer, T. (2014). Development of rapid temporal processing and its impact on literacy skills in primary school children. *Child Development*, 85(4):1711–1726.

Stevenson, J., Langley, K., Pay, H., Payton, A., Worthington, J., Ollier, W., and Thapar, A. (2005). Attention deficit hyperactivity disorder with reading disabilities: Preliminary genetic findings on the involvement of the ADRA2A gene. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 46(10):1081–1088.

Stollman, M. H., Kapteyn, T. S., and Sleeswijk, B. W. (1994). Effect of time-scale modification of speech on the speech recognition threshold in noise for hearing-impaired and language-impaired children. *Scandinavian Audiology*, 23(1):39–46.

Stoodley, C. J., Hill, P. R., Stein, J. F., and Bishop, D. V. M. (2006). Auditory event-related potentials differ in dyslexics even when auditory psychophysical performance is normal. *Brain Research*, 1121(1):190–199.

Strong, G., Torgerson, C., Torgerson, D., and Hulme, C. (2011). A systematic meta-analytic review of evidence for the effectiveness of the 'Fast ForWord' language intervention program. *Journal of child psychology and psychiatry, and allied disciplines*, 52(3):224–235.

Stuart, G. W., McAnally, K. I., McKay, A., Johnston, M., and Castles, A. (2006). A test of the magnocellular deficit theory of dyslexia in an adult sample. *Cognitive Neuropsychology*, 23(8):1215–1229.

Swanson, H. L. (1993). Working memory in learning disability subgroups. *Journal of Experimental Child Psychology*, 56(1):87–114.

Talcott, J. B., Witton, C., Hebb, G. S., Stoodley, C. J., Westwood, E. A., France, S. J., Hansen, P. C., and Stein, J. F. (2002). On the relationship between dynamic visual and auditory processing and literacy skills; results from a large primary-school study. *Dyslexia*, 8(4):204–225.

Talcott, J. B., Witton, C., Mclean, M. F., Hansen, P. C., Rees, A., Green, G. G. R., and Stein, J. F. (2000). Dynamic sensory sensitivity and children's word decoding skills. *Proc. Natl. Acad. Sci. U.S.A.*, 97(6):2952–2957.

Tallal, P. (1980). Auditory temporal perception, phonics, and reading disabilities in children. *Brain and Language*, 9(2):182–198.

Tallal, P., Miller, S. L., Bedi, G., Byma, G., Wang, X., Nagarajan, S. S., Schreiner, C., Jenkins, W. M., and Merzenich, M. M. (1996). Language comprehension in language-learning impaired children improved with acoustically modified speech. *Science*, 271(5245):81–84.

Thomson, J. M., Fryer, B., Maltby, J., and Goswami, U. (2006). Auditory and motor rhythm awareness in adults with dyslexia. *Journal of Research in Reading*, 29(3):334–348.

Thomson, J. M. and Goswami, U. (2008). Rhythmic processing in children with developmental dyslexia: Auditory and motor rhythms link to reading and spelling. *Journal of Physiology Paris*, 102(1-3):120–129.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., and Imai, K. (2014). mediation : R Package for Causal Mediation Analysis. *Journal of Statistical Software*.

Torgesen, J. K. (2000). Individual Differences in Response to Early Interventions in Reading: The Lingering Problem of Treatment Resisters. *Learning Disabilities Research and Practice*.

Torgesen, J. K., Wagner, R., and Rashotte, C. (2011). *TOWRE 2: Test of word reading efficiency*. Pearson Clinical Assessment.

Toscano, J. C., McMurray, B., Dennhardt, J., and Luck, S. J. (2010). Continuous Perception and Graded Categorization. *Psychological Science*, 21(10):1532–1540.

Treutwein, B. and Strasburger, H. (1999). Fitting the psychometric function. *Perception and Psychophysics*, 61(1):87–106.

Van Beinum, F. J., Schwippert, C. E., Been, P. H., Van Leeuwen, T. H., and Kuijpers, C. T. (2005). Development and application of a /bAk/-/dAk/ continuum for testing auditory perception within the Dutch longitudinal dyslexia study. *Speech Communication*, 47(1-2):124–142.

Van Ingelghem, M., Boets, B., Van Wieringen, A., Onghena, P., Ghesquière, P., and Wouters, J. (2005). An auditory temporal processing deficit in children with dyslexia. In Ghesquière, P. and Ruijssenaars, A. J., editors, *Learning disabilities: A challenge to teaching and instruction.*, chapter 3, pages 47–63. Leuven University Press.

Van Ingelghem, M., Van Wieringen, A., Wouters, J., Vandenbussche, E., Onghena, P., and Ghesquière, P. (2001). Psychophysical evidence for a general temporal processing deficit in children with dyslexia. *NeuroReport*, 12(16):3603–3607.

Van Zandt, T. (2011). How to fit a response time distribution. *Psychonomic Bulletin & Review*, 7(3):424–465.

Vandermosten, M., Boets, B., Luts, H., Poelmans, H., Golestani, N., Wouters, J., and Ghesquière, P. (2010). Adults with dyslexia are impaired in categorizing speech and nonspeech sounds on the basis of temporal cues. *Proceedings of the National Academy of Sciences*, 107(23):10389–10394.

Vandermosten, M., Boets, B., Luts, H., Poelmans, H., Wouters, J., and Ghesquière, P. (2011). Impairments in speech and nonspeech sound categorization in children with dyslexia are driven by temporal processing difficulties. *Research in Developmental Disabilities*, 32(2):593–603.

Vandermosten, M., Wouters, J., Ghesquière, P., and Golestani, N. (2018). Statistical Learning of Speech Sounds in Dyslexic and Typical Reading Children. *Scientific Studies of Reading*, 23(1):116–127.

Vargo, F. E., Grosser, G. S., and Spafford, C. S. (1995). Digit span and other WISC-R scores in the diagnosis of dyslexia in children. *Perceptual and motor skills*, 80(3 Pt 2):1219–29.

Vidyasagar, T. R. (2019). Visual attention and neural oscillations in reading and dyslexia: Are they possible targets for remediation? *Neuropsychologia*.

Waber, D., Weiler, M., Wolff, P., Bellinger, D., Marcus, D., Ariel, R., Forbes, P., and Wypij, D. (2001). Processing of rapid auditory stimuli in school-age children referred for evaluation of learning disorders. *Child Development*, 72(1):37–49.

Wagenmakers, E. J. and Brown, S. (2007). On the Linear Relation Between the Mean and the Standard Deviation of a Response Time Distribution. *Psychological Review*.

Wagenmakers, E. J., Grasman, R. P., and Molenaar, P. C. (2005). On the relation between the mean and the variance of a diffusion model response time distribution. *Journal of Mathematical Psychology*.

Wagner, R. K. and Torgesen, J. K. (1987). The Nature of Phonological Processing and Its Causal Role in the Acquisition of Reading Skills.

Wald, A. (1947). *Sequential Analysis*. Wiley.

Wandell, B. A. and Yeatman, J. D. (2013). Biological development of reading circuits. *Current Opinion in Neurobiology*, 23(2):261–268.

Wang, S. and Gathercole, S. E. (2013). Working memory deficits in children with reading difficulties: Memory span and dual task coordination. *Journal of Experimental Child Psychology*, 115(1):188–197.

Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*.

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence-Second Edition, manual.* Pearson Clinicial Assessments, 2 edition.

Whalen, D. H. (1991). Subcategorical phonetic mismatches and lexical access. *Perception & Psychophysics*, 50(4):351–360.

White, S., Milne, E., Rosen, S., Hansen, P., Swettenham, J., Frith, U., and Ramus, F. (2006). Target article with commentaries and response: The role of sensorimotor impairments in dyslexia: A multiple case study of dyslexic children. *Developmental Science*, 9(3):237–255.

Wichmann, F. A. and Hill, N. J. (2001a). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, 63(8):1293–1313.

Wichmann, F. A. and Hill, N. J. (2001b). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics*, 63(8):1314–1329.

Wiecki, T. V., Sofer, I., and Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in Neuroinformatics*.

Winn, M. B. and Litovsky, R. Y. (2015). Using speech sounds to test functional spectral resolution in listeners with cochlear implants. *The Journal of the Acoustical Society of America*, 137(3):1430–1442.

Witton, C., Stein, J. F., Stoodley, C. J., Rosner, B. S., and Talcott, J. B. (2002). Separate influences of acoustic AM and FM sensitivity on the phonological decoding skills of impaired and normal readers. *Journal of cognitive neuroscience*, 14(6):866–874.

Witton, C., Talcott, J. B., Hansen, P. C., Richardson, A. J., Griffiths, T. D., Rees, A., Stein, J. F., and Green, G. G. (1998). Sensitivity to dynamic auditory and visual stimuli predicts nonword reading ability in both dyslexic and normal readers. *Current Biology*, 8(14):791–797.

Wolf, M. and Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology*.

Wolf, M. and Bowers, P. G. (2000). Naming-speed processes and developmental reading disabilities: An introduction to the special issue on the double-deficit hypothesis. *Journal of learning disabilities*.

Wolf, M. and Katzir-Cohen, T. (2001). Reading Fluency and Its Intervention. *Scientific Studies of Reading*, 5(3):211–239.

Wolfgang, S., Ellen, R., and Marco, E. (2000). Training phonological skills and letter knowledge in children at risk for dyslexia: A comparison of three kindergarten intervention programs. *Journal of Educational Psychology*.

Wright, B. A., Lombardino, L. J., King, W. M., Puranik, C. S., Leonard, C. M., and Merzenich, M. M. (1997). Deficits in auditory temporal and spectral resolution in language-impaired children. *Nature*, 387(6629):176–178.

Yeatman, J. D., Dougherty, R. F., Ben-Shachar, M., and Wandell, B. A. (2012). Development of white matter and reading skills. *Proceedings of the National Academy of Sciences*.

Yeatman, J. D., Rauschecker, A. M., and Wandell, B. A. (2013). Anatomy of the visual word form area: Adjacent cortical circuits and long-range white matter connections. *Brain and Language*.

Zhang, Y., Zhang, L., Shu, H., Xi, J., Wu, H., Zhang, Y., and Li, P. (2012). Universality of categorical perception deficit in developmental dyslexia: An investigation of Mandarin Chinese tones. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 53(8):874–882.

Ziegler, J. C. (2008). Better to lose the anchor than the whole ship. *Trends in Cognitive Sciences*, 12(7):244–245.

Ziegler, J. C. and Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, 131(1):3.

Ziegler, J. C., Pech-Georgel, C., George, F., and Lorenzi, C. (2009). Speech-perception-in-noise deficits in dyslexia. *Developmental Science*, 12(5):732–745.

Ziegler, J. C., Perry, C., and Zorzi, M. (2019). Modeling the Variability of Developmental Dyslexia. In *Developmental Dyslexia across Languages and Writing Systems*, chapter 16, page 350. Cambridge University Press.

Appendix A
SUPPLEMENTARY MATERIAL FOR O'BRIEN ET AL. 2018

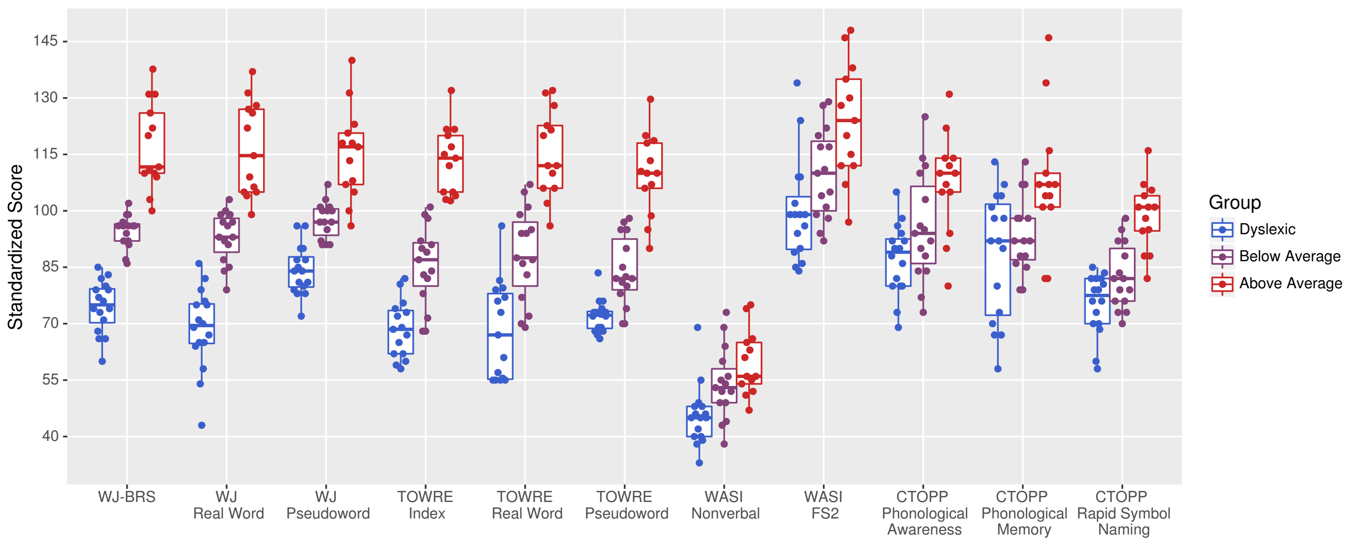
A.0.1 Stepwise model selection procedure

Fixed-effect predictors with sum coding were used for both the continuum (static /ʃa/∼/sa/ versus dynamic /ba/∼/da/) and paradigm (ABX versus single-stimulus) variables. Reading ability (WJ-BRS score) was entered as a continuous fixed-effect predictor except where otherwise stated. Nuisance predictors were added for presence/absence of ADHD diagnosis (treatment coding) and for non-verbal IQ (WASI-II Matrix Reasoning; continuous predictor). A random effect for participant was also included.

Model of psychometric slope

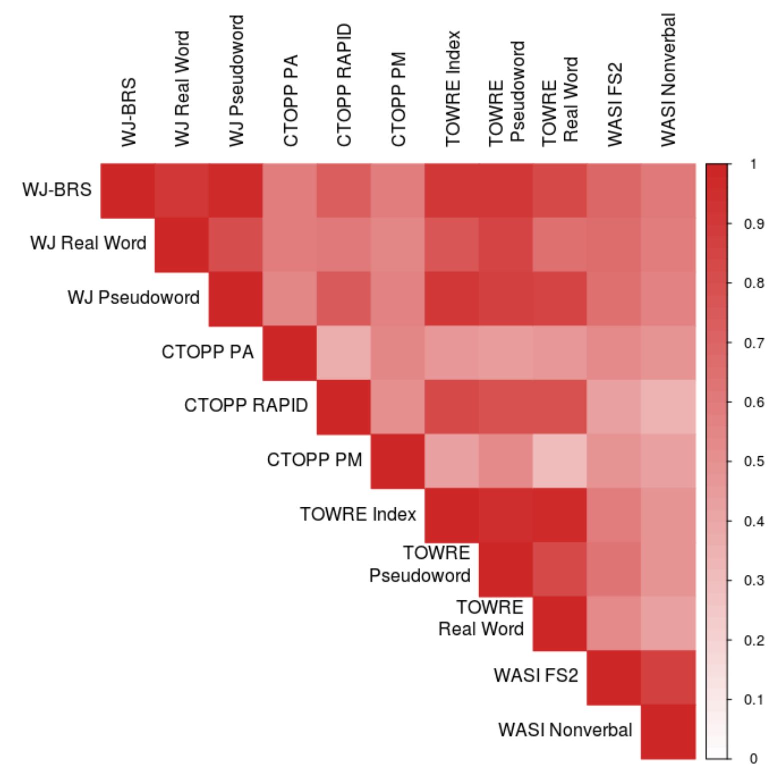
The fully-specified model of psychometric slope indicated that only one predictor—WJ-BRS reading score—showed a significant effect on slope estimates. To refine this model, first the nuisance parameters (ADHD diagnosis and nonverbal IQ) were each removed, and those subset models were tested against the full model using likelihood ratio tests and a conservative threshold for parameter exclusion of $p > 0.10$. The tests indicated no improvement in model fit from the inclusion of the nuisance predictors. Next, the predictors for continuum and paradigm (and their interactions with each other and with reading ability) were removed from the reduced model. Again, likelihood ratio tests indicated no significant improvement of model fit from the inclusion of those predictors, suggesting that the most parsimonious model of psychometric slope contained only a continuous predictor of reading ability (WJ-BRS) and a random intercept for each participant.

Figure A.1: Demographic information for 44 participants.



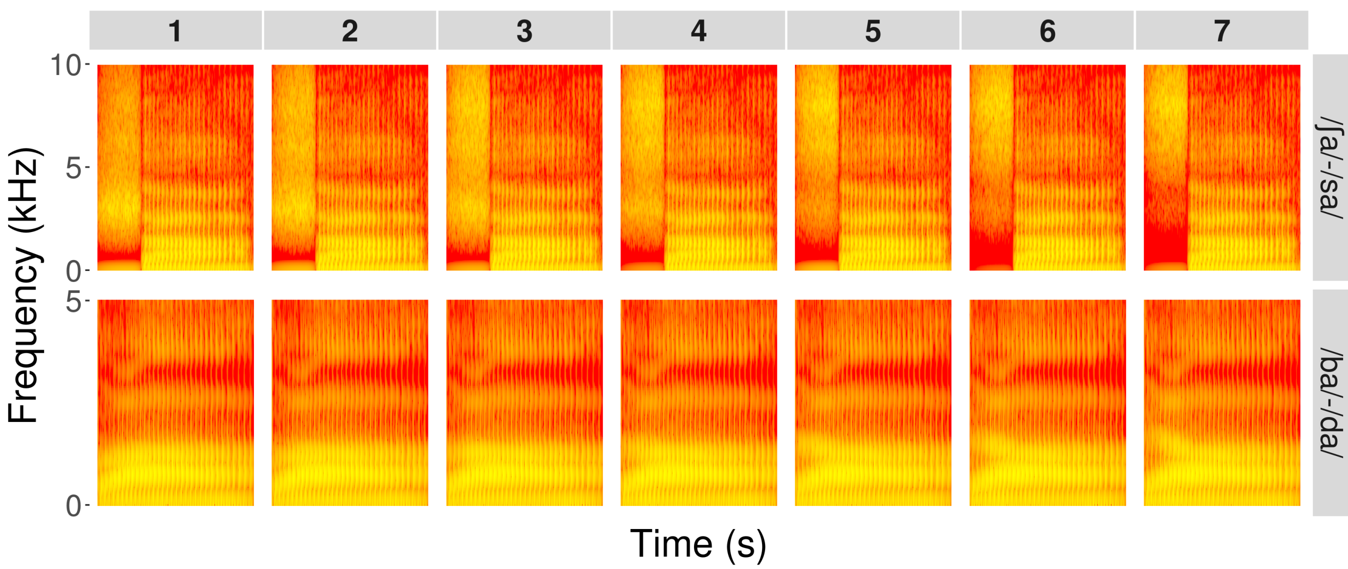
Demographic information for all 44 participants on the study on various behavioral measures, including the Woodcock-Johnson IV (WJ), Test of Word Reading Efficiency (TOWRE), Weschler Abbreviated Scale of Intelligence (WASI), and the Comprehensive Test of Phonological Processing (CTOPP). Each dot represents an individual subject, box plots denote the 25^{th}, 50^{th}, and 75^{th} percentiles, and color marks the group assigned to each subject. Note that all standardized scores are on the same scale ($\mu = 100$) except the WASI Nonverbal measure ($\mu = 50$).

Figure A.2: Correlations between behavioral measures.



Correlation plot for behavioral measures relating to reading, phonological processing, and cognitive abilities. All correlations are highly significant after correcting for multiple comparisons ($p < 1e-8$).

Figure A.3: Stimulus spectrograms.



Spectrograms of the endpoints of the two speech continua. Top row: the endpoints of the /ʃa/∼/sa/ continuum, which differ based on the relative amplitudes of spectral peaks in the initial fricative. Bottom row: the endpoints of the /ba/∼/da/ continuum, which differ based on starting frequency of the initial F2 transition. Note the different y-axis scales in the top and bottom rows.

Model of lapse rate

Following the same procedure as above, both nuisance predictors (ADHD diagnosis and non-verbal IQ) were eliminated, as were the predictor for continuum type and its interactions with reading ability and paradigm, and the interaction between reading ability and continuum. This yielded a final model of lapse rate containing a continuous predictor for reading ability, a categorical predictor for continuum, and a random effect for participant.

A.0.2 Quadratic and Linear Discriminant Analysis

To assess whether individuals in the Dyslexic group could be identified as such based on their psychometric functions, we first computed each subjects average psychometric function by averaging the slope and two asymptote parameters across all six blocks of the experiment. Using a leave-one-out approach, a quadratic discriminant analysis (or linear discriminant analysis, where noted) of group identity using slope and the average lapse rate was performed on a training set comprised of all data except the held-out point. Then the class of the held-out point was predicted. This was repeated holding out each subjects average psychometric function, and in this way every subject could be assigned a label without overfitting.

Appendix B

SUPPLEMENTARY MATERIAL FOR O'BRIEN ET AL. 2019

B.1 Supplemental Analysis: group level statistics

B.1.1 Drift rate

As in the main manuscript, we used mixed model selection to identify the most parsimonious model of drift rate. Subject was included as a random effect, as our data contained four drift rate estimates per participant (one at each coherence level). The selected model included significant main effects of stimulus coherence and age, plus a significant interaction of stimulus coherence and group (Table B.7). The main effect of group was not significant ($p = 0.117$), but the direction of the relationship was the same as in the model where reading is treated as a continuous variable. Therefore, our results from both models are in qualitative agreement. The fact that reading skill was a significant main effect as a continuous measure but not as a discrete one is likely the result of reduced statistical power: there are 90 subjects in the group analysis, but 104 in the continuous-measure analysis.

B.1.2 Decision criterion parameters

The parameter a, representing an individuals threshold of evidence for initiating a decision, was modeled as the dependent variable next. Linear mixed model selection dropped all three covariates, leaving only a main effect of group ($\beta = 0.351$, SE $= 0.130$, $p = 0.009$).

The parameter sz, representing trial-to-trial variability in the drift process starting point, was modeled similarly. The selected model contained only a main effect of group, although this effect missed the standard threshold of significance ($\beta = 0.105$, SE $= 0.060$, $p = 0.085$).

Table B.1: Correlations with composite reading score

Measure	β	SE	p
Age	1.08	1.40	0.443
ADHD diagnosis	-6.91	4.45	0.124
Gender	1.01	3.78	0.780
CTOPP-2			
Phonological Awareness	0.65	0.11	< 0.001
Phonological Memory	0.57	0.10	< 0.001
Rapid Automatic Naming	0.88	0.08	< 0.001
TOWRE-2			
Pseudoword Decoding	1.02	0.03	< 0.001
Sight Word Reading	0.82	0.03	< 0.001
WASI-II			
Full-scale IQ	0.75	0.09	< 0.001
Matrix Reasoning	0.95	0.16	< 0.001
Vocabulary	0.97	0.13	< 0.001
Woodcock Johnson IV			
Letter Word Identification	0.91	0.03	< 0.001
Word Attack	0.94	0.05	< 0.001

Table B.2: Group demographic measures

	Control	Dyslexic	Significance
	n =48	n =43	
	28 male, 20 female	25 male, 18 female	
CTOPP 2			
Phonological Awareness	98.3 (15.0)	86.4 (12.7)	< 0.001
Phonological Memory	98.6 (17.3)	84.9 (12.8)	< 0.001
Rapid Naming	99.4 (12.5)	79.0 (9.6)	< 0.001
TOWRE 2			
TOWRE Index	106.8 (11.0)	68.4 (8.0)	< 0.001
Pseudoword Decoding	104.6 (11.4)	71.6 (6.3)	< 0.001
Sight Word Efficiency	108.2 (11.7)	68.5 (11.3)	< 0.001
WASI-III			
Full-scale IQ	115.8 (16.0)	96.3 (9.9)	< 0.001
Matrix Reasoning	55.3 (10.8)	46.6 (7.3)	< 0.001
Vocabulary	63.2 (11.0)	49.1 (7.8)	< 0.001
Woodcock-Johnson IV			
Basic Reading Score	110.3 (12.7)	77.2 (10.5)	< 0.001
Letter Word Identification	109.8 (11.8)	74.4 (12.1)	< 0.001
Word Attack	109.4 (14.9)	82.4 (10.8)	< 0.001

Table B.3: DDM parametr reliability estimates

Parameter	Split-half reliability	Adjusted reliability
v_6	0.093	0.170
v_{12}	0.397	0.568
v_{24}	0.353	0.522
v_{48}	0.512	0.677
a	0.531	0.693
t	0.594	0.737
sv	0.141	0.248
sz	0.279	0.436

Split half reliability is calculated by partitioning each subjects responses into two by random assignment, estimating the DDM parameters on each half, and measuring the Pearsons correlation between parameter estimates. Adjusted reliability is calculated with the Spearman-Brown prophecy formula. This adjustment is an estimate of the reliability had the estimates been computed on twice as many observations as in split-half reliability.

Table B.4: Selected model of reaction time on the motion discrimination task

	β	SE	p
(Intercept)	4.101	0.316	< 0.001
Coherence	-0.173	0.009	< 0.001
Age	-0.059	0.021	0.007
Reading skill	-0.006	0.001	< 0.001

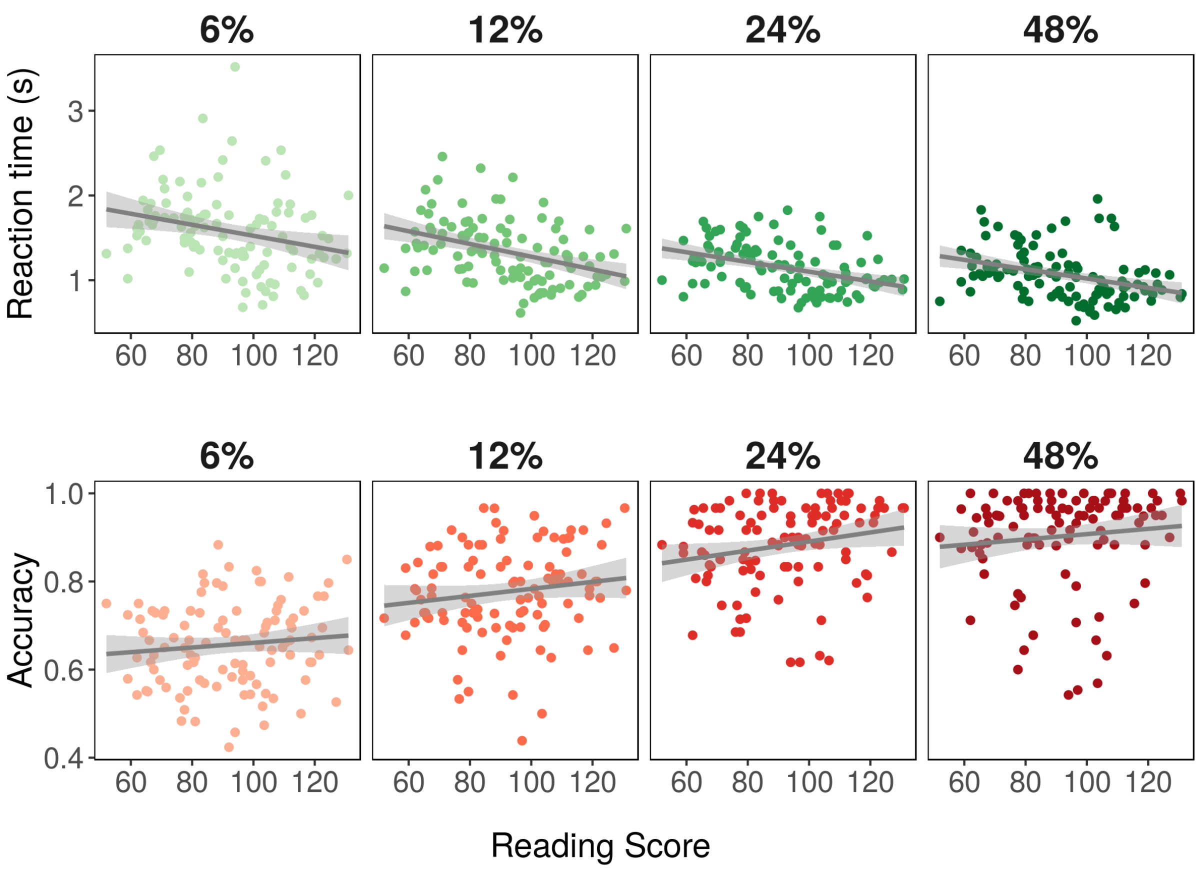
Table B.5: Selected model of accuracy on the motion discrimination task

	β	SE	p
(Intercept)	0.218	0.062	< 0.001
Coherence	0.084	0.003	< 0.001
Age	0.024	0.006	< 0.001

Table B.6: Selected model of the ratio of error-to-correct reponse times on the motion discrimination task.

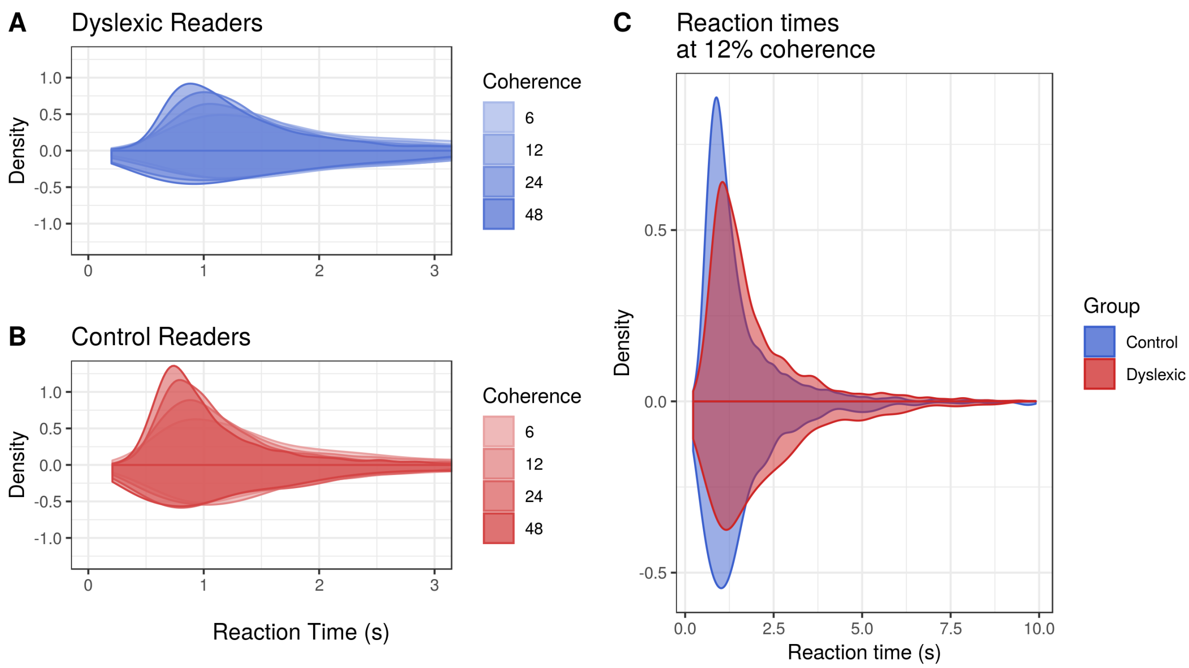
	β	SE	p
(Intercept)	0.555	0.350	0.115
Reading skill	0.004	0.002	0.050
Age	0.080	0.028	0.005
Nonverbal IQ	-0.007	0.004	0.096

Figure B.1: Reaction times and accuracy on random dot motion task.



Median reaction times (top row) and accuracy (bottom row) for each individual as a function of reading score. Panels show each stimulus coherence.

Figure B.2: Reaction time distributions for Control and Dyslexic groups.



Panels A-B: density plots of reaction times at each coherence level for the Dyslexic and Control groups. On the positive axis, correct response time distributions are shown, and on the negative axis error responses are shown. Plots are truncated at 3 seconds for ease of viewing. Panel C: overlaid density plots of reaction times at 12% coherence for the Dyslexic and Control groups.

Table B.7: Selected model of drift rate

	β	SE	p
(Intercept)	1.630	0.097	< 0.001
Coherence	0.772	0.027	< 0.001
Age	0.240	0.072	0.002
Group	-0.227	0.143	0.117
Coherence:Group	-0.133	0.053	0.014

Table B.8: Selected model of residual time t

	β	SE	p
(Intercept)	0.459	0.025	< 0.001
Group	0.092	0.039	0.021
Nonverbal IQ	0.032	0.020	0.104
Age	-0.31	0.018	0.085

Table B.9: Selected model of trial-to-trial variability in residual time st

	β	SE	p
(Intercept)	0.280	0.042	< 0.001
Group	0.185	0.062	0.004
Age	-0.064	0.031	0.041

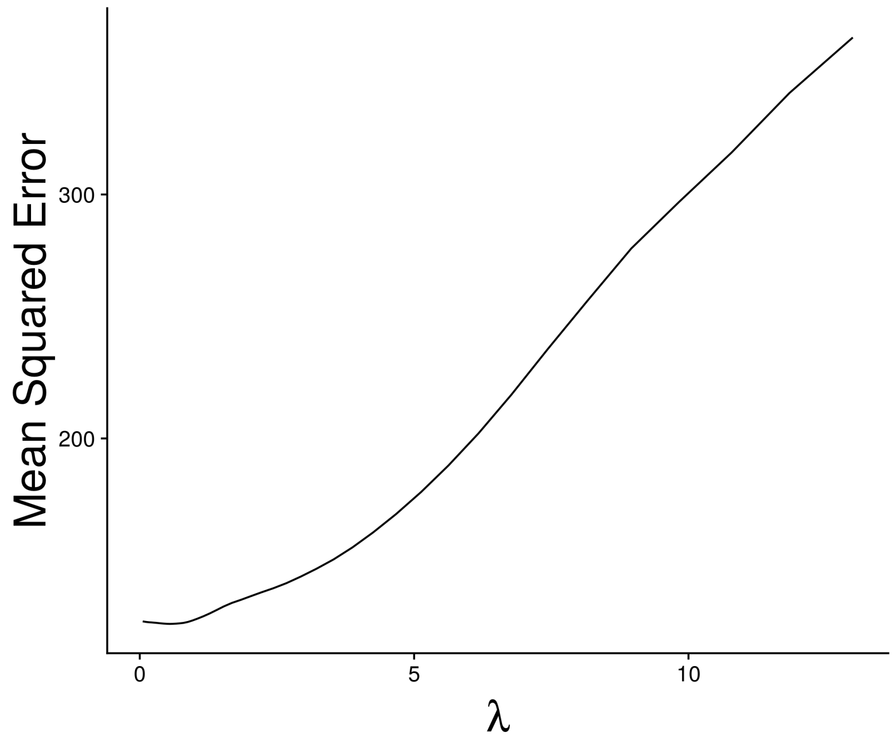
B.1.3 Non-decision time parameters

The selected model for residual non-decision time t contained three main effects: group, nonverbal IQ and age, of which only group fell below the standard threshold of significance (Table B.8).

Lastly, we considered the trial-to-trial variability in non-decision time st. The selected model contained two predictors: group and age at testing (Table S6).

We can therefore see that the relationships between estimated DDM parameters and reading skill are qualitatively consistent regardless of whether reading disability is treated as a categorical or continuous variable.

Figure B.3: Lasso regression mean squared error as a function of regularization.

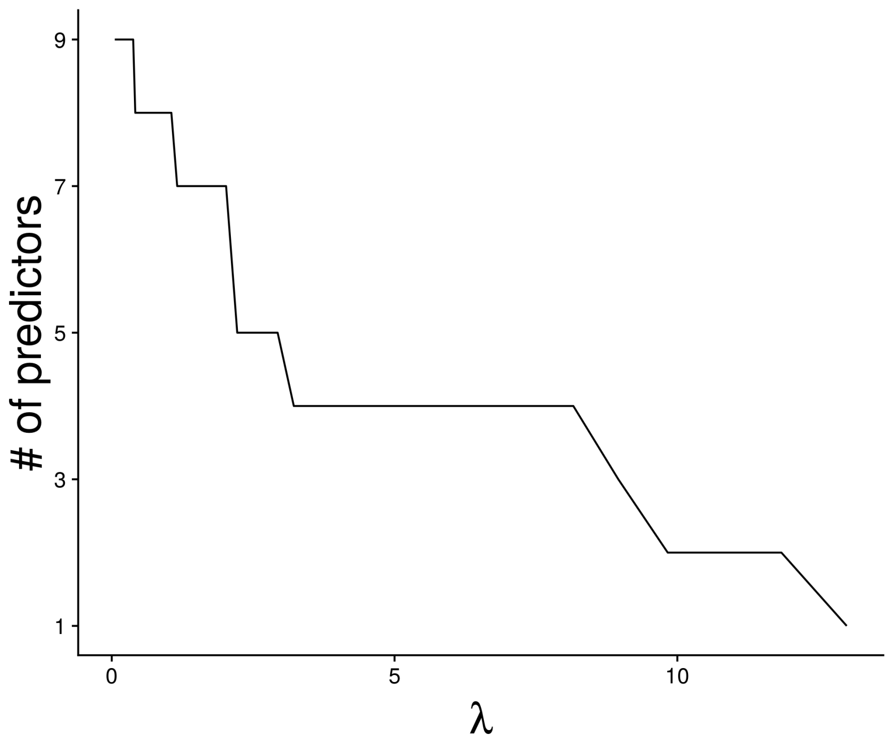


Mean squared error (MSE) of lasso regression as a function of the regularization parameter λ with 10-fold cross validation.

Table B.10: Lasso model of reading skill

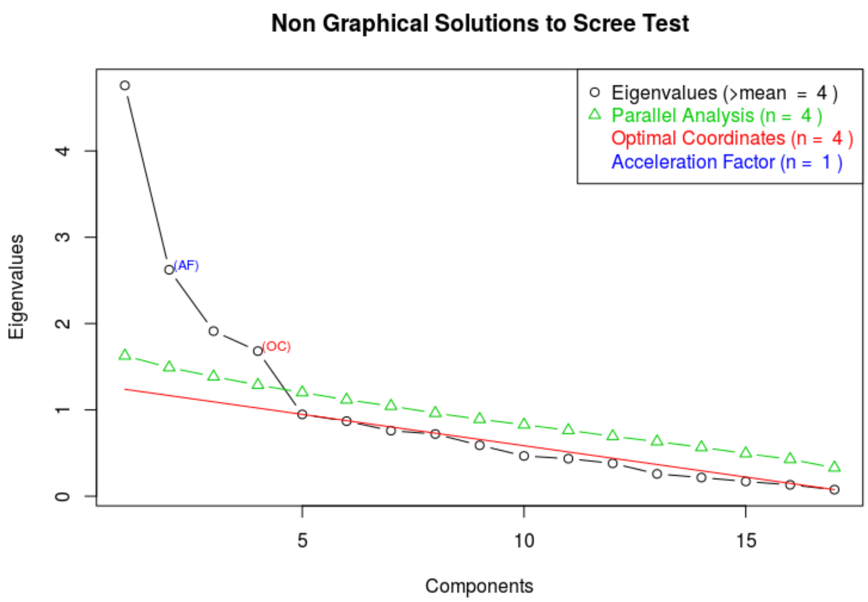
	β
(Intercept)	-3.28e-16
a	-0.092
v_{comp}	-0.090
d_{comp}	0.118
Nonverbal IQ	0.317
CTOPP RAN	0.512
CTOPP PA	0.169
ADHD	-0.002

Figure B.4: Number of predictors retained by lasso regression as a function of regularization.



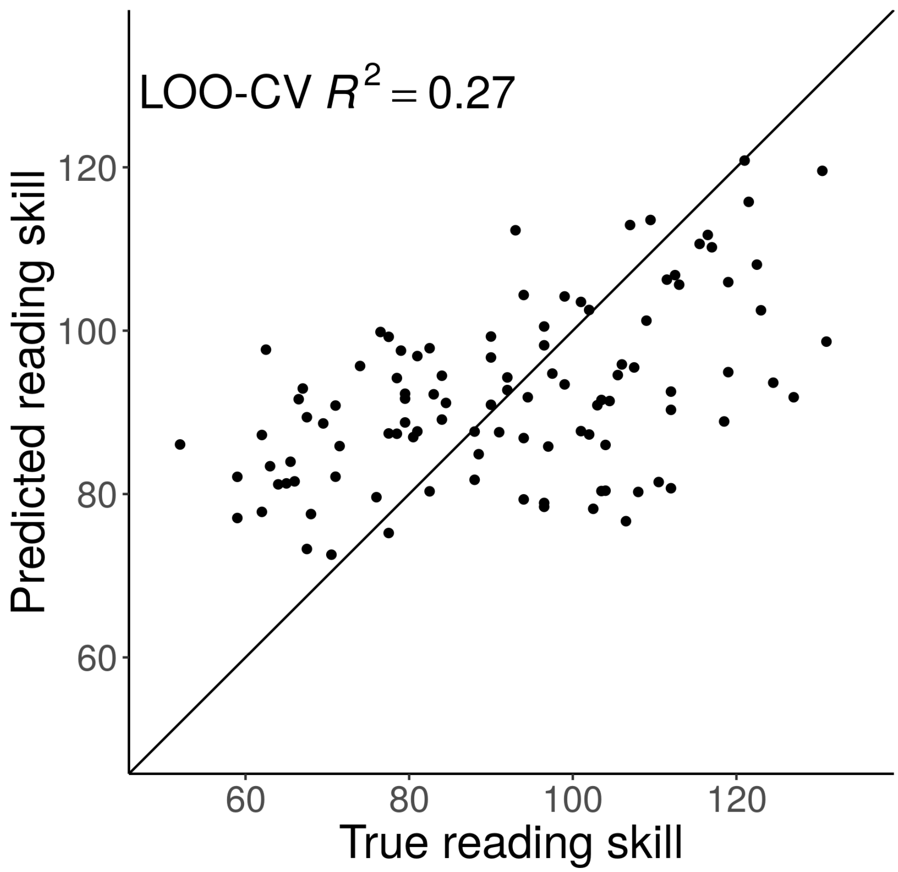
Number of predictors retained by lasso regression as a function of the regularization parameter λ with 10-fold cross validation.

Figure B.5: Scree test for exploratory factor analysis.



Scree test for exploratory factor analysis. Four standard measures of model fit are given: eigenvalues, parallel analysis, the optimal coordinates metric and acceleration factor metric.

Figure B.6: One-factor model predictions with cross-validation.



Comparison of true versus predicted reading skill for the single-factor model. Point estimates are computed with leave-one-out cross validation (LOO-CV).