Spring 2022

Using Item-Response Analysis to Evaluate the Task-Demands of a Pictural-Response Based Test of Word Reading

Michael Eason

USING ITEM-RESPONSE ANALYSIS TO EVALUATE THE TASK-DEMANDS OF A PICTURAL-RESPONSE BASED TEST OF WORD READING

by

Michael Eason

Bachelor of Science
University of South Carolina, 2017

Submitted in Partial Fulfillment of the Requirements

For the Degree of Master of Arts in

School Psychology

College of Arts and Sciences

University of South Carolina

2023

Accepted by:

Scott L. Decker, Director of Thesis

Samuel McQuillin, Reader

Tracey L. Weldon, Interim Vice Provost and Dean of the Graduate School

ACKNOWLEDGEMENTS

There are several people I would like to thank for their involvement and support in the completion of this thesis. I would first like to thank my mentor, Dr. Scott Decker, who has challenged and supported me since beginning in his lab. You have always valued my interests and thoughts, and you have always encouraged me in whatever direction I chose. I would like to thank my reader, Dr. Samuel McQuillin, whose supportive nature and encouragement was clear at every corner. I would like to thank my lab-mates for their emotional support and their constant willingness to listen to my ramblings, as well as the support of my mother, whose emotional support was gravely needed, and sisters, who have always shown interest in what I do. Finally, I would also like to thank William Hoskins, without your efforts and dedication to the technological development of our testing platform, this project would not be possible. Your psychological grit and our friendship are truly appreciated. To anyone else that I might have missed that has supported me in my efforts to complete this milestone, thank you too.

ABSTRACT

While some children's reading difficulties are a consequence of socio-economic factors and inadequate instruction, other children have extreme difficulties learning to read despite having adequate educational opportunity and adequate levels of intelligence. These children are classified under the Individuals with Disabilities Education Act (IDEA) as having a learning disability specific to reading (SLD-R). The identification of children with SLD-R is costly on time, personnel, and other school resources, and due to an insufficient number of resources generally available to schools, the methods implemented for identifying children are alarmingly inconsistent. The aim of the current study is to investigate the task demands of a recently developed computerized test of isolated word reading that was developed to address the resource barriers seen in schools. The task demands were explored by using both person and item characteristics to predict response-likelihood across items. An initial multiple logistic regression analyses revealed that a person's phonological awareness and oral vocabulary abilities directly predicted the likelihood of success across items. Item characteristic variables and interactions terms between item characteristics were added in subsequent models. Overall, the analyses suggest that the word reading test most demands phonological ability and phonetics knowledge, particularly feedbackward phonetics. Results also show that task demands are significantly moderated by the linguistic parameters required as part of the test design. Limitations of the study are discussed and suggestions for test optimization are given.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BOI......Body Object Interaction

CARE...... Carolina Automated Reading Evaluation

WC......Word Concreteness

Freq......Frequency

LTPF......Least Transparent Phoneme Forward Probability

LTPB......Least Transparent Phoneme Backward Probability

MSD...... Mean Serialized Distance

PSW......Patterns of Strengths and Weaknesses

SLD-R......Specific Learning Disability in Reading

CHAPTER 1

INTRODUCTION

Learning to read is an important part of education and its acquisition acts as a catalyst for success in modern society. Those with higher reading ability have a greater likelihood of accessing subsequent educational material (Ozuru, Dempsey, & McNamara, 2009), demonstrate a greater likelihood of succeeding in other academic areas (Bastug, 2014; Boonen et al., 2016; Ozuru, Dempsey, & McNamara, 2009; Meneghetti, Carretti, & De Beni, 2006), and generally have less difficulty in the educational system. Unfortunately, the converse is also true. Children with a lower reading ability have a more difficult time accessing subsequent educational material (Ozuru, Dempsey, & McNamara, 2009; Boonen et al., 2016), exhibit socio-emotional barriers specific to academic attainment (Eissa, 2010; Sideridis, 2007), are more likely to have stigmatizing and isolating interactions with the school system (i.e., exclusion and dropout; Barga, 1996; Elksnin & Elksnin 1996; Christle, Jolivette, & Nelson, 2005), and must frequently develop compensatory strategies to overcome reading-related difficulties in college (Adelman, 1990; Heiman & Precel, 2003; Shaywitz et al., 2020).

Reading achievement gaps can be seen in as early as six years old (Muter & Snowling, 2009; Ferrer et al., 2015) and when not treated or intervened on these gaps consistently persist into later years (Snowling, Muter, & Carroll, 2007; Muter & Snowling, 2009; Ferrer et al., 2015). Several studies have shown that early reading ability is an important predictor in later reading ability; naturally, those with more fluent reading have an easier time accessing reading materials and benefit more from exposure to print (Cunningham & Stanovich, 1997), while those with less fluent reading have a more difficult time and tend to avoid the matter altogether (Sideridis, 2003; Sideridis et al., 2006). Fortunately, those receiving the appropriate services during the earlier years of reading acquisition demonstrate significantly better outcomes than those not receiving specialized services (Suggate, 2016). Needless to say, the identification of those with reading difficulties is an imperative in the school system.

Reading acquisition and development is a multi-faceted issue, significantly interacting with both socio-emotional and environmental-contextual factors. However, even with adequate instruction and adequate general cognitive ability children can still have extreme difficulties learning. Historically, these children have been classified with a learning disability specific to reading (SLD-R) and typically make up the largest portion of students in special education services. Because of the historical lack of understanding surrounding the underlying causes of SLD-R, children are most often identified by determining whether their reading difficulties are lower than what is expected from their overall level of cognitive (i.e., intellectual) functioning; scores on reading assessments are determined as either statistically expected or unexpected given a composite on a range of measures of their cognitive functioning. The use of intellectual quotients to determine whether a student is fulfilling their reading potential, and classifying the student accordingly, is referred to as the IQ-Discrepancy method. The use of IQ-Discrepancy is appealing because it offers a means of making group distinctions that is more objective, being based on statistical convention rather than social value. The use of IQ-Discrepancy requires the assumption that achievement increases as a function of intelligence, except in SLD-R individuals, meaning that the cognitive processes underlying SLD-R must be non-influential in the calculation of IQ. While this proposition can hold true (Ferrer et al., 2010), the use of IQ-Discrepancy also presumes that individuals with SLD-R are qualitatively distinct from those with reading difficulties and converging IQ scores. However, little research, if any, has demonstrated that there are qualitative distinctions in reading acquisition between those with SLD-R with and without converging IQ scores, and has actually shown the contrary (Stanovich, 1994; Stuebing et al., 2002; Toth & Siegel, 1994). The most notable issue with the IQ-Discrepancy model is that it operates on the premise that students' performances are *unexpected*, and that their causes are *unknown*. While this was once the case, research efforts aimed at understanding the underlying causes of reading difficulties have produced a wealth of information that have helped define SLD-R not by its lack of expected performance but the nature of its identified causes.

Reading difficulties are most often conceptualized as the inability to determine the sounds (i.e., phonemes) that are made by the letters, or groups of letters, (i.e., graphemes) in a specific word and pronounce them in a fluid manner. In some languages, such as German, the sounds that each grapheme makes have a one-to-one ratio, meaning the phoneme-grapheme correspondence is consistent across word usage. In English, however, a considerable amount of graphemes (i.e., most vowels, some consonants, and select digraphs and trigraphs) have a one-to-many ratio, meaning that a single grapheme can have several different corresponding phonemes across words. The knowledge of which sounds can be made by each grapheme (i.e., phoneme-grapheme correspondence) is referred to as phonetic knowledge. Research has revealed that specific elements of oral language that are neurological in origin, such as phonological awareness (Badian, 1995; Catts et al., 2017; Furnes & Samuelsson, 2011; Hogan et al., 2018; MacDonald & Cornwall, 1995), auditory temporal detection (Au & Lovegrove, 2001; Mundy & Carroll, 2012), and lexical access (Georgiou et al., 2008; Manis et al., 1999, 2000; Mesman & Kibby, 2011; Savage & Frederickson, 2005) are the primary abilities responsible for the adequate acquisition of phonetic knowledge and basic reading ability. An individual's oral vocabulary may also have a role in the acquisition of word-specific phonetic knowledge (Hintze et al., 2003; Kim et al., 2014).

Consequently, SLD-R efforts have been made, and have been largely successful, at recharacterizing SLD-R as the presence of extreme difficulties in the decoding of words (reading aloud) and accurate spelling that are neurobiological in origin (Lyon et al., 2003). Under the revised definition provided by Lyon, Shaywitz, and Shaywitz (2003), practitioners assess children in several developmentally appropriate areas of reading and use statistical procedures to evaluate whether the reading scores obtained are expected given their phonological and auditory ability estimates and not better explained by differences on other domains, such as general language (Catts et al., 2005; McArthur et al., 2000), attention (Ehm et al., 2016; Rucklidge & Tannock, 2002), or emotional distress (Wanzek et al., 2014). If a child's reading performance is expected given their phonological and auditory ability estimates and their phonological and auditory ability estimates are not what is expected given the rest of their cognitive profile, the child is likely to meet criteria for SLD-R. This method of identification is termed Patterns of Strengths and Weaknesses (PSW). While PSW has strong theory and an adequate amount of data to warrant its use, it is often difficult to implement due to high demands on time and even greater demands on professional expertise compared to that of IQ-Discrepancy, placing an even greater emphasis on professional training within a field that is already unable to provide an adequate amount of personnel.

Attempts have been made to provide schools with assessments that can be administered in short periods of time and still provide information on a student's current and future achievement level, as well as response to intervention. However, these assessments are often limited in scope and are more concerned with outcome measurement than they are the differential identification of causes.

The ideal solution to identifying children with SLD-R is a system that can quickly assess for both reading difficulties and the corresponding cognitive deficits responsible, while still maintaining a level of diagnostic expertise expected from that of a highly trained professional. With the growing capabilities of technology in modern society and its increasing prevalence in schools, computerized assessments have become a viable option for the universal implementation of PSW models. Computerized assessments offer reliable presentation of stimuli and collection of responses across administrations, the opportunity for both national and local normative data, automation of data analysis, and assistance with interpretive reports and recommendations. Computerized assessments are also capable of individual test administration across a group of students at a single time point, thus even if the test itself is only equally as long as its paper-pencil counterparts more children can be tested across the same period of time. Researchers have demonstrated that computerized tests are comparable to their paper-pencil counterparts (Wang et al., 2008) and have demonstrated that computerized assessments can be effective assessments for reading comprehension (Davison et al., 2018; Förster &

Souvignier, 2011), phonological ability (Steacy & Compton, 2019), and knowledge of phoneme-grapheme correspondences (McBride et al., 2010). However, while the use of computerized assessments is becoming more widespread, there are no computerized assessment systems that offer a collection of reading tasks that have been normed together and can be used in combination with normed measures of reading-related abilities (e.g., phonology, vocabulary, listening comprehension, etc.). To date, no computerized batteries have been developed and validated that can differentiate between the underlying causes of reading difficulties to the extent that commercial batteries have offered, particularly for use across the most vital age span for identifying children with reading difficulties (kindergarten through third grade). This may partly be because the tests that are most commonly used to assess for reading difficulties are pronunciation based (i.e., word identification, nonsense word reading, oral reading fluency) and the use of computerized tests in measuring reading ability poses its own challenge for tests typically requiring pronunciation and articulation. Unless test developers are able to program speech recognition into the computerized assessment and account for varying dialects and regional speech patterns, pronunciation-based tests cannot be administered through an automated computerized format. Their designs will have to undergo a shift in response-modality, such as that seen in FastBridge (Christ, 2017), where researchers opt to measure phonetic knowledge and word-specific spelling via the orthographic choice task (i.e., children are provided the pronunciation of the word and are required to choose the correct spelling). Even then, FastBridge supplements its testing with paper-pencil administrations, limiting the effectiveness of its usage.

A promising solution is one developed by Hoskins, Hobbs, Eason, and Decker (2021). Hoskins and colleagues outline the technological development of a computerized platform for assessing a range of reading and reading-related abilities, the Carolina Automated Reading Evaluation (CARE), and provide preliminary analyses that suggest comparable utility of their system with universal screening systems standardly used in schools. The CARE is a collection of automated computerized assessments involving letter-naming, phonetics knowledge, word identification, sentence reading, cloze reading (i.e., fill-in-the-blank), reading fluency, and several phonological, oral language, and attentional tests. The extensive inclusion of reading and reading-related measures is both exciting and promising given the lack of comprehensive fully computerized assessments currently available for use in schools. Given the promise of computerized assessments, the theoretical alignment of the CARE with the PSW model of identification, and the potential widespread dissemination of the CARE, a further investigation into its specific task elements and demands is warranted.

In navigating the problem of response-modality for their word identification test, the developers opted to use pictures as the response-modality, which naturally raises concerns over confounding demands of expressive vocabulary and the potential for students with specific language impairment to be identified as a “poor reader” on the basis of inadequate picture-knowledge alone. Conversely, students with an adequate expressive vocabulary and picture-knowledge base may be assisted in the identification of words by the multiple-choice picture-format. It is well known that words with a higher imageability are more easily learned and acquired as it relates to text-reading (Steacy & Compton, 2019) and that the acquisition of the verbal representation of a word (i.e., oral vocabulary) is directly related to the likelihood of learning the specific orthography of the corresponding written word (Kim et al., 2013, 2014). Research has suggested that the relationship of vocabulary with word identification could be due to the activation of phonemic components in learning the semantic components of vocabulary (Ehri, 2014). Perhaps a stronger semantic component may be responsible for a greater phonemic activation during orthography learning and during word identification performance. For this reason, it is suspected that the use of pictures as a response-modality in the CARE would aid in the word reading process and that those who would normally not be able to read a word from orthography alone may be seen as performing more adequately in word reading than would be reflected in pronunciation-based text-reading formats. In other words, using pictures as a response-modality may allow for the opportunity of different test-takings strategies.

In typical pronunciation-based reading tasks, the only performance strategy typically available to readers is “spelling-to-sound” reading, commonly known as “decoding.” Decoding requires the use of feedforward phonetics, which is the determination of the correct phoneme from the given grapheme. The converse of feedforward phonetics is feedbackward phonetics, otherwise known as “sound-to-spelling,” or more simply “spelling,” and is the first of two additional strategies that may be made available to the students in the word identification test design in the CARE. Feedbackward phonetics is characterized by the process of determining the correct grapheme correspondences given the phoneme of the word. Given that students in the CARE word reading test have four available pictures, it is assumed that they have four corresponding words from which they can employ feedbackward phonetic strategies and determine which corresponds with the presented word. The third reading strategy is partial word reading. Partial word reading strategy also utilizes feedbackward phonetics and uses the phonemic information made available by the pictures, except that those using this strategy do not necessarily determine the phoneme-grapheme correspondences for the full word. Instead, students select the answer from partial knowledge and make their final decision using process of elimination from the information that is known and the information obtained from the pictures available. This strategy is assumed to be largely available to the student when the pictures made available as options are less similar in phonology and orthography than the word presented. Additionally, since the sound-to-spelling and partial word reading strategies require the use of phonemic information obtained from the pictures, the use of these two are dependent on the extent to which this information can be accessed and used for performance.

The purpose of the current study is to identify whether the task-demands of the Word Identification test in the CARE assessment battery are consistent with what would be expected of a pronunciation-based word reading test; specifically, the study will identify whether participants' phonological awareness abilities predict item-response and whether the items' feedforward orthographic consistencies predict item-response beyond that of the alternatives. Clarifying the task-demands will yield greater interpretability of the scores on the Word Identification test and examine its appropriateness for use in a school setting.

CHAPTER 2

METHOD

Participants

The current study examined the effects of person- and item-level characteristics on response-success probability on a pictural-response based word reading test in a sample of children ages five to nine years old. Participants in this study included children in general education and tiered support classrooms (i.e., not self-contained) across three different elementary schools in the state of South Carolina. Children were recruited for participation through existing collaborations with school psychologists, reading interventions, and RtI team members at the respective schools.

The personnel responsible for students' data management at each school provided the researchers with the appropriate demographic information, including date of birth, race, sex, and grade. For children to be considered eligible for the study, they must have been in grades Kindergarten through Third and must have not been identified by their school with a significant cognitive impairment (i.e., IDEA classification under Intellectual Disability) or a significant developmental delay (i.e., IDEA classification under Developmental Delay). Once children were identified as qualifying for the study, they were given unique identifiers for the purposes of the study and all identifying information (i.e., names, school identification numbers) was removed. The mean age for participants for the entire sample (*N* = 570) was 6.77 (*SD* = .90) years, the number of females in the study was 289 (51%), and the primary ethnic group for the sample was

White (68.6%). The total number of children in kindergarten, first grade, second grade, and third grade was 157, 331, 50, and 32, respectively.

Data Collection Procedures

All children participating in this study were evaluated as part of another study aimed at the development and validation of a computer-based psychoeducational test battery. The psychoeducational battery used was non-commercial, research-based, and consisted of measures of basic and advanced reading skills, phonological awareness, oral language ability, and other select cognitive abilities.

Children were tested individually within a group setting (i.e., each child was at their own computer in the classroom and entire classrooms of children were tested a time). For data collection at each school, touchscreen computers and headphones were placed within a single designated classroom and placed as far apart as possible to prevent children from interacting with each other while testing. Children were directed to listen to the audio on the computer and to simply following instructions given to them. Once children were sitting and the volume was set at a comfortable level, the research personnel launched the test administration and monitored the children appropriately. Children were directed back to their classrooms after they finished participating.

The tests administered to children at each elementary school varied according to the school's needs and the time limitations imposed but always consisted of the tests used in the current study. All procedures conducted within this study were approved by the University of South Carolina Institutional Review Board (IRB) and approved by the appropriate governing research committees for each elementary school.

Materials

CARE Word Identification (WI) Test. The CARE WI test is administered by presenting a word on the screen (e.g., *hat*) for each item and simultaneously presenting four pictures to choose from (*e.g., bat, hat, cat, ham*), one of which corresponds to the presented word. Ideally, the remaining pictures, *distractors*, would be chosen on the basis of similarity on phoneme-grapheme correspondences and orthographic similarity. However, in the construction of item distractors the developers found that the options were limited in which distractors could be chosen based on the chosen similarity metrics while also being reliably recognizable to children (i.e., the language demands for picture recognition were developmentally appropriate). Therefore, the distractors were chosen using a multi-level multi-dimensional scheme intended to maintain a balance between the types of features each item relates to and reduce the likelihood that the participants can arrive to the correct response by process of elimination, at least with non-phonetic strategies. The distractors were therefore chosen based on their variability and similarity on two dimensions. Phonetic Similarity and Physical Appearance. When possible, each distractor was chosen such that it maximized similarity to the distractor based on both physical appearance and phonetic sequence. However, distractors were chosen such that there was at least one picture with similar physical appearance as the correct picture and at least one picture with phonetic similarity as the correct picture. The fourth distractor was defaulted to another phonetically similar picture or, if necessary, a picture whose corresponding word was unidentifiable or ambiguous.

A total of 72 items, with their corresponding distractors, were created. Two items were removed after data collection due to their misalignment with item construction definitions (they had morphemes; climbing, laughing). The words used in the test were chosen to maximize variability in difficulty, ideally up to what would be expected for those in third grade. Words were chosen from the English Lexicon Project (Balota et al., 2007). The English Lexicon Project is considered a mega-study and provides a database of 40,481 words in the English language and their respective lexical and sublexical features. In order to maximize this variability and range, while maintaining phonetic skill as the primary demand, words were chosen based off broader lexical and sublexical characteristics (i.e., frequency, length, number of syllables and number of letters in each constituent grapheme, as well as their suitability for administration in the given test format, they must be illustratable and have lower perceived demands on picture vocabulary). To facilitate the selection of words in ascending anticipated difficulty, words in the database were first ordered according to the most encompassing characteristics, which were frequency and length. Next, words were characterized according to their number of syllables and number of digraphs (graphemes with 2 letters), trigraphs (graphemes with 3 letters), and quadrigraphs (graphemes with 4 letters). Two to three words were chosen for each combination of features found in the list so long as the words were perceived as easily illustratable and having minimum picture vocabulary demands.

An example item of the test administration can be found in Figure 2.1 below. Participants were instructed to select the picture that matched the presented word. If the participants at any point in time demonstrated prolonged duration on any single item-attempt or expressed confusion over a certain item, they were encouraged to try their best and if needed to select the answer to their best guess. Responses were collected as a dichotomous variable, either being incorrect (value of 0) or correct (value of 1).

Participant Phonological Awareness. Phonological Awareness (PA) was measured using the Sound Blending test on the CARE. Participants are presented with a screen of four pictures, one of which corresponds to the correct answer. Audio recordings of phonemes are presented one-at-a-time (i.e., */k/-/a/-/t/*) with 1 second intervals of silence between each. When presented non-segmentally (i.e., combined; blended) these phonemes form a real word (i.e., cat). Participants must determine what word the combined phonemes represent and select the picture that most appropriately corresponds to the word. Sound Blending specifically measures the ability for participants to detect phonemes in isolation, maintain these phonemes in their phonological loop and determine the word represented by their combined presentation. Sound blending is a measure of phonological processing. The traditional implementation of Sound Blending does not require the use of pictures, instead requiring pronunciation of the word; this format is used in several standardized test batteries (e.g., WJ-Cog, CTOPP, etc.; Schrank, Mather, & McGrew, 2014; Wagner et al., 1999) and has an established role in the research base as a predictor of word decoding ability and phonics. To avoid confounding phonological awareness with general age-related development, Sound Blending scores were regressed onto the child's age. The metric used in this study is the difference between the predicted logit value and the actual logit value for each observation, the residual.

Participant Receptive Vocabulary. Receptive Vocabulary (RV) was measured using the Oral Vocabulary test on the CARE. Participants taking the Oral Vocabulary test are presented with a screen of four pictures, one of which corresponds to the correct answer. An audio recording of the correct picture's name is presented. A button on the bottom left of the screen with an audio-speaker symbol is made available so that the audio presentation can be repeated as many times as needed by the child. Participants must choose the picture on the screen that matches that of the audio recording. Oral vocabulary specifically measures whether children have acquired the semantic meaning for a specific noun, verb, or abstraction, and can then recognize the word in the form of a picture. RV has been shown to be an important indicator of verbal intelligence and predictor of reading comprehension and language acquisition. This test was chosen as a covariate in predicting word reading response-success due to the anticipated confounds of vocabulary in the WI test. To avoid confounding receptive vocabulary with general age-related development, Oral Vocabulary scores were regressed onto the child's age. The metric used in this study is the difference between the predicted logit value and the actual logit value for each observation, the residual.

Item Least Transparent Phonemic Probabilities. Each phoneme-grapheme correspondence in a word has an estimated frequency of occurrence in the English language. The Least Transparent Phoneme (LTP) is a metric of frequency for the least frequently occurring phoneme-grapheme correspondence in the word for each item. Hanna (1996) established an estimate for these frequencies based off a list of 44 phonemes and 120 graphemes, and Fry (2004) succinctly summarized these findings in a paper to be more accessible to researchers. The frequencies for each of the 192 phoneme-grapheme correspondences is summarized in Fry (2004). LTP has been shown in the research base to be a strong predictor of word reading difficulty, most likely because of its robustness to the rest of the word composition and therefore being more reliable and “pure” indicator of orthographic consistency. A lower value of LTP is expected to translate to a greater difficulty in phonetic translation in word reading. Two specific types of metrics were derived from the LTP and used in the analysis for the current study: LTP Feedforward Probability (LTPF) and LTP Feedbackward Probability (LTPB). The probability of each phoneme-grapheme correspondence is calculated by dividing the estimated frequency of the phoneme-grapheme correspondence by the summation of all their possible counterparts. When the reader is given the grapheme only, their counterparts can consist of all possible corresponding phonemes and the resulting metric is LTPF. When the reader is given the phoneme only, their counterparts can consist of all possible corresponding graphemes and the resulting metric is LTPB. Each phoneme-grapheme unit in the word has its own feedforward probability and feedbackward probability. This study conceptualizes LTPF as representing the spelling-to-sound difficulty for the word and LTPB as the sound-to-spelling difficulty for the word. A lower probability in either of the two metrics is expected to translate to a greater difficulty in word reading in their respective strategies. A visual explanation of LTPF and LTPB as it relates to the other frequency metrics can be found in Figure 2.2, where the LTPF value for “Turtle” is .98 and the LTPB value for “Turtle” is .01.

Item Mean Serialized Distance. Mean Serialized Distance (MSD) is a metric formulated by the current researchers. Mean Serialized Distance represents the average distance between each combination of two phoneme-grapheme correspondences when accounting for frequency, grapheme-length, and positional distance between the graphemes within the word. The absolute difference for the frequencies were multiplied by the absolute difference of the positions and multiplied by the average letter length of the two graphemes (to account for difficulty of the length of the grapheme). MSD can be conceptualized as the likelihood of cumulative amount of phonetic information offered to the reader in the word. For example, the word "Turtle" has four phoneme-grapheme units (i.e., "t" for /t/ twice, "ur" for /er/, and "le" for /l/; Figure 2.2). The first value (a) is the absolute difference between the base rates (i.e., frequency) of "t" and "le" is 6,908 and represents some relative difference in how frequently children are exposed to the associations. The second value (b) is the absolute difference between the letter position of "t" and "le" is 1, indicating the smallest distance possible between the two grapheme units. The third value (c) is the average grapheme length is 1.5, indicating that the number of letters used to obtain the appropriate information. The three values are multiplied (a*b*c) to indicate the cumulative, or chunked, information that is available for a single combination of grapheme units. These calculations are performed for every combination of graphemes in the word and averaged for the word. A larger MSD is thought to represent a lesser likelihood of information available for partial word reading. A visual explanation of MSD as it relates to the other frequency metrics can be found in Figure 2.2.

Item Whole Word Frequency. The logarithmic frequency (Freq) of each whole word was used as a metric of exposure that children are anticipated to have to a word. The logarithmic frequency for each word was obtained from a HAL word corpus (i.e., UseNet) snapshot, taken by Lund & Burgess (1996) and made available by the English Lexicon Project. The logarithmic frequency was calculated by taking the natural logarithm of the number of occurrences of the word in the corpus snapshot, excluding homophones. Frequency effects have shown large effect sizes in relatedness to word reading difficulty because of the effects of word exposure on word-specific phonetic knowledge. A visual explanation of logarithmic frequency as it relates to the other frequency metrics can be found in Figure 2.2, where the value of Freq for "Turtle" is indicated as 9.109.

Item Word Concreteness. Word Concreteness (WC) is a linguistic feature of the target word that represents the level of word abstraction on a scale of one to five, with one representing language-based definitions and five representing experience-based definitions (Brysbaert, Warriner, & Kuperman, 2014). WC has shown to have a direct effect on oral vocabulary acquisition (Groot & Keijzer, 2000), and subsequently vocabulary on word reading (Ouellette, 2006; Kim et al., 2014). Word imaginability has shown to have a direct effect on word reading accuracy and acquisition (Steacy & Compton, 2019). WC ratings for the current study were obtained from those provided by Brysbaert, Warriner, & Kuperman (2014) and made available by the English Lexicon Project. Higher WC is expected to significantly predict response-success likelihood through more frequent semantic exposure, improved lexical access, and greater likelihood of sound-to-spelling reading strategy and lower likelihood of spelling-to-sound reading strategy.

Body Object Interaction (BOI) is a linguistic feature of the target word and was obtained from (Pexman et al., 2019) and made available by the English Lexicon Project. Body Object Interaction is rated on a scale from one to seven, seven being the highest degree of body-object-interaction. Higher BOI ratings have shown to correspond with a higher likelihood of accuracy and reaction time in semantic decision tasks (Pexman et al., 2019). It is conceptualized in the current study as measure of the estimated strength of semantic representation for each word. It is expected to significantly predict response-likelihood through easier semantic representation and improved lexical access. For this reason, it is also expected that BOI ratings will moderate the extent to which feedbackward and feedforward reading strategy is used.

Planned Analysis

To investigate the extent to which the CARE WI test demands cognitive constructs expected from pronunciation-based word reading tasks and whether the shift to pictures for response-modality introduce other pathways (i.e., strategies) of reading performance, an approach will be used that is similar to the explanatory item response model used by Compton and colleagues (2020), with the exception that person and item characteristics will not be crossed into interaction terms and random effects will not be included in the model. Person ability estimates for PA and RV will be used to predict the observed response for each person across all items, therefore accounting for both person and item variability for each response. The resulting models are a series of binary multiple logistic regression models and are used hierarchically to determine the influence of the variables and interaction terms as it relates to the predictive power of each other. All variables were scaled to a mean of zero and a standard deviation of 1 prior to analyses. The descriptions of models are presented below and in Table 2.1.

Model 1: The current study makes the assumption that if the accuracy of a person's response to an item is uniquely predicted, at least in part, by phonological ability, then the requirement for phonetic knowledge acquisition will be higher than if phonological ability did not contribute a statistically significant amount of unique variance. Additionally, if a person's phonological ability estimates uniquely predict their response to word reading items even in the presence of their vocabulary estimate, then it will be assumed that the test requires phonics uniquely beyond that of vocabulary ability alone. To determine the extent to which each item's response is primarily rooted in phonetic knowledge, a binary multiple logistic regression will be used in which the accuracy of the participants' responses was predicted by both phonological awareness and receptive vocabulary. Oral language was included as a covariate because the test administration format for the isolate word reading test allows the confounding of oral vocabulary and general object knowledge with word reading performance. Including both phonological awareness and oral language into the regression allows for the identification of items whose requirements are unique to phonological demands and not just picture vocabulary or general language ability.

Model 2: Each of the sublexical frequency characteristics of the items (LTPF, LTPB, MSD) in the CARE WI test were added into the model to test the extent to which each reading strategy is used across persons and items. The word's frequency (Freq) was also added into the model to ensure that the sublexical frequency variables used in the model predicted variability uniquely beyond that of whole-word frequency.

Model 3: The third model introduces the word-level features that relate to the use of pictures as a response-modality (WC, BOI). The features are added as both main effects and interaction terms onto the LTPF and LTPB models to determine the extent to which WC and BOI act as "gatekeepers" for each strategy. LTPF is conceptualized as the decoding of a word without the phonemic representation and therefore requires an adequate degree of familiarity. It is expected that the statistical contribution of LTPF to response-likelihood is maximized by increased levels of BOI. LTPB is conceptualized as the spelling of the word given the phonemic representation and we expect that abstraction of the word serves to limit or enhance the accessibility to the phonemic representation; the statistical contribution of LTPB to the response-likelihood will be maximized by increased levels of WC. While we do not anticipate that WC has a direct effect on LTPF in terms of causality, we expect that it will has a statistically moderating effect in the model, since the reading strategies represented by LTPF and LTPB are mutually exclusively within a single item-response.

Model 4: The last model introduces the interaction terms of WC and BOI on the MSD and Freq variables as additional "gatekeepers" of information specific to the statistical analysis. The purpose is to examine whether the predictive power of the sound-to-spelling reading strategy are dependent on the extent to how much information is made available and the familiarity of the word. It is expected that the addition of these interaction terms will increase the statistical effects of the sound-to-spelling terms introduced in previous models, but no specific hypotheses are made on the interaction terms themselves.

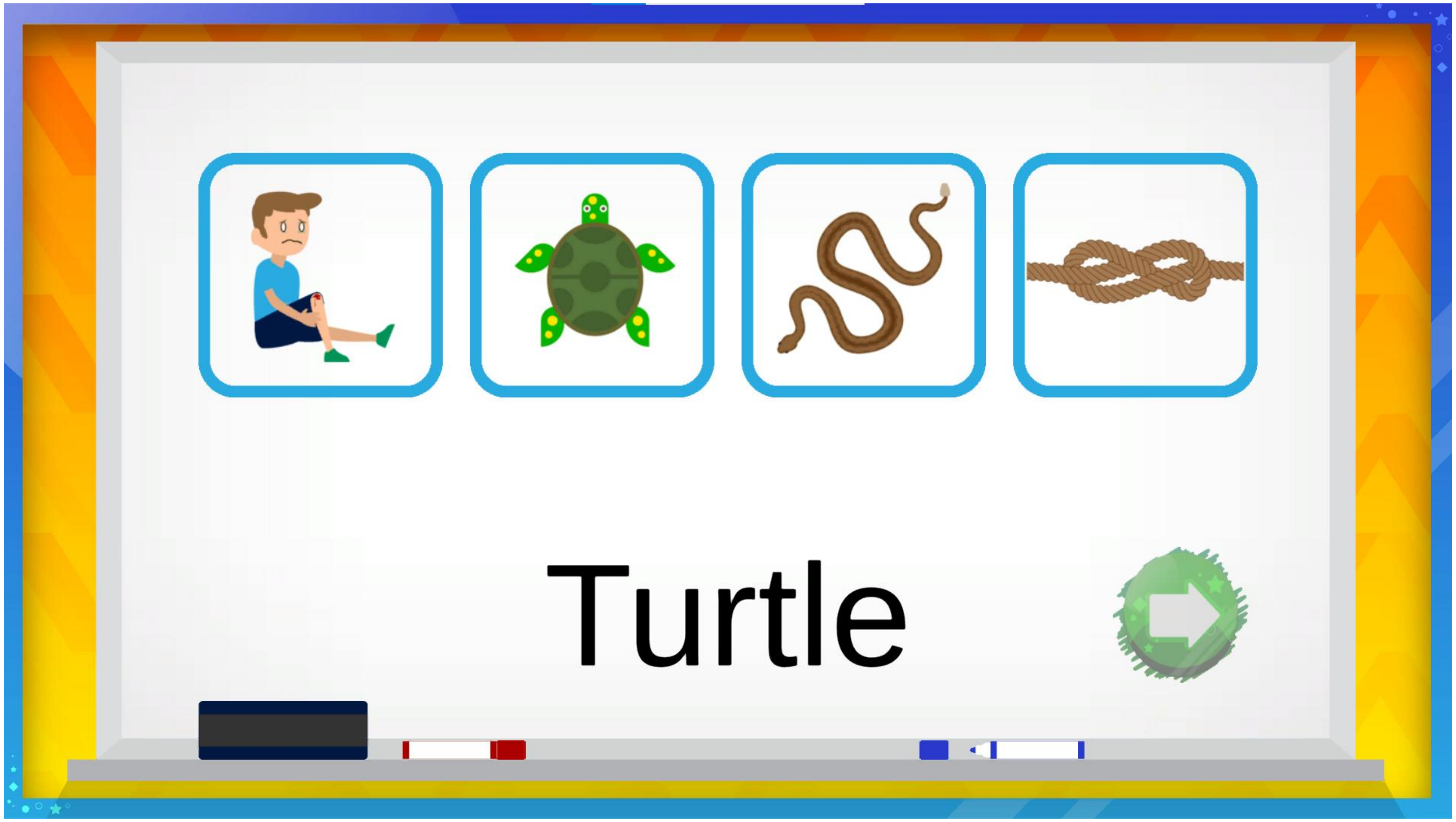

Figure 2.1. An example of an item from the CARE Word Identification test

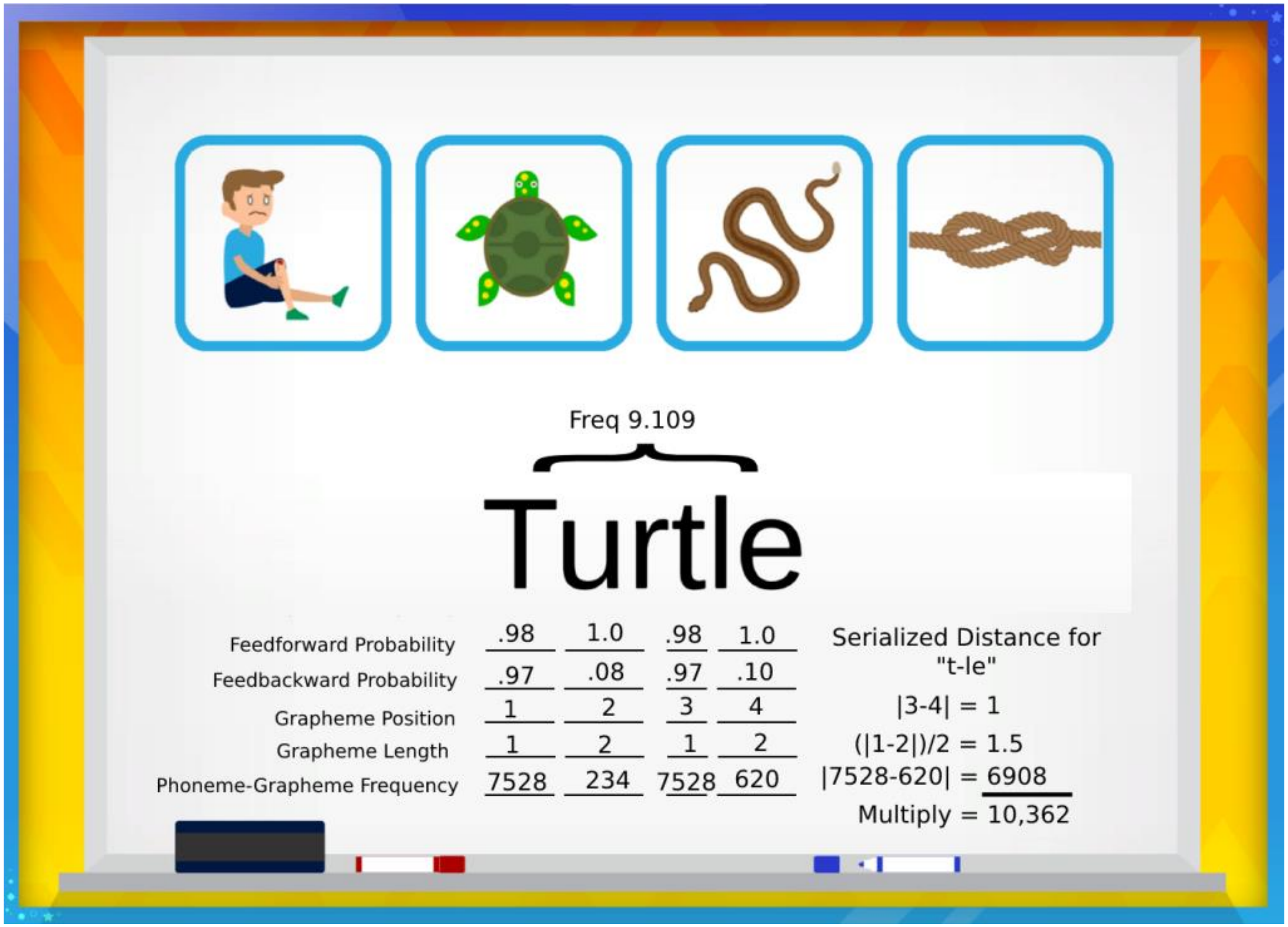

Figure 2.2. A Visual Explanation of the Frequency Metrics

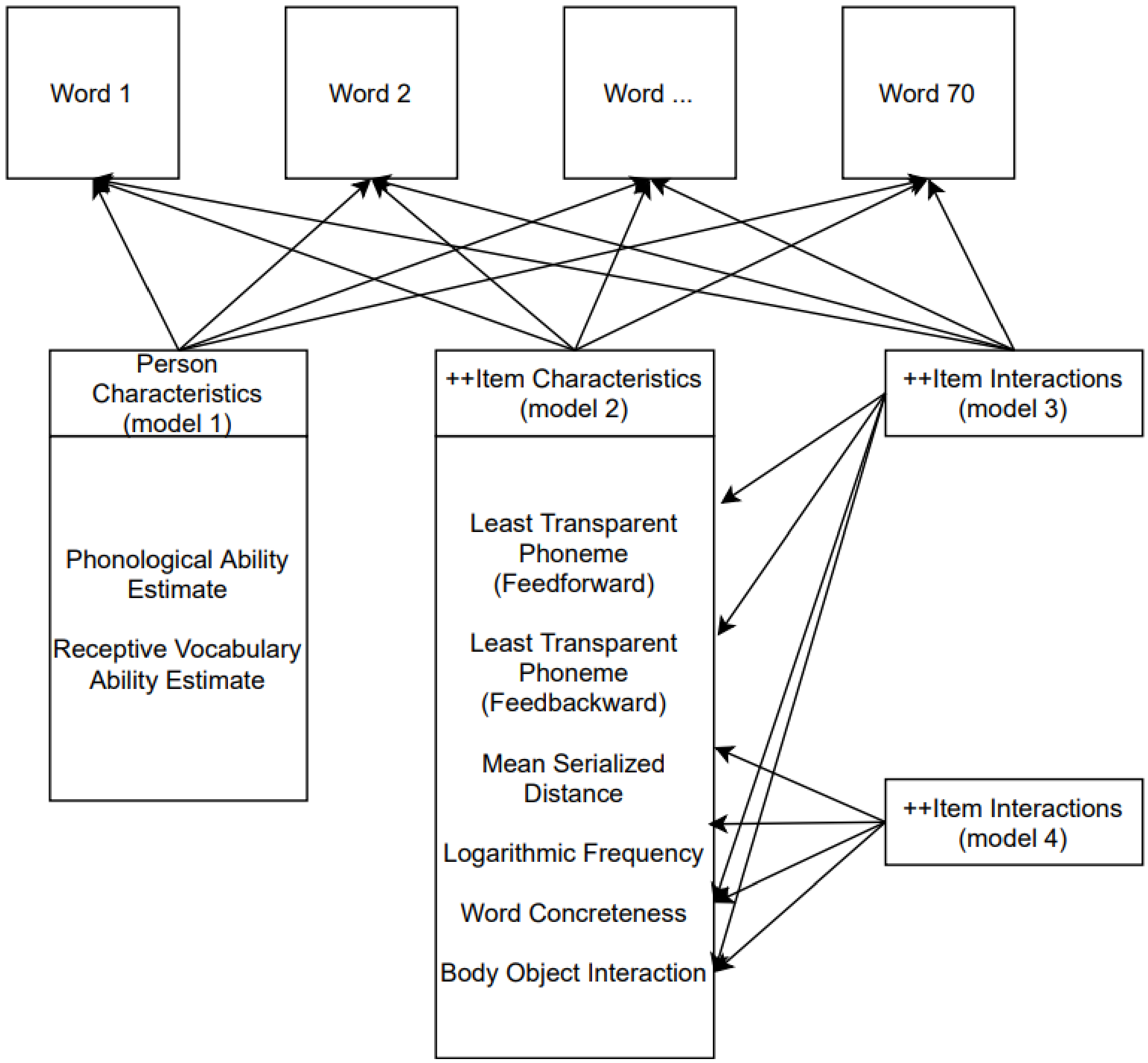

Figure 2.3 Person, Item, and Item Interactions for Predicting Item Response

Table 2.1 Overview of Models in Hierarchical Multiple Regression

Model	**Description**	**Purpose and Expected Results**
1	Person Characteristics	Identify the base extent to which the task requires phonology (translating to phonics) and vocabulary (translating to picture recognition).
2	Addition of Main Effects for Frequency Metrics	Identify the base extent to which different strategies are involved in the task without considering the parameters of the response-modality.
3	Addition of Linguistic Features as main effects and interaction terms with LTPF and LTPB	Identify the extent to which the parameters facilitate or hinder performance on the task and the extent to which they moderate the involvement of different reading strategies.
4	Addition of Interaction terms of Linguistic features with MSD and Freq	Identify the extent to which the inclusion of word-specific knowledge and partial-word information predicts task performance and changes the predictive value of other variables.

CHAPTER 3

RESULTS

The descriptive statistics for all predictor variables can be found in Table 3.1. Pearson's correlations were used to examine the extent of collinearity among predictor variables. Crossing the person and item features is required to calculate correspondence between the person features and item features. This procedure necessarily results in almost a zero-correlation, so correlations were only obtained for within item features and within person features. A Pearson's pairwise correlation was conducted between the PA and RV variables. The analysis revealed a moderate correlation between the two ($r = .57$, $p < .001$). Pearson's pairwise correlations were also conducted amongst the item-level variables. Results indicate statistically significant correlations between LTPF and LTPB ($r = .25$, $p < .05$), LTPB and MSD ($r = -.27$, $p < .05$), MSD and BOI ($r = .38$, $p < .01$), and WC and BOI ($r = .68$, $p < .01$). The correlation matrix between the item-level predictors can be found in Table 3.3. The results of each of the models are presented in Table 3.3.

Model 1: To determine whether a person's phonological ability and receptive vocabulary ability significantly predicted the likelihood of response-success across items on the test of word reading, a binary multiple logistic regression analysis was conducted. The results of the analysis revealed an overall significant model, indicated by the Likelihood Ratio test (Log-likelihood Difference = -1799.3, $p < .001$). A Pseudo R-squared was calculated using the McFadden Pseudo R-squared (.06). The results indicated that a person's phonological and oral language ability estimates significantly predicted a person's likelihood of success. Specifically, the results revealed that an increase in 1 standard deviation in phonology predicted an increase in the odds of response-success by 1.64 ($p < .001$) while an increase in 1 standard deviation in oral language ability predicted an increase in the odds of response-success by 1.29 ($p < .001$).

Model 2: To determine whether an item's sublexical frequencies and word-specific frequencies predict the likelihood of response-success across items, additional terms were added to the binary multiple logistic regression analysis in Model 1. The analysis was conducted in which both the participant's ability estimates on PA and RV were included with the sublexical frequency variables. The results of the analysis revealed an overall significant model, indicated by the Likelihood Ratio test (Log-likelihood Difference = -3304.1, $p < .001$). A Pseudo R-squared was calculated using the McFadden Pseudo R-squared (.12). The results of the analysis revealed that three of the predictors demonstrating statistical significance. PA and RV maintained their effects while LTPF, LTPB and Freq showed statistically significant effect sizes. Specifically, an increase in 1 standard deviation of the of LTPF increased the odds that a participant demonstrated a successful response on an item by 1.68 ($p < .001$) and 1.47 (LTPB; $p < .001$) and 1.08 (Freq; $p < .001$). MSD did not demonstrate any significant statistical contributions to response-likelihood.

Model 3: To determine whether an item's WC and BOI weakened or strengthened the relationship between feedforward phonetic strategy (LTPF) and feedbackward phonetic strategy (LTPB), a binary multiple logistic regression analysis was conducted in which WC and BOI were added as interaction terms onto LTPF and LTPB as "gatekeeping" terms. The results of the analysis revealed an overall significant model, indicated by the Likelihood Ratio test (Log-likelihood Difference = -5960, $p < .001$). A Pseudo R-squared was calculated using the McFadden Pseudo R-squared (.21). The results of the analysis revealed no statistically significant changes in the contribution of individual predictor estimates when compared to the previous model. Conversely, all newly added variables from the current model, except for main effect of BOI, added statistically significant contributions. Specifically, an increase in 1 standard deviation of CR decreased the odds that a participant demonstrated a successful response on an item by .91 ($p < .001$) and .58 (LTPB:WC; $p < .001$), an increase in odds by 1.30 (LTPF:BOI; $p < .001$) and 2.89 (LTPB; $p < .001$), and a decrease in odds by .66 (LTPB:BOI; $p < .001$).

Model 4: To examine the conditions under which the accessibility of the information required for feedbackward reading strategy predicts response-likelihood is maximized, a binary multiple logistic regression was conducted in which a second wave of gatekeeping terms were introduced, specifically, WC and BOI on MSD and Freq. The results of the analysis revealed an overall significant model, indicated by the Likelihood Ratio test (Log-likelihood Difference = -6318.7, $p < .001$). A Pseudo R-squared was calculated using the McFadden Pseudo R-squared (.22). with a statistically significant increase in odds ratio for LTPF and a statistically significant decrease in odds ratio for LTPB when compared to Model 2. The results also revealed a statistically significant decrease in odds ratio for CR, a statistically significant increase in odds ratio for BOI and LTPB:WC, and a statistically significant decrease in odds ratio for LTPB:BOI. The interaction terms introduced in this model were also found to be statistically significant such that an increase in 1 standard deviation of WC lead to an increase in the predictive strength of MSD by 1.28 ($p < .001$) and an increase in the predictive strength of Freq by 1.44 ($p < .001$). Conversely, BOI was found to lead to an increase in the predictive strength of MSD and Freq but in the other direction, by .67 ($p < .001$) and .78 ($p < .001$), respectively. The graph for model 5 can be found in Figure 3.1 below. Each of the observed item-responses is found on the y-axis, values limited to 0 (incorrect) or 1 (correct), and the estimated likelihood of each response being correct along the x-axis. The logistic curve obtained through Model 4 is provided in figure 3.1.

Table 3.1 Descriptive Statistics for Variables used for Analysis Variables

Variable	M	SD	Max	Min
Person (I = 570)				
PA	3.35e-17	1.57	4.75	-6.175
RV	-5.20e-17	1.27	4.74	-8.025
Item (J = 70)				
LTPF	.38	.32	1.00	.01
LTPB	.23	.28	.97	.02e-1
MSD	6,940	418	18,755	541.5
Freq	9.88	1.25	12.32	6.56
WC	4.70	.444	5	2.53
BOI	4.91	1.40	6.88	1.65
Item-Person (R = 35,319)				
WI	.55	.50	0.00	1.00

Table 3.2 Pearson's Pairwise Intercorrelations for the Item Predictor Variables

	1	2	3	4	5	6
1. LTPF		.25*	-.04	-.07	.05	.19
2. LTPB			-.27*	-.08	.03	.12
3. MSD				-.18	-.23	-.38**
4. Freq					.14	.20
5. WC						.62**
6. BOI						

**. Correlation is significant at the .01 level
*. Correlation is significant at the .05 level

Table 3.3 Results from Regression Models

	Model 1		Model 2		Model 3		Model 4	
Term	OR	CI	OR	CI	OR	CI	OR	CI
Intercept	1.19	(1.17, 1.22)	.93	(.90, .96)	.95	(.91, .98)	.83	(.80, .87)
Person								
PA	1.64	(1.6, 1.68)	1.66	(1.62, 1.7)	1.65	(1.61, 1.70)	1.67	(1.63, 1.72)
RV	1.29	(1.26, 1.32)	1.32	(1.29, 1.36)	1.32	(1.29, 1.36)	1.33	(1.30, 1.37)
Item								
LTPF			1.68	(1.57, 1.8)	1.83	(1.69, 1.99)	2.06	(1.89, 2.24)
LTPB			1.47	(1.35, 1.6)	1.28	(1.17, 1.41)	1.26	(1.14, 1.39)
MSD			.99[ns]	(0.97, 1.01)	1.00[ns]	(.98, .1.03)	.95[ns]	(.93, .98)
Freq			1.08	(1.06, 1.1)	1.10	(1.07, 1.12)	1.21	(1.18, 1.24)
CR					.91	(.88, .95)	.71	(.68, .75)
BOI					.99[ns]	(.94, 1.04)	1.23	(1.16, 1.30)
Item-Item								
LTPF:WC					.58	(.50, .67)	.65	(.56, .76)
LTPF:BOI					1.30	(1.17, 1.44)	1.47	(1.32, 1.63)
LTPB:WC					2.89	(2.48, 3.37)	5.34	(4.42, 6.46)
LTPB:BOI					.66	(.58, .74)	.35	(.31, .41)
MSD:WC							1.28	(1.20, 1.36)
MSD:BOI							.67	(.64, .70)
Freq:WC							1.44	(1.38, 1.50)
Freq:BOI							.78	(.74, .81)

[ns] Indicates odds ratios that have a non-significant p-value (>.001), all others are significant

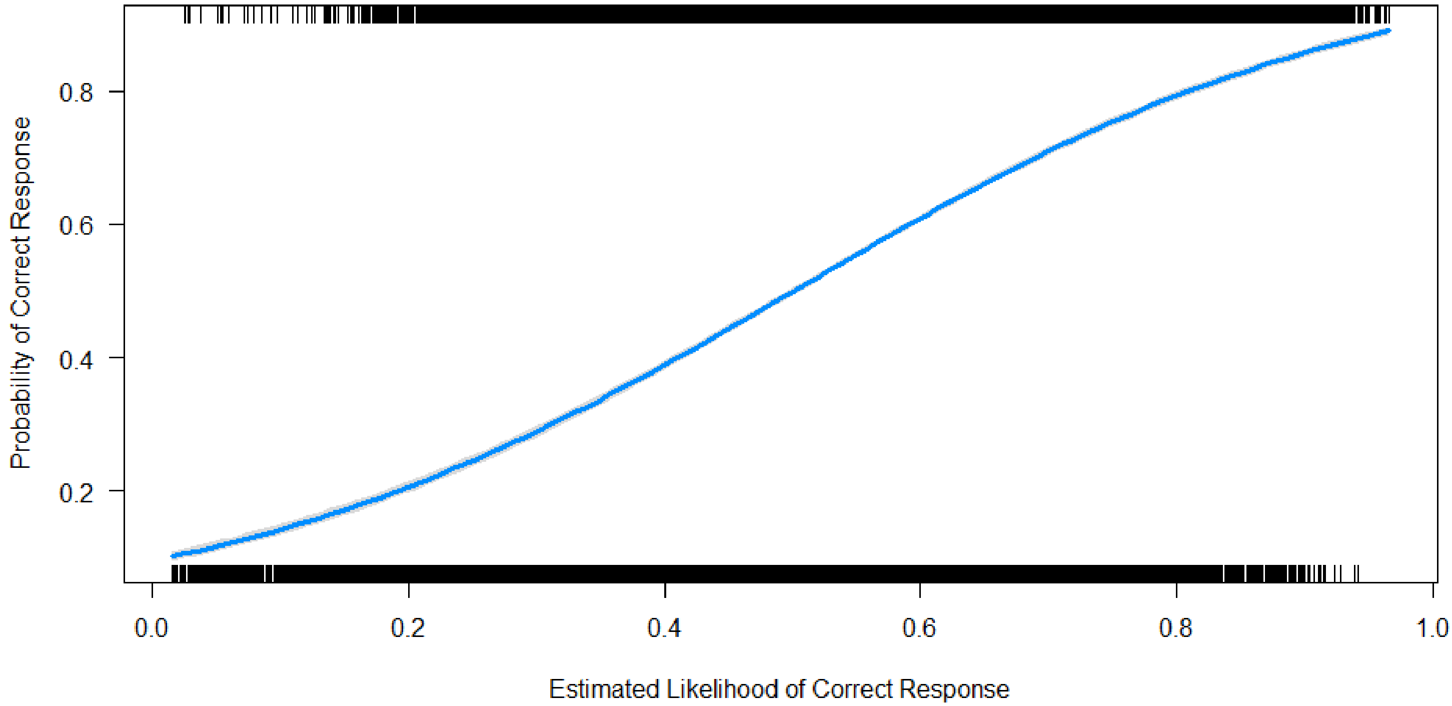

Figure 3.1. Binary Logistic Curve for Predictors of Item-Response (Model 4)

CHAPTER 4

DISCUSSION

Learning to read is vital to success in the educational system and in modern society. Those who do not have difficulties learning to read have enhanced educational outcomes, access to a greater variety of occupational opportunities, and are less likely to have adverse interactions with the educational system itself. Those who struggle to learn to read have more difficult with subsequent educational material and are more likely to find navigating the educational system stigmatizing and isolating. While some children fail to learn to read due to socio-economic issues and lack of adequate educational opportunity, others struggle despite adequate educational opportunity. The identification of children who struggle to learn to read has been an ongoing mission and the methods used to identify children have developed from achievement only gaps, to discrepancies between reading achievement and what would be expected from overall intelligence (i.e., IQ-Achievement Discrepancy), to the inclusion of specific cognitive deficits that are thought to be neurobiological in origin (i.e., Patterns of Strengths and Weaknesses). Identifying children struggling to learn to read due to neurobiological and cognitive reasons earlier in the educational process result in much more effective remediation efforts and outcomes for children, so the identification of these children as early as possible is imperative. Researchers have identified tests that can be used to adequately identify children with learning disabilities specific to reading, such as tests of word reading, reading fluency, orthographic knowledge, and phonological awareness and access measures. However, the implementation of these methods varies widely across settings, with more historically established methods having continued use due to the lack of available expertise and resources available in schools. Efforts have been made to limit the demand of resources on schools by using group administration of rapid fluency-based measures. While these measures have shown promise, they exhibit their own weaknesses, such as the lack of discrimination between those struggling to read due to instructional opportunity and those struggling due to neurological impairment, as well as psychometric issues related to reliability of measurement. Computerized tests have demonstrated the ability to effectively assess for reading comprehension, orthographic knowledge, and spelling abilities. Hoskins and colleagues have created such a battery and have been conducting research on its usability and external construct validity. Computerized batteries like the CARE offer a potential solution to the resource barriers seen in schools. While their tests show adequate convergent validity with other screening measures, the response-modalities for some of the tests deviate from the typical pronunciation- or verbal-based responses required in their paper-pencil counterparts, which might shift the actual task-demands and pathways of response.

The aim of the current study was to identify whether the introduction of picture-based responses the Word Identification test introduced opportunities for unintended test-taking strategies. The study used an explanatory item-response analysis to investigate this question. Each of the variables included represented either a demand on one of the three specific reading strategies thought to be possible in this test, a moderator of the demand for one of the three specific reading strategies, or a general demand involved in the acquisition of reading. A higher degree of statistical influence from LTPF, LTPB, and

MSD would indicate a more frequently used spelling-to-sound strategy, sound-to-spelling strategy, and partial-word reading strategy across items, respectively.

The phonological awareness and receptive vocabulary estimates from the base model (Model 1) uniquely contribute to response-likelihood across models and do not indicate statistically significant changes as item-level variables are introduced, meaning the effects of phonology and receptive vocabulary are relatively consistent in light of added orthographic information. The consistency of these effects likely reflect their importance in some broader reading acquisition process outside of the test-taking environment.

The results from Model 2 suggest that LTPF and LTPB have comparable effect sizes in contributing to the prediction of response-likelihood, with LTPF having a slightly higher statistical contribution when considered the main effects alone. These results suggest that when not accounting for other variables the use of spelling-to-sound and sound-to-spelling strategies are comparable. The MSD of a word has no statistical contribution when considering only the main effects. These results suggest that when not accounting for other variables the clustering of more or less accessible individual phoneme-grapheme units is not involved in task demands. Word-specific frequency (Freq) has a statistically significant contribution in Model 2, indicating that word-specific knowledge contributes to the likelihood of response. However, its contribution is smaller relative to that of the sublexical frequency variables.

The results from Model 3 suggest that of the selected parameters introduced by the response-modality (WC, BOI), only WC has a statistically significant main effect, at least when also considering the interaction of WC and BOI on the reading strategies. The statistical interaction terms introduced in Model 3 also reveal insights into the usage of different reading strategies in the CARE word reading test. When examining the main effects of these three variables without consideration into how they may interact with other elements of the test design (Model 2), the analysis indicates that the first strategy, spelling-to-sound, is used most frequently, with sound-to-spelling being the second-most prominent and the third, partial-word reading, being not statistically significant at all. These results would suggest that this test is primarily a test of typical spelling-to-sound word reading, yet with almost equal amounts of demands of word reading strategies that are specific to this test design. Additionally, the effects of the phonological person-ability estimate on likelihood of response are maintained, even after adding items, item-interactions, which suggest that a person's phonological ability contributes to the word reading test beyond what is required by phonics usage, this is consistent with literature and findings that sight word usage and automatic word recognition are a viable path to word reading as well. However, these results occur when interaction terms are not included within the model and given the "closed-system" nature of the test-taking environment, interactions are almost definite.

The interaction terms in Model 3 were introduced to specifically test for whether the extent that each reading strategy is used is also dependent on the phonemic information available to the participant from the pictures provided. WC served as an indicator of how easily accessible the phonemic information was. WC was allowed to interact with the variables for each strategy. BOI served as an indicator of specific experience with a word and a proxy for the strength of semantic representation each word had. It was also allowed to interact with the variables for each strategy. The researchers expected that WC would reflect a greater facilitation of sound-to-spelling strategy and that BOI would reflect a greater facilitation of spelling-so-sound strategy. Additionally, the researchers anticipated that WC and BOI would have inverse effects on the other reading strategy and therefore provide both convergent and divergent evidence for the constructs that LTPF and LTPB represent. The results of Model 3 were consistent with the hypotheses and revealed that the likelihood of spelling-to-sound reading strategy was statistically significantly increased by a higher rating on BOI, suggesting that words with a stronger semantic representation of the word are more likely to read using spelling-to-sound strategy (LTPF) and are perhaps more likely to be read using spelling-to-sound due to greater familiarity with the word. Conversely, words with a higher WC were less likely to use spelling-to-sound strategy. The results of the interactions on LTPB were also consistent with the hypotheses and revealed that words with less abstract semantic representations (and therefore more easily identifiable) are more likely to be read using the sound-to-spelling strategy and that words with a higher BOI were less likely to use sound-to-spelling.

The current study hypothesized that the extent of moderation itself would be higher after accounting for the availability of partial-word reading and word-specific knowledge. The availabilities of these factors were similarly assumed using the interaction of WC and BOI with MSD and Freq. However, because MSD and Freq are not metrics specific to feedforward or feedbackward consistency, the researchers did not make specific hypotheses to which variables would interact. The results from Model 4 revealed that the interactions of MSD and BOI with WC and BOI were in the same directions as those found in the interactions with LTPB. The consistency of these interactions with LTPB provide further support for the hypothesis MSD is an indicator of partial-word reading and might suggest that the frequency effects typically seen in reading are most facilitating sound-to-spelling strategy.

Overall, the study findings indicate that of the possible strategies of performance on the CARE's word reading test, spelling-to-sound is least context dependent and has the second largest statistical contribution, with sound-to-spelling having the largest contribution after accounting for the statistical interaction terms. The robustness of spelling-to-sound combined with the conditional effect of sound-to-spelling is promising. It means that the researchers can reasonably modify parameters of the items to minimize the effect of sound-to-spelling and maintain the effect of spelling-to-sound strategies.

An additional concern with the introduction of pictures as a response-modality and the possibility for three different strategies across responding for any given item is that the same score across several individuals may represent varying levels of reading ability. Consequently, the normative data collected for the test may not be representative of a single population, but rather several groups with functional distinctions, and that these populations may be indistinguishable through word identification score alone. This may lead to relatively weaker or inconsistent associations with other reading-related constructs.

Limitations

The R-squared for the final analysis indicates a relatively small amount of variability among the responses was explained by the model. The primary limitation of this study is that the variables introduced in this model only account for a modest portion of the available variance and the introduction of additional relevant factors may change the statistical influence of the variables in the current study. For example, the introduction of a variable representative of the fluidity of lexical access may also account for a significant portion of variance and even result in an increase or decrease in statistical effect from the phonological ability or receptive vocabulary ability variables. Another limitation of this study is that the variables and interaction terms used in this study are only a selection of those that could be made available for analysis, particularly if given as part of a massed administration, as is possible with a computerized assessment such as the CARE. For example, the number of vowels in a word typically indicate more opportunity for inconsistent orthographic information and the average length of graphemes may affect the difficulty of acquisition of individual phoneme-grapheme units and may be influential in the model, or at least moderating of the other terms. It is also important to note that the current analysis only takes the parameters of the actual target word into consideration and uses some of these parameters to estimate the effect of the pictural response. This method does not account for the qualities within the target pictures themselves or the qualities of pictures of the alternative responses for each item.

Future Research

Future research on the validity and utility of this test could involve the inclusion of more person- and item-level predictors. It is recommended that analyses be conducted in which other reading-related cognitive constructs are added into the model, ones that would be expected to be more directly involved in the test-taking process, such as lexical access and rapid automatic naming. Additional item-level features could be ones in which the features of the pictures, such as abstraction and phonemic and graphemic overlap, and physical appearance are introduced as moderating effects would help determine how the effects of sound-to-spelling and partial word reading are involved in the test-taking process and how to minimize the effects of partial word reading.

The developers of the CARE Word Identification test may want to consider making modifications to their test items to minimize the potential for sound-to-spelling test performance. The findings of the current study suggest that the sound-to-spelling test performance could be minimized, and consequently maximizing the spelling-to-sound, by choosing words that are less concrete. Less concrete, and more abstract words would offer less phonemic information and therefore require the reader to be more familiar with the word; though using more abstract words may ultimately require more vocabularic ability than what would be optimally desired. So, the researchers may want to consider a different test design altogether, as the use of pictures as a response-modality may limit the extent to which the spelling-to-sound test-taking strategy can be maximized at all.

REFERENCES

Adelman, P. B., & Vogel, S. A. (1990). College graduates with learning disabilities—Employment attainment and career patterns. *Learning Disability Quarterly*, *13*(3), 154-166.

Au, A., & Lovegrove, B. (2001). The role of visual and auditory temporal processing in reading irregular and nonsense words. *Perception*, *30*(9), 1127–1142. https://doi.org/10.1068/p3025

Badian, N. A. (1995). Predicting reading ability over the long term: The changing roles of letter naming, phonological awareness and orthographic processing. *Annals of Dyslexia*, *45*(1), 79–96. https://doi.org/10.1007/BF02648213

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., ... & Treiman, R. (2007). The English lexicon project. *Behavior research methods*, *39*(3), 445-459.

Barga, N. K. (1996). Students with learning disabilities in education: Managing a disability. *Journal of learning disabilities*, *29*(4), 413-421.

Bastug, M. (2014). The structural relationship of reading attitude, reading comprehension and academic achievement. *International Journal of Social Sciences and Education*, *4*(4), 931-946.

Boonen, A. J., de Koning, B. B., Jolles, J., & Van der Schoot, M. (2016). Word problem solving in contemporary math education: A plea for reading comprehension skills training. *Frontiers in psychology*, *7*, 191.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior research methods*, *46*(3), 904-911.

Catts, H. W., Adlof, S. M., Hogan, T. P., & Weismer, S. E. (2005). Are Specific Language Impairment and Dyslexia Distinct Disorders? *Journal of Speech, Language, and Hearing Research*, *48*(6), 1378–1396. https://doi.org/10.1044/1092-4388(2005/096)

Catts, H. W., McIlraith, A., Bridges, M. S., & Nielsen, D. C. (2017). Viewing a phonological deficit within a multifactorial model of dyslexia. *Reading and Writing*, *30*(3), 613–629. https://doi.org/10.1007/s11145-016-9692-2

Christ, T. J. (2017). FastBridge learning. Retrieved September 6, 2017.

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental psychology*, *33*(6), 934.

Ehm, J.-H., Kerner auch Koerner, J., Gawrilow, C., Hasselhorn, M., & Schmiedek, F. (2016). The association of ADHD symptoms and reading acquisition during elementary school years. *Developmental Psychology*, *52*(9), 1445–1456. https://doi.org/10.1037/dev0000186

Elksnin, N., & Elksnin, L. (1996). Making the transition from high school to work or postsecondary training. *Diagnosis and Management of Learning Disabilities: An Interdisciplinary/Lifespan Approach, 3rd edition. San Diego: Singular Publishing Group*.

Ferrer, E., Shaywitz, B. A., Holahan, J. M., Marchione, K. E., Michaels, R., & Shaywitz, S. E. (2015). Achievement gap in reading is present as early as first grade and persists through adolescence. *The Journal of pediatrics*, *167*(5), 1121-1125.

Ferrer, E., Shaywitz, B. A., Holahan, J. M., Marchione, K., & Shaywitz, S. E. (2010). Uncoupling of reading and IQ over time: Empirical evidence for a definition of dyslexia. *Psychological science*, *21*(1), 9101.

Furnes, B., & Samuelsson, S. (2011). Phonological awareness and rapid automatized naming predicting early development in reading and spelling: Results from a cross-linguistic longitudinal study. *Learning and Individual Differences*, *21*(1), 85–95. https://doi.org/10.1016/j.lindif.2010.10.005

Georgiou, G. K., Parrila, R., Kirby, J. R., & Stephenson, K. (2008). Rapid Naming Components and Their Relationship With Phonological Awareness, Orthographic Knowledge, Speed of Processing, and Different Reading Outcomes. *Scientific Studies of Reading*, *12*(4), 325–350. https://doi.org/10.1080/10888430802378518

Heiman, T., & Precel, K. (2003). Students with learning disabilities in higher education: Academic strategies profile. *Journal of learning disabilities*, *36*(3), 248-258.

Hintze, J. M., Ryan, A. L., & Stoner, G. (2003). Concurrent Validity and Diagnostic Accuracy of the Dynamic Indicators of Basic Early Literacy Skills and the Comprehensive Test of Phonological Processing. *School Psychology Review*, *32*(4), 541–556. https://doi.org/10.1080/02796015.2003.12086220

Hogan, T. P., Catts, H. W., & Little, T. D. (2005). The relationship between phonological awareness and reading.

Hoskins, W. H., Hobbs, W. I., Eason, M. J., Decker, S., & Tang, J. (2021). The design and implementation of the Carolina Automated Reading Evaluation for reading deficit screening. *Computers in Human Behavior Reports*, 100123.

Kearns, D. M. (2020). Does English have useful syllable division patterns?. *Reading Research Quarterly*, *55*, S145-S160.

Kim, Y.-S., Otaiba, S. A., Puranik, C., Folsom, J. S., & Gruelich, L. (2014). The contributions of vocabulary and letter writing automaticity to word reading and spelling for kindergartners. *Reading and Writing*, *27*(2), 237–253. https://doi.org/10.1007/s11145-013-9440-9

Lund, K., & Burgess, C. (1996). Hyperspace analogue to language (hal): A general model semantic representation. In *Brain and Cognition* (Vol. 30, No. 3, pp. 5-5). 525 B ST, STE 1900, SAN DIEGO, CA 92101-4495: ACADEMIC PRESS INC JNL-COMP SUBSCRIPTIONS.

MacDonald, G. W., & Cornwall, A. (1995). The Relationship Between Phonological Awareness and Reading and Spelling Achievement Eleven Years Later. *Journal of Learning Disabilities*, *28*(8), 523–527. https://doi.org/10.1177/002221949502800807

Manis, F. R., Doi, L. M., & Bhadha, B. (2000). Naming Speed, Phonological Awareness, and Orthographic Knowledge in Second Graders. *Journal of Learning Disabilities*, *33*(4), 325–333. https://doi.org/10.1177/002221940003300405

Manis, F. R., Seidenberg, M. S., & Doi, L. M. (1999). See Dick RAN: Rapid Naming and the Longitudinal Prediction of Reading Subskills in First and Second Graders.

Scientific Studies of Reading, *3*(2), 129–157. https://doi.org/10.1207/s1532799xssr0302_3

McArthur, G. M., Hogben, J. H., Edwards, V. T., Heath, S. M., & Mengler, E. D. (2000). On the “Specifics” of Specific Reading Disability and Specific Language Impairment. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *41*(7), 869–874. https://doi.org/10.1111/1469-7610.00674

Meneghetti, C., Carretti, B., & De Beni, R. (2006). Components of reading comprehension and scholastic achievement. *Learning and individual differences*, *16*(4), 291-301.

Mesman, G. R., & Kibby, M. Y. (2011). An Examination of Multiple Predictors of Orthographic Functioning. *Journal of Learning Disabilities*, *44*(1), 50–62. https://doi.org/10.1177/0022219410371675

Mundy, I. R., & Carroll, J. M. (2012). Speech prosody and developmental dyslexia: Reduced phonological awareness in the context of intact phonological representations. *Journal of Cognitive Psychology*, *24*(5), 560–581. https://doi.org/10.1080/20445911.2012.662341

Muter, V., & 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.

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of educational psychology, 98*(3), 554.

Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and instruction*, *19*(3), 228-242.

Pexman, P. M., Muraki, E., Sidhu, D. M., Siakaluk, P. D., & Yap, M. J. (2019). Quantifying sensorimotor experience: Body–object interaction ratings for more than 9,000 English words. *Behavior research methods*, *51*(2), 453-466.

Rucklidge, J. J., & Tannock, R. (2002). Neuropsychological profiles of adolescents with ADHD: Effects of reading difficulties and gender. *Journal of Child Psychology and Psychiatry*, *43*(8), 988–1003. https://doi.org/10.1111/1469-7610.00227

Savage, R., & Frederickson, N. (2005). Evidence of a highly specific relationship between rapid automatic naming of digits and text-reading speed. *Brain and Language*, *93*(2), 152–159. https://doi.org/10.1016/j.bandl.2004.09.005

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

Shaywitz, S. E., Holahan, J. M., Kenney, B., & Shaywitz, B. A. (2020). The Yale Outcome Study: outcomes for graduates with and without dyslexia. *Journal of Pediatric Neuropsychology*, *6*(4), 189-197.

Sideridis, G. D. (2003). On the origins of helpless behavior of students with learning disabilities: avoidance motivation?. *International Journal of Educational Research*, *39*(4-5), 497-517.

Sideridis, G. D., Morgan, P. L., Botsas, G., Padeliadu, S., & Fuchs, D. (2006). Predicting LD on the basis of motivation, metacognition, and psychopathology: An ROC analysis. *Journal of Learning Disabilities*, *39*(3), 215-229.

Snowling, M. J., Muter, V., & Carroll, J. (2007). Children at family risk of dyslexia: a follow-up in early adolescence. *Journal of child psychology and psychiatry*, *48*(6), 609-618.

Steacy, L. M., & Compton, D. L. (2019). Examining the role of imageability and regularity in word reading accuracy and learning efficiency among first and second graders at risk for reading disabilities. *Journal of experimental child psychology*, *178*, 226-250.

Stuebing, K. K., Fletcher, J. M., LeDoux, J. M., Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2002). Validity of IQ-discrepancy classifications of reading disabilities: A meta-analysis. *American Educational Research Journal, 39*(2), 469-518.

Suggate, S. P. (2016). A meta-analysis of the long-term effects of phonemic awareness, phonics, fluency, and reading comprehension interventions. *Journal of learning disabilities, 49*(1), 77-96.

Toth, G., & Siegel, L. S. (2020). A critical evaluation of the IQ-achievement discrepancy based definition of dyslexia. In *Current directions in dyslexia research* (pp. 45-70). Garland Science.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (1999). *Comprehensive test of phonological processing: CTOPP*. Austin, TX: Pro-ed.

Wanzek, J., Otaiba, S. A., & Petscher, Y. (2014). Oral Reading Fluency Development for Children with Emotional Disturbance or Learning Disabilities. *Exceptional Children*, *80*(2), 187–204. https://doi.org/10.1177/001440291408000204