EXAMINING THE PSYCHOMETRIC PROPERTIES OF A MORPHOLOGICAL AWARENESS MEASURE IN ELEMENTARY SCHOOL STUDENTS

by

KATHRYN ELIZABETH BANGS

(Under the Direction of Scott Ardoin)

ABSTRACT

The purpose of the current study is to examine the psychometric quality of a frequently employed morphological awareness (MA) measure, the Test of Morphological Structure (TMS; Carlisle, 2000) as well as discuss the implications of altering a measure on research result. Ninety-six fourth grade students from a southeastern elementary school were administered the TMS, a phonological awareness measure, and a reading comprehension measure. The TMS using three item orders (i.e., original, theoretical based on the developmental progression discussed in prior research, and data-based yielded from Rasch analyses) and an applied discontinue rule was evaluated. Results from the current study suggest that the TMS regardless of item order has adequate psychometric properties, but the measures construction needs further investigation with a more diverse population. Additionally, results indicate that prior research (e.g., Carlisle & Nomanbhoy, 1993; Fracasso et al., 2014; Herman et al., 2013; Tighe & Binder, 2013) using the original item order with the applied discontinue rule may not provide as accurate a picture of an individual's MA ability as researchers believe.

INDEX WORDS: Morphological awareness, reading, measurement, Rasch analysis

EXAMINING THE PSYCHOMETRIC PROPERTIES OF A MORPHOLOGICAL AWARENESS MEASURE IN ELEMENTARY SCHOOL STUDENTS

by

KATHRYN ELIZABETH BANGS

B.A., Mount Holyoke College, 2011

M.A., Mount Holyoke College, 2013

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2018

EXAMINING THE PSYCHOMETRIC PROPERTIES OF A MORPHOLOGICAL AWARENESS MEASURE IN ELEMENTARY SCHOOL STUDENTS

by

KATHRYN ELIZABETH BANGS

Major Professor: Scott Ardoin
Committee: Amy Reschly
Ashley Harrison
George Engelhard

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
August 2018

ACKNOWLEDGEMENTS

During the process of creating this work, I have learned so many lessons from many supportive and wonderful people. This dissertation and my graduate schooling would not have been possible without each and every one of them. Each has provided me guidance, advice, and encouragement. I cannot even begin to thank them for everything they have done and how much it means to me. This acknowledgement page is never going to encompass my gratitude for all those who have guided me on my journey.

My dissertation committee could not have been better! Thank you to Drs. Ashley Harrison, Amy Reschly, and George Engelhard. I would especially like to thank my committee chair, advisor, and mentor, Dr. Scott Ardoin. Without his support and direction, I would not have learned and grown as much over the past several years. I am very thankful for all that my advisor and committee members have done for me over the last five years and for our future friendship and continued mentorship. I am so grateful that they were willing to share your time, wisdom, and expertise with me while guiding me in my work.

There were many people that helped me conduct my research including the Ardoin Lab and the school district. Everyone was very supportive and receptive of my work. The research assistants, administrators, staff, and teachers made the process easy for me. I also would like to thank my masters' advisor, Dr. Katherine Binder, without whom I would have never made it to UGA, into the field of school psychology, or into the literacy research realm. I want to thank them for being so gracious with their time and support!

I was very fortunate to have a large support system that supported me and cheered as I went through the dissertation process. Thank you to all those who listened to my woes, reminded me to engage in self-care, maintained good spirits and humor, and provided friendship throughout my life.

Last, but not least - Thank you to my family as without them I would never have considered taking this challenging journey. They have all given me constant and unwavering support. They reminded me to persevere when things were difficult and encouraged me to work hard in pursuing my dreams. I want to especially thank my mother for her undying support, encouragement, and guidance throughout my life. Without her, I would have never come this far.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Many public schools utilize the Response-To-Intervention (RTI) model to aid their decision making and assist in providing needed services to their students. RTI is a multi-tiered system of service delivery in which data are collected to screen for at-risk students, monitor progress, and aid in making appropriate educational decisions (Reschly & Bergstrom, 2009). If a student is identified as struggling via screening, they are further assessed in the particular area in which they are struggling. The resulting data helps inform appropriate instructional decisions including moving a student up or down a tier in order to provide a free, appropriate education and to match the instructional level that the student needs to assist his or her learning. Furthermore, the data resulting from assessments, universal screening, and progress monitoring are used by schools to meet requirements set by federal laws such as Individuals with Disabilities Education Improvement Act (IDEIA, 2004) and the Every Student Succeeds Act (ESSA) of 2015, the newest amendment to the No Child Left Behind (NCLB) Act of 2001.

Legislation, Its Evolution, and How It Affects Assessment

IDEIA (2004) is a law that requires certain protections and services to be provided to children with disabilities and stems from Public Law (PL) 94-142 established in 1975. The first edition of IDEIA occurred in 1990 when PL-94-142 was combined with PL-99-457 and was called the Individuals with Disabilities Act (IDEA, 1990). The two public laws, PL-94-142 and PL-99-457, provided the same protection of a free and appropriate public education (FAPE) for children with disabilities of different ages, ages 6-21 and ages 3-6, respectively. The law was amended again in 1997 to emphasize the role of parents in decision making, the use of Individual Education Plans (IEP) with measureable goals for academic problems and Functional Behavior Assessments (FBAs) for behavioral problems, and to promote access to the general education curriculum (Jacob, Decker, & Hartshorne, 2011; Sattler, 2008). The most recent amendment, renamed IDEIA (2004), highlighted that written plans require measureable goals that are monitored and reported along with complete evaluations every three years (Jacob et al., 2011; Sattler, 2008).

IDEIA (2004) inspired the use of the RTI system and evidence-based practices and interventions. Specifically, IDEIA (2004) highlighted three essential components to identify and serve struggling students successfully. First, the law recommended that all students be screened to identify those who were at-risk before problems took hold. Second, identified students would then be monitored for a minimum of twelve weeks. The progress monitoring data collected during those weeks would then be used to inform educational and instructional decisions such as movement through the multi-tiered system and identification of disabilities.

Whereas IDEIA was amended to encourage earlier identification and early intervention, NCLB (2001) holds school districts responsible for students' status and progress towards standards of achievement. NCLB (2001) was written as a general education law that amended the Elementary and Secondary Education Act of 1965 rather than as a law designed to address students with disabilities such as IDEIA (2004). Thus, NCLB (2001) held school districts responsible for the performance of *all* students within their district and required schools to monitor progress and publicly report the school's overall assessment data. The 2004 amendment of IDEIA helped align the accountability implications dictated by NCLB (2001) in a stronger way.

The newest amendment of NCLB, ESSA (2015), aligns more closely with IDEIA (2004) than its predecessor. Like IDEIA (2004), ESSA (2015) highlights the use of a multi-tiered system that utilizes universal screening, progress monitoring, and data-based decision making. In terms of accountability, it reiterates that schools need to measure performance annually. It adds that a supplementary reliable and valid measure needs to be administered to students in addition to typical state assessments to demonstrate student progress and growth (“The Every Student”).

With increasing legislation (e.g., ESSA, IDEIA, etc.) and advances in data systems and analytic methods, accountability for student performance and mandated screening and progress monitoring in grades Kindergarten through 12^{th} has amplified the importance of using reliable and valid assessment and screening measures. Data resulting from assessments and screening are used for many educational decisions including assessing teacher effectiveness, special education eligibility determinations, and individualized educational plans. The interpretations of the assessment and screening results are not only affecting the assessed students but also future students, examiners, teachers, and other service providers. With a wide range of students being assessed and the results impacting numerous types of decisions, the measures used need to represent students’ skills consistently and accurately.

Screening for Literacy Skills

Literacy skills are a particular target for screening among educators. Educators want to identify potential problems early in the development of literacy skills to prevent future educational difficulties. Learning to read in English is a long and difficult process (Bowers, Kirby, & Deacon, 2010; Mesmer & Griffith, 2005). Numerous years of instruction and practice are needed to become a proficient reader. Some of the difficulty in learning to read is due to the structure of English as a morphophonemic language. This means that it is influenced by two structures: 1) phonological structure and 2) morphological structure (Bowers et al., 2010; Mesmer & Griffith, 2005).

The phonological structure consists of sound-units (e.g., /*m*/ and /*ch*/) whereas the morphological structure is comprised of units of meaning, called morphemes. There are several types of morphemes including root words (e.g., fish, plant, etc.) and affixes, which are further categorized as prefixes (e.g., *in*-, *pre*-, etc.) and suffixes (e.g., -*ing*, -*able*, etc.). It is the interplay between the two structures that makes English challenging (Carlisle, 2012).

Teachers devote the majority of instruction time on developing literacy and language skills. Reading comprehension and critical thinking are also considered highly important. Research has demonstrated the significance of rudimentary abilities, such as phonological awareness, in the process of reading comprehension (Adams, 1990). Schools spend millions of dollars on universal screeners, interventions, and instructional materials targeting skills such as phonological awareness, reading fluency, and reading comprehension to help improve student performance (Sharpe, 2012). Specifically, significant time is devoted to the screening for the understanding of the phonological structure of English to glean information in order to improve instruction and student performance. Although phonological awareness aids students' understanding of a large percentage of the English language (Mesmer & Griffith, 2005), the phonological structure is only part of the battle, particularly in later school years.

As children mature and become proficient readers, they encounter more morphologically complex words (Anglin, 1993; Nagy, Berninger, & Abbott, 2006). Phonological awareness aids students in making estimated pronunciations of unknown words. However, the morphological structure also influences the pronunciations of words. For example, the /*sh*/ sound in *dishonest* and in *dish* are pronounced differently due to the phonological structure influencing *dish* and the phonological and morphological structures influencing *dishonest* (Carlisle, 2012). Furthermore, research provides evidence that morphological awareness (MA) also influences students' reading achievement (Deacon & Kirby, 2004; Kirby et al., 2011; Wolter, Wood, & D'zatko, 2009). Despite the importance of MA, schools do not generally assess students' MA skills as part of their current universal screening process, possibly because a reliable and valid measure for assessing students MA skills does not currently exist. Research is needed to develop and refine such a measure in order to fully capture a student's skills and to inform educational practice.

When creating, considering, or evaluating a measure, theory and prior research need to be explored in order to understand the target construct fully before analyzing measurement issues (Wilson, 2005). To do this, the construct must first be defined along with its factors. MA is defined as the "conscious awareness of the [morphological] structure of words and the ability to reflect on and manipulate their structure" (Carlisle, 1995, p. 194). The definition of MA also encompasses the integration of semantic and phonological knowledge. Although semantic and phonological knowledge are integral aspects of MA, the construct of MA is impacted by various other factors, which need to be understood in order to evaluate or create an effective measure. One factor that must be considered in understanding the construct of MA is that morphemes exist in a hierarchy with the distinction of root words and affixes (i.e., prefixes and suffixes) being first. Each level within the hierarchy contributes to the complexity of MA and its development in students. A second factor to consider is the clarity of the relationship between the root word and the morphologically complex word (i.e., root words with added prefixes and/or suffixes). The relationships between words add to the complexity of MA. Morphologically complex words that preserve the sound and spelling of the root word are acquired before those that do not preserve the sound and spelling (Ku & Anderson, 2003). The ability to recognize relationships between words that are disguised by sound and spelling shifts increases with age (Carlisle, 1995; Carlisle & Nomanbhoy, 1993; Clark & Cohen, 1984; Jones, 1983; Mahony, Singson, & Mann, 2000; Wolter et al., 2009). Word frequency is a third factor that affects the development of MA and influences students' ability to recognize relationships between words. The more frequent a word is employed in the English language, the sooner a word is acquired (Clark & Cohen, 1984; Gordon, 1989) and this acquisition begins as early as infancy (Lieven, 2010). Thus, frequency makes certain morphologically complex words easier to learn due to the effects of exposure.

As briefly described, MA has several factors that influence its complex developmental progression including the types of morphemes, relationship clarity, and word frequency. The complexity of MA needs to be reflected within measures assessing MA. Wilson (2005) recommended that understanding theory and prior research is vital to creating or evaluating measures as it allows educators and researchers to adequately represent a construct within a measure.

Purpose

Inconsistency in the measures used within prior MA research suggests that researchers do not believe an adequate measure of MA exists as they frequently adapt or modify measures from one study to the next. The focus of this study was to investigate the psychometric properties of a commonly used, but modified, measure of MA as a first step in identifying a potential screener and/or diagnostic measure of students' MA skills. The limitations of modifying a measure without inspecting the reliability and validity of the data produced by that measure were examined. In addition, the impact of a seemingly small modification may impact the results of a study investigating the relationship between MA and reading comprehension was explored.

Chapter 2 discusses the previous literature regarding the development of MA and its contribution to reading achievement. First, the chapter reviews the importance of using universal screeners within the educational system and in aiding in plans to prevent and remediate potential reading problems. Next, a discussion of the definition, conceptualization, and factors that contribute to the complexity of MA ensues. It is imperative to understand this literature before delving into any measurements issues when creating or evaluating a measure (as recommended by Wilson, 2005). Importantly, the chapter highlights the existing measurement issues in current measures assessing MA as well as discussed the potential impact those issues have on results within prior research.

The specific research questions, methodology, and analyses for the present study are presented in Chapter 3. Fourth grade students enrolled in a public elementary school were recruited for the study. Each student completed several measures assessing phonological decoding, reading comprehension, and MA. The items in the two morphological tasks that comprise the Test of Morphological Structure (Carlisle, 2000) were administered randomly to each student. After administration is complete, the data were analyzed based upon three potential item-orderings each with a discontinue rule applied: original order (i.e., items ordered in the way the test was designed), theoretical order (i.e., items ordered based upon the developmental trajectory of MA), and data-based (i.e., items ordered based upon item analysis, specifically Rasch analysis).

Analyses were conducted to, first, examine the reliability and validity of these item-ordering procedures with an applied discontinue rule and to examine the extent to which the resulting scores from each item-order procedure differ from one another. Next, regression analyses were conducted to understand how each of these item-ordering procedures may impact the amount of variance MA explains in reading comprehension after phonological awareness is controlled for. Specific hypotheses included (a) that the item-ordering procedures based on prior research and data would have stronger psychometric properties than the original, non-hierarchical item-order after a discontinue rule has been applied; (b) all three item-ordering procedures of the Test of Morphological Structure (Carlisle, 2000) have significant relationships with phonological awareness and reading comprehension; and (c) the data produced from each item-order explains variance within reading comprehension above the effects of phonological awareness with the item-ordering procedures based on prior research and data explaining more variance than the original, non-hierarchical item-order.

CHAPTER 2

REVIEW OF THE LITERATURE

The National Association of School Psychologist (NASP) states that decisions made in professional practice need to be driven by data (McDonald, 2013; "Model for Comprehensive", 2010). Data provides evidence for or against a particular decision, such as providing intervention services. This evidence often comes in the form of assessment data. Thus, assessment comprises a major portion of school activity and has several purposes, including screening, diagnosis, and intervention planning and evaluation (Sattler, 2008). Universal screening is particularly important in comprehensive assessment systems (Hosp, Hosp, & Dole, 2011). It includes the assessment of academic, behavioral, social, and/or emotional outcomes of all students within a district (Hosp et al., 2011). In addition, it is often considered the first step of the problem-solving process (Petscher, Kim, & Foorman, 2011).

Universal screening is the most common method of both preventing and remediating academic and/or behavior problems in the school setting (Cook, Volpe, & Livanis, 2010; McDonald, 2013). Screening allows for preventative interventions to disrupt an undesirable academic or behavior path (Eklund et al., 2009) as well as provide data regarding the effectiveness of the core curriculum (McDougal, Leblanc, & Hintze, 2010). Universal screening also allows for appropriate interventions to be chosen and delivered in order to remediate problems (Cook et al., 2010; Hosp et al., 2011).

Federal legislation supports universal screening and assessment practices in the educational system. The No Child Left Behind Act of 2001 (NCLB, 2001) highlighted the importance of assessment and specifically endorsed the need to identify at-risk students through the use of universal screening (Cook et al., 2010). NCLB's replacement, the Every Student Succeeds Act of 2015 (ESSA, 2015), maintains the importance of assessment and notes that a reliable and valid academic measure needs to be used to demonstrate student growth in addition to required state assessments ("The Every Student"). Additionally, the Individuals with Disabilities Education Improvement Act (IDEIA, 2004) mandates child-find procedures and screening for potential disabilities (Cook et al., 2010).

Due to the wide use of universal screening and the endorsements of its use by several educational and psychological associations (e.g., NASP, American Psychological Association, Council for Exceptional Children, etc.), screening measures need to be valid and reliable. Petscher et al. (2011) state that screening measures should be brief, easy to use, and able to predict student success with high accuracy. Eklund et al. (2009) warn against using measures that may over-identify students. Although over-identification is a concern, universal screening measures help identify more at-risk students than relying on teacher referrals and tend to be more accurate and consistent indicators (Eklund et al., 2009). It is, however, important to recognize that not just any assessment measure can be used for universal screening purposes. It is essential that prior to a measure being used for universal screening purposes, its technical adequacy and utility for identifying students with instructional needs be examined and documented.

Numerous universal screening and early identifications measures are being used within the educational setting, most of which focus on reading achievement and skills such as letter-sound recognition, phonological awareness (PA), and reading fluency. These measures are given to help identify students who may develop or who have difficulties in reading in order to prevent or remediate future reading problems. In the early grades, pre-reading skills such as letter naming fluency, letter sound fluency, and nonsense word fluency are assessed (Joseph, 2006). The universal screening measures in reading within later grades are, however, somewhat limited. Unless schools employ a computer adapted measure, which measures a combination of skills (e.g., phonics, vocabulary, reading comprehension), the only reading skill typically assessed is students' oral reading fluency (Sharpe, 2012).

One reading skill that is typically not assessed by universal screeners is morphological awareness (MA). MA is the awareness and ability to manipulate the morphological structure of words (Carlisle, 1995). Assessing students' MA could aid in identifying struggling students. The purpose of this paper is to describe the construct of MA, research describing its relationship with reading achievement (particularly comprehension), and measurement concerns associated with existing MA measures. Prior to delving into measurement concerns and components of MA, literature describing the importance of MA within reading achievement will be discussed.

Language and Literacy Development

Educators are particularly concerned with screening for and identifying potential problems in literacy skills. Literacy skills (i.e., reading and writing) are generally complex and require years of learning and practice before students become proficient in reading and writing. The English language, a morphophonemic language, is particularly difficult to learn and measure because it is influenced by both phonological and morphological structures (Bowers et al., 2010; Mesmer & Griffith, 2005). Words are constructed first by their phonological structure and, for some complex words, a morphological structure is then superimposed (Mesmer & Griffith, 2005; Nagy et al., 2006). Phonological structure refers to sound-units (i.e., phonemes), such as */t/* and */sh/,* whereas morphological structure refers to the smallest units of meaning (i.e., morphemes), such as root words (i.e., *cat*, *tree*, etc.) and affixes (prefixes and suffixes; See Figure 1; Apel,

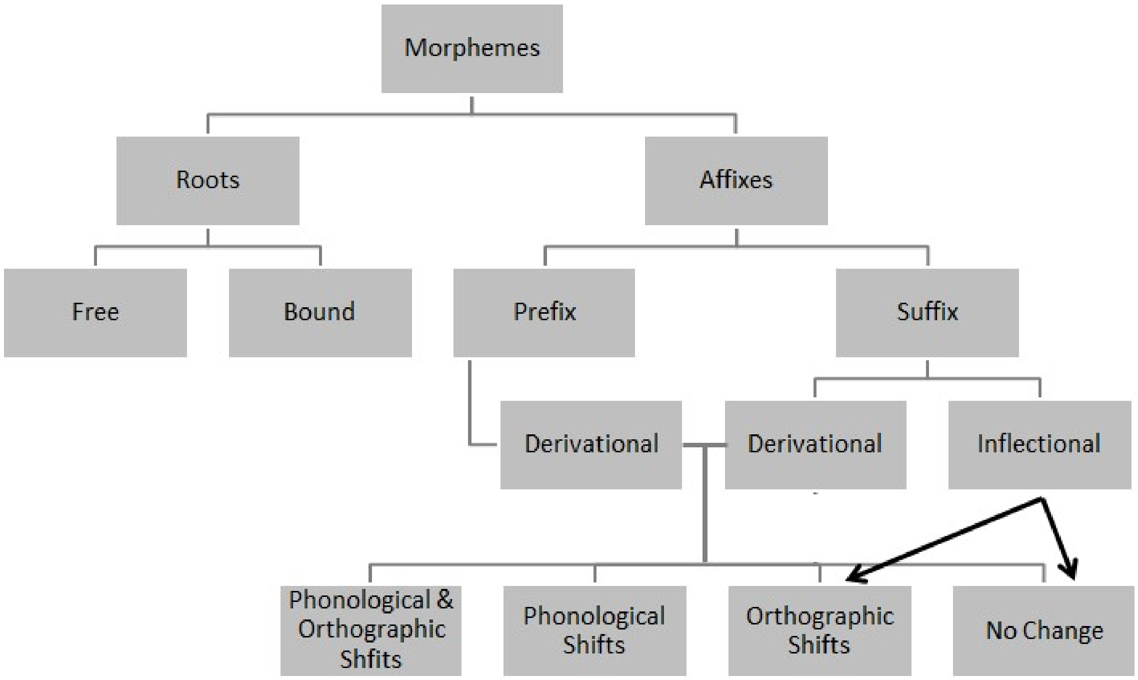

Figure 1. Morpheme hierarchy

Dieham, & Apel, 2013; Carlisle & Katz, 2006; Carlisle & Stone, 2005). This mixed structure makes the English language fairly unpredictable due to the various language and grammar rules and conventions that may be applied to any given word.

The pronunciation of a word could follow the phonological structure predominantly or use a combination of phonological and morphological structures (Cazden, 1968; Deacon & Kirby, 2004); thus, making a word more challenging to decode. For example, the /*sh*/ sound in *misheard* and in *fish* are pronounced differently due to the pronunciation of fish being based strictly upon its phonological structure whereas the pronunciation of misheard is based both upon its phonological and morphological structure (Carlisle, 2012). Furthermore, phonology does not require the comprehension or understanding of word definitions, whereas morphology is concerned with the definitional aspects of word comprehension and vocabulary development (Nagy, Berninger, Abbott, Vaughan, & Vermeulen, 2003). Thus, mastering the phonological structure is only half of the battle for reading acquisition.

As children become more proficient readers, they encounter more morphologically complex words since approximately half of all English words contain more than one morpheme (Anglin, 1993; Nagy et al., 2006). Thus, children must gain an understanding of morphemes (i.e., MA) in order to decode and understand the meaning of morphologically complex words. Knowledge of affixes (i.e., prefixes and suffixes) is especially important as it allows students to break words down into components (e.g., root words and affixes) similar to how knowledge of phonics is necessary to break words into sound components. MA aids students in their decoding and word recognition skills by augmenting their decoding processes; thus, making MA an important reading skill for educators to measure.

The Definition, Conceptualization, and Components of MA

When considering or evaluating a measure, Wilson (2005) suggests that research, theory, and other literature needs to be explored and understood before diving into measurement issues and concerns. The first step in measurement is identifying and defining the construct and its components established on the research-base. Wilson discussed that when designing items of a measure, two fundamental aspects need to be considered: content components and distributional components. Content components are the aspect of the construct of which a researcher is attempting to measure, whereas the distributional components are the ways in which items are classified, such as difficulty level (Wilson, 2005). This section will: (a) provide a description and overview of the complex construct of MA and (b) discuss various components of MA as identified by existing research (e.g., Ku & Anderson, 2003; Kuo & Anderson, 2006; Mahony et al., 2000).

MA is often defined as a metalinguistic ability (Apel et al., 2013; Wolter & Pike, 2015) that encompasses the "conscious awareness of the [morphological] structure of words and the ability to reflect on and manipulate their structure" (Carlisle, 1995, p. 194). Researchers have suggested that orthographic, semantic, phonological, and morphological abilities are interdependent (Carlisle, 2003; Fowler & Liberman, 1995). Kuo and Anderson (2006) discussed a model conceptualizing the interdependence of these abilities (Figure 2). Their model consists of three main abilities: PA, semantic awareness, and orthographic awareness. These three abilities overlap, creating four intersecting areas: (a) grapho-morphological awareness, the overlap of PA, semantic awareness, and orthographic awareness; (b) MA, the overlap of semantic awareness and PA; (c), grapho-semantic awareness, the overlap of orthographic awareness and semantic awareness; and (d) grapho-PA, the overlap of orthographic awareness

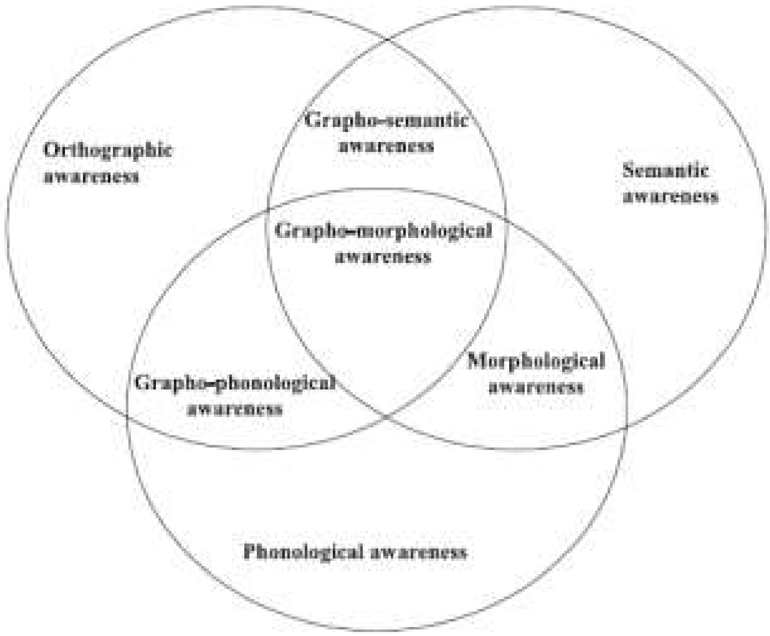

Figure 2. Interrelation between Different Aspects of Metalinguistic Awareness as presented by Kuo and Anderson (2006).

Table 1

Kuo & Anderson's (2006) Model Definitions (p.162)

Term	Definition
Orthographic Awareness	the ability to manipulate and reflect upon the written representations of one's language
Phonological Awareness	the ability to manipulate and reflect upon the sounds of one's language
Semantic Awareness	the knowledge about how meanings are organized in a language
Grapho-Phonological Awareness	the knowledge about the grapheme-phoneme and phoneme-grapheme conversion rules of a language
Grapho-Semantic Awareness	the knowledge about how semantic information is encoded in the orthography and how orthography provides cues to meaning
Morphological Awareness	the knowledge about the pairings of sound and meaning in a language
Grapho-Morphological Awareness	the ability to coordinate orthographic, phonological, and semantic information during reading

and PA, which may also be conceptualized as phonics. See Table 1 for definitions of each of these skills from Kuo and Anderson. This manuscript will presume that MA is the overlap of semantic and phonological information without orthographic knowledge as conceptualized by Kuo and Anderson. They defined MA as "the knowledge about the pairings of sound and meaning in a language" (Kuo & Anderson, 2006, p.162). Semantic awareness and PA are the aspects that researchers have generally agreed upon whereas the integration of orthographic skills is debated (i.e., grapho-morphological awareness; Jannedy, Poletto, & Weldon, 1994; Kuo & Anderson, 2006).

It is important to understand that, although semantic and phonological knowledge are integral aspects of MA, they are not the only constituents within the construct of MA. Other components that impact the measurement of MA include *morpheme type* (e.g., inflectional, derivational, etc.), *levels of transparency* (e.g., phonological transparent, phonologically opaque, etc.), and *word frequency*[1]. A discussion of each of these components and their related literature are discussed below.

Morpheme type. There is a hierarchy of morpheme types (see Figure 1) existing of two primary types of morphemes: root morphemes (e.g., dog, crawl) and affixes (e.g., pre-, -ed). Root morphemes carry the main component of a word's meaning and, in some cases, are a word itself. Root morphemes that can stand alone as a word are called free morphemes (e.g., sing) whereas bound morphemes cannot stand on their own without the addition of other morphemes (e.g., struct as in construct; Bowers & Kirby, 2010).

[1] Other general measurement issues include item presentation, question focus (relational, semantic, or distributional components), and response requirement (oral, written, etc.). However, these general aspects will not be the focus of this paper. Rather, the aspects specific to MA will be explored within this section.

The second type of morphemes, affixes, consist of prefixes and suffixes, which are typically easier than root morphemes for students to recognize and manipulate, and are the building blocks for creating morphologically complex words. Affixes are always bound morphemes and are attached to root morphemes either directly, as in the word *singing*, or indirectly by being attached to another affix, such as the *–ly* in *interestingly* (Bowers & Kirby, 2010). Affixes assist in the meaning of words but are not the primary part of a word's definition. Although affixes are often easy for students to identify and manipulate when learning new words, they also exist on a hierarchy of difficulty or complexity (Bowers & Kirby, 2010). There are two types of morphological complexity: inflectional and derivational (Carlisle, 2000). Understanding these types and their components helps researchers develop measures to assess students' abilities.

Inflectional morphology. Inflectional morphemes typically preserve the root morpheme but change the quantity or tense of that morpheme by adding a suffix (Berko, 1958; Kuo & Anderson, 2006). According to Anglin (1993), there are eight inflectional morphemes:

> "the plural inflection (e.g., the *–s* in *cats),* the possessive inflection (e.g., the *-'s* in *mother's),* the third person-singular verb inflection (e.g., the *-s* in *jumps),* the progressive inflection (e.g., the *-ing* in *soaking),* the past-tense inflection (e.g., the *-ed* in *jumped),* the past participle (e.g., the *-en* in *fallen),* the comparative inflection (e.g., the *-er* in *sourer),* and the superlative inflection (e.g., the *-est* in *fairest*)" (p. 18).

Sixty-five percent of all suffixed words are comprised of the -ed, -ing, and plural (-s and -es) suffixes (White, Sowell, & Yanagihara, 1989). Just by having an adequate understanding of these three suffixes, students have a strong knowledge base of morphologically complex words.

An understanding and usage of inflectional morphology is often the first to develop during the process of morphological awareness acquisition. Numerous studies have demonstrated that inflectional knowledge develops before reading instruction begins and follows a relatively fixed developmental course: (a) children learn a simple concept, (b) overgeneralize that concept, and, then, (c) learn to distinguish proper versus improper uses (Cazden, 1968; Kuczaj, 1977, 1978; Kuo & Anderson, 2006; Marcus et al., 1992). Researchers have suggested that inflectional understanding and usage often begins in children as young as 18 months old (Cazden, 1968; Kuo & Anderson, 2006), and significantly increases in early elementary school with the use of plural, possessives, and past tense (Cazden, 1968; Kuo & Anderson, 2006; Wolter et al., 2009). Evidence of this knowledge may be observed through children's manipulation of inflections in oral language production (Kuo & Anderson, 2006).

Based on a longitudinal study of three participants' spontaneous speech, Cazden (1968) stated that when looking at noun and verb inflections, plural forms appeared prior to possessive use, and that progressive tense (i.e., action that is in progress) tended to emerge before past and present verb tenses. Similarly, Berko (1958) used nonwords paired with pictures to assess young children's inflectional usage. Preschool and first grade students were presented with a nonword, such as *wug*, paired with a picture (see Appendix A). They were then presented with a picture of two *wugs* and asked to name the picture. Results indicated that the students demonstrated understanding of basic morphological rules (i.e., plurals, verb tense, compounds, possessives, etc.) suggesting that young children acquired some inflectional knowledge and were able to apply that knowledge to novel words (Berko, 1958; Ku & Anderson, 2003; Kuo & Anderson, 2006).

Derivational morphology. Derivational morphemes are comprised of both prefixes and suffixes and alter the meaning of the root morpheme or change the part-of-speech (Anglin, 1993; Carlisle, 2003; Ku & Anderson, 2003; Kuo & Anderson, 2006; Tyler & Nagy, 1989). For example, adding *-ness* to the root word *happy* changes the part of speech from an adjective to a noun, whereas adding *un-* to *happy* changes the definition to the opposite of happy. Unlike inflectional morphemes, there are many more derivational morphemes in the English language, making this area of morphology more difficult to master (Kuo & Anderson, 2006). Derivational morphology comprises a large portion of the English language. In fact, there are approximately 14,000 derivatives compared to 4,000 root words (Ku & Anderson, 2003). Many have argued that because of the number of words that are derivational both in written and oral language, this area of morphology is particularly important (Carlisle, 1995; Ku & Anderson, 2003). However, with the variety and number of derivational affixes, many studies found that a student's development of derivational knowledge had a longer progression (Derwing & Baker 1979; Nagy, Diakidoy & Anderson 1993; Singson, Mahony, & Mann, 2000; Tyler & Nagy 1989). Generally, oral and written language derivational knowledge increased gradually across grade levels (Mahony et al., 2000; Singson et al., 2000; Tyler & Nagy, 1989).

Evidence of the gradual derivational development begins around kindergarten/first grade (Anglin, 1993; Gordon, 1989). However, Anglin (1993) and Carlisle (1995) suggested in their syntheses of the literature that the usages and development of derivational knowledge truly commences around third grade and has a sharp increase around fifth grade. In third or fourth grade, knowledge and awareness of derivational structure and meaning becomes more explicit (Kuo & Anderson, 2006). Mahony et al. (2000) assessed the derivational knowledge of students in third to sixth grade. Results revealed that derivational knowledge increased with grade level

(Mahony et al., 2000), with performance for all suffix types having improved equally over time (as demonstrated in their additional study; Singson et al., 2000). Generally, younger students were more likely to accept false relationships (i.e., words that appear to have the same root word) between words than older students (Mahony et al., 2000). By sixth grade, children were approximately 90% correct when identifying derivational relationships (Singson et al., 2000). Accuracy and usage continued to develop throughout middle and high school (Kuo & Anderson, 2006).

Types of derivational morphology knowledge. One of the reasons it takes children longer to develop derivational morphology knowledge is due to multiple types of derivational morpheme knowledge. The development of derivational morphology knowledge includes relational, syntactic, and distributional knowledge of derivational morphemes. Each knowledge type relies on the knowledge type that precedes it. *Relational knowledge* tends to appear first, followed closely by *syntactic knowledge* and, finally, *distributional knowledge*.

Relational knowledge is the recognition of root words and how root words are related to suffixes (Kuo & Anderson, 2006). This knowledge includes rejecting false relationships as well as identifying correct relationships between two morphologically complex words (Derwing & Baker, 1979; Mahony et al., 2000). For example, students may believe the words *passable* and *passion* are related because they both have *pass* in them (i.e., a false relationship). Relational knowledge appears when individuals are four to five years of age (Clark & Cohen, 1984) and continually increases through eighth grade (Derwing & Baker, 1979; Mahony et al., 2000; Ku & Anderson, 2003; Kuo & Anderson, 2006; Tyler & Nagy, 1989). Mahony et al. (2000) pointed out that the relationships between root words and complex words are often obscured by sound and spelling shifts. This complexity may require the continued development of MA throughout a student's education. Young students have more difficulty identifying false relationships, especially when multiple spelling and/or sound shifts occur. Older students are less likely to accept false relationships and are more likely to identify correct relationships between words that they learned in school such as "*moon* and *month"* even when both phonological and orthographic shifts exist (Mahony et al., 2000, p. 203).

Syntactic knowledge is the understanding that particular suffixes are associated with particular parts-of-speech (Tyler & Nagy, 1989). Unlike relational knowledge, syntactic knowledge is not developed until formal instruction in reading and writing begins. Because the syntactic property and the sentence's syntactic structure are typically a redundancy, readers are able to read through material faster and comprehend the material with greater ease (Rayner & Pollatsek, 1989). Kuo and Anderson (2006) suggested that children may be able to recognize the syntactic properties of morphologically complex words around fourth grade. Younger students may not recognize that the verb tense of a sentence is incorrect. For example, an elementary student may view this sentence as correct: "Yesterday, I walk to school." Syntactic knowledge allows students to understand that, when a sentence is past tense (e.g., *I walked to school*), a different suffix may be added to a verb then when a sentence is in the present tense (e.g., *I walk to school*). It is common for younger students to struggle with part-of-speech knowledge, and Kuo and Anderson (2006) stated that many older students also have difficulty with understanding these properties. Therefore, it seems that syntactic knowledge has a long developmental period (Kuo & Anderson, 2006).

Distributional knowledge is the understanding of how the syntactic properties of the root word constrains affixes (Kuo & Anderson, 2006). For example, the suffix *–ly* cannot be attached to nouns (e.g., successly), but can be attached to adjectives (e.g., successfully; Kuo & Anderson, 2006). This knowledge is considered the most challenging aspect of derivational morphology. It is the last aspect to develop because it relies on the understanding of syntactic (i.e., part-of-speech) and relational (i.e., recognizing root words) knowledge; thus, the development of relational and syntactic knowledge has to occur prior to distributional knowledge development (Kuo & Anderson, 2006).

To examine how relational, syntactic, and distributional knowledge impact derivational morpheme acquisition, Tyler and Nagy (1989) had fourth, sixth, and eighth grade students complete tasks targeting these three knowledge areas. The relational knowledge measure consisted of multiple choice items that asked students to define a given word in a content sentence. For example, "'I'm in a celebratory mood,' Mary announced. Mary felt like: (a) having a party, (b) being alone, (c) going to sleep, (d) having a fight, and (e) don't know" (Tyler & Nagy, 1989, p. 654). The syntactic knowledge measure was also comprised of multiple choice items in which students were asked to complete a context sentence. For example, "You can ____ the effect by turning off the lights. Fill in the blank: (a) intensify, (b) intensification, (c) intensity, and (d) intensive" (Tyler & Nagy, 1989, p. 656). The distributional knowledge measure asked students to indicate whether they knew the definition of words presented in a list by circling *Yes* or *No* (e.g., "Do you know what *tameness* means?;" Tyler & Nagy, 1989, p. 660). Overall, Tyler and Nagy (1989) found that knowledge of relational, syntactic, and distributional properties increased with age. Their findings suggested that students were able to distinguish between properly and improperly formed derived words and were more likely to accept words in which suffixes were attached to free root morphemes (i.e., root morphemes that can stand on his own) rather than bound root morphemes (i.e., root morphemes that cannot stand without an affix attached; Bowers & Kirby, 2010); thus, demonstrating the intricate nature of derivational knowledge. Additionally, Tyler and Nagy reported that fourth grade students were able to apply their derivational knowledge. Specifically, students in fourth grade appeared to recognize relationships between root words and derived forms that are unfamiliar and complex, which included knowledge of both bound and free root morphemes as well as understanding appropriate syntactic contexts (Tyler & Nagy, 1989).

In a similar study, McCutchen, Green, and Abbott (2008) asked fourth and sixth grade students to complete a relational morphology task (i.e., the "Comes From" task; Derwing & Baker, 1979), a sentence completion task (i.e., Test of Morphological Structure: Production task; Carlisle, 1995), and an additional task that required students to identify correct root-suffix combinations to evaluate relational, syntactic, and distributional knowledge, respectively. Results revealed that students had greater accuracy on the relational items than on the syntactic and distributional items, with sixth graders out performing fourth graders overall.

In summary, a hierarchy of morphemes exist beginning with the distinction between root morphemes and affixes (Figure 1). In terms of root morphemes, students have an easier time recognizing free morphemes (i.e., can stand alone without affixes) than bound morphemes (i.e., cannot exist without affixes attached; Bowers & Kirby, 2010). For affixes, two types exist, inflectional and derivational, with inflectional morphemes being acquired first (Berko, 1958; Ku & Anderson, 2003; Kuczaj, 1977, 1978; Marcus et al., 1992). Inflectional morphemes are less complicated in that they include suffixes that change the quantity or tense of the root word (Berko, 1958; Kuo & Anderson, 2006) whereas derivational morphemes consist of prefixes and suffixes that alter the root word's definition or part-of-speech (Anglin, 1993; Carlisle, 2003; Ku & Anderson, 2003; Tyler & Nagy, 1989). Knowledge of both types of affixes increases with age (Mahony et al., 2000; Singson et al., 2000; Tyler & Nagy, 1989) with the developmental progression occurring longer for derivational morphemes due to three aspects (e.g., relational, syntactic, and derivational knowledge; Derwing & Baker 1979; Nagy et al., 1993; Singson et al., 2000; Tyler & Nagy 1989). The developmental progression of derivational morphology typically begins with evidence of relational knowledge appearing around four to five years old (Clark & Cohen, 1984; Kuo & Anderson, 2006). When formal reading and writing instruction begins, syntactic knowledge begins to emerge (Berko, 1958) followed by distributional knowledge (Kuo & Anderson, 2006). Overall, the intricate developmental progression of MA needs to be reflected in measures of MA.

Levels of transparency. Like morpheme type, level of transparency also contributes to the complexity of MA and likely should be considered when measuring students' MA. Level of transparency refers to the way in which root words are altered to more complex forms and how clear the relationships between words are. For example, *walk* to *walked* is transparent whereas *run* to *ran* is more ambiguous. Since an affix may alter root morphemes in numerous ways, these alterations are difficult for students to understand, especially when the addition of affixes make phonological and/or orthographic shifts. Phonological shifts preserve the spelling of the root morpheme when an affix is added but they change its sound whereas orthographic shifts change the root morpheme's spelling but preserve its sound (Carlisle, 2003; Carlisle & Stone, 2005). To add further complication, however, some root morphemes are shifted in both ways, making it even more difficult for students to break words down into their components (Nagy et al., 2006). For example, *width* is a word that includes both phonological and orthographical shifts. The root morpheme, *wide*, is altered when the suffix *–th* is added whereas *wider* is phonologically and orthographically transparent (i.e., not requiring a shift other than the addition of the suffix).

Students often struggle with understanding that *wide* and *width* have similar meanings because of these shifts.

Unlike derivational morphemes, inflectional morphemes are easier to define due to the alterations of the root word being limited to changes in quantity or tense, and, typically, tend to be more phonologically and orthographically transparent (Anglin, 1993). Although most inflections tend to be easier to understand and show a predictable developmental path (i.e., plurals, possessives, past tenses, etc.), the development of irregular inflections and derivations are less predictable. The systematic relationships of past tense forms to their present tense forms appear to be difficult for first graders to understand (Kuo & Anderson, 2006). This difficulty is particularly true for irregular past tense forms and morphemes with various phonological pronunciations (Cazden, 1968). For example, the suffix *-ed* has three phonological variations: /t/ (as in the word *helped*), /d/ (as in *stayed*), and /id/ (as in *wanted*). However, the age at which knowledge of these systematic relationships in irregular inflections and morphemes with various phonological pronunciations is acquired is vague and needs to be further investigated, according to Kuo and Anderson (2006). Thus, it is unclear when full mastery of morphology is reached.

Evidence of derivational morpheme acquisition was first seen in early elementary school when students demonstrated competence with transparent derivational words (Clark & Cohen, 1984; Jones, 1983; Wolter et al., 2009). Similarly, first graders' derivational knowledge included transparent words and high frequency derivations (Carlisle, 1995; Carlisle & Nomanbhoy, 1993; Wolter et al., 2009). When Carlisle and Nomanbhoy (1993) asked first grade students to produce a morphologically complex word to complete a context sentence, the students' accuracy was 40.9% when the derived form was phonologically transparent. However, it dropped to 11.2% when the derived form was phonologically opaque (i.e., having various phonological pronunciations; Carlisle & Nomanbhoy, 1993; Ku & Anderson, 2003). Carlisle (1995) stated that, although first grade students may be learning derivational morphemes, their knowledge was limited to transparent relationships with common affixes. This is commensurate with research conducted by Clark and Cohen (1984) and Jones (1983) who found that kindergarten and first grade students' knowledge of derivational morphemes was constrained to transparent relationships. However, results from a study conducted by Wolter et al. (2009) indicated that first grade students were able to provide both transparent and opaque derivatives when given a context sentence. This study suggested that understanding complex relationships may begin as early as first grade, even when morphological relationships or strategies are not explicitly taught.

With mixed results for first grade students, Mahony et al. (2000) investigated grade level increases of derivational knowledge in third through sixth grade students based on levels of transparency. Mahony et al. revealed that younger students accepted false relationships based on level of transparency. For example, young students may likely believe *except* and *acceptance* are related because the two words are phonologically similar. Likewise, older students were able to differentiate between more complex relationships disguised by phonological and orthographic shifts more effectively (e.g., "Halloween/holy, moon/month… break/breakfast, and cup/cupboard"; Mahony et al., 2000, p. 203), and were able to recognize transparent derivations faster than opaque derivations when reading word lists (Carlisle & Stone, 2005). Similarly, McCutchen et al. (2008) investigated whether phonological transparency interacted with grade level. Results revealed that fourth and sixth grade students performed similarly on transparent items, but sixth grade students outperformed fourth graders on phonologically opaque items (McCutchen et al., 2008).

Overall, levels of transparency contribute to the complexity of MA greatly by making relationships between words ambiguous. Children generally try to utilize strategies that require as few changes to the root word as possible (Tyler & Nagy, 1989). These strategies make the transparent words easier to acquire (Ku & Anderson, 2003). Recognition and awareness of morphological relationships, including those that involve shifts, improves with grade level (Mahony et al., 2000) with transparent relationships being acquired and identified first (Carlisle, 1995; Carlisle & Nomanbhoy, 1993; Clark & Cohen, 1984; Jones, 1983; Wolter et al., 2009). Phonological and orthographic shifts make it difficult for students to understand many derivational morphemes whereas inflectional morphemes tend to be more transparent in their relationships (Anglin, 1993; Kuo & Anderson, 2006). Additionally, younger students are more likely to accept a false relationship based on the level of transparency whereas older students outperform younger students on opaque items (Mahony et al., 2000). However, older students recognize relationships between words faster when they are transparent (Carlisle & Stone, 2005; McCutchen et al., 2008).

Word frequency. Children find more frequently used morphemes easier to remember, thus, acquiring more frequent morphemes and words first (Clark & Cohen, 1984; Gordon, 1989). In her review of the literature, Lieven (2010) discussed frequency effects on language acquisition. She stated that word frequency affects how strong the response pattern is when compared to all possible response patterns. In other words, higher frequency words will take less effort to recognize than low frequency words. Frequency effects influencing language first begins with a child's exposure to their parents' speech, to learning stimuli (i.e., books, print based toys, etc.), and to early school instruction such as child care and preschool. The words a child hears most frequently will be acquired first (Lieven, 2010). For example, "no" is a high frequency word that parents, teachers, and siblings say regularly. Due to the frequent exposure, a child will often begin using "no" early on. Infrequently used words and morphemes tend to be the last to be mastered (Ku & Anderson, 2003; Kuo & Anderson, 2006). Derivational morphemes include a wider range of frequency than inflectional morphemes, making the acquisition of derivational morphemes complex.

Numerous studies have examined word frequency effects on MA for people in various age groups. Nagy, Anderson, Schommer, Scott, and Stallman (1989) examined the effects of word family size and word frequency on adults' word recognition. Theoretically, a larger word family size should increase the chances of a student being exposed to a root morpheme, which would then increase the root word frequency. Therefore, family size and root word frequency effects are conceivably linked. To explore these corresponding effects, college students were asked to differentiate root words from nonwords. Nagy et al. found that the number of words in a word family and the frequency of those words affected recognition speed of the root word. The results were affected by the part-of-speech and the age in which the words were likely acquired (Nagy et al., 1989). This supports the developmental progression of morpheme type in which inflectional morphemes, often consisting of verbs, nouns, and simple adjectives, are acquired before most derivational forms (Anglin, 1993; Cazden, 1968; Kuczaj, 1977, 1978; Kuo & Anderson, 2006; Marcus et al., 1992). Further, age of acquisition would also influence the amount of exposure to a given morpheme. As Lieven (2010) discussed, more encounters with a given morpheme reinforces its usage and influences its identification speed.

Researchers have also studied the effect of word frequency on the development of derivational morphemes (Carlisle & Stone, 2005; Ku & Anderson, 2003; Mahony et al., 2000). During their examination of derivational knowledge, Mahony et al. (2000) suspected that item difficulty of the morphological task was influenced by root word frequency, complex word frequency, and affix frequency. Due to this concern, Carlisle and Stone (2005) examined whether familiar morphemes facilitated word reading in lower (e.g., second and third grade) and upper (e.g., fifth and sixth grade) elementary school students. Students were asked to read two sets of words: (a) a morphologically complex word list (one root morpheme and one suffix) and; (b) a single morpheme word list. The words on each list were matched based on the difficulty of spelling, word length, and word frequency. When compared, students read the two-morpheme words more accurately than the one-morpheme words, with the upper elementary school students being more accurate than the lower elementary school students (Carlisle & Stone, 2005). Similarly, when working with second, fourth, and sixth grade students, Ku and Anderson (2003) found that older students were able to interpret infrequent, derived words composed of high frequency morphemes more accurately than younger students. Overall, several studies suggest that the frequency of the derived word may moderate the root word frequency effects on word reading (Carlisle & Katz, 2006; Carlisle & Stone, 2005; Ku & Anderson, 2003).

In terms of word frequency, higher frequency words are more recognizable to students and tend to be acquired first (Clark & Cohen, 1984; Gordon, 1989). In fact, Lieven (2010) stated that frequency effects commence as soon as a child begins hearing words. Root word recognition is affected by word frequency as well as by part-of-speech and age in which the word was likely acquired (Nagy et al., 1989). Evidence has been provided that students may even be able to read two-morpheme words more accurately than one-morpheme words (Carlisle & Stone, 2005). Furthermore, the frequency of the complex word may moderate the root word frequency effects on word reading (Carlisle & Katz, 2006; Carlisle & Stone, 2005; Ku & Anderson, 2003).

In summary, MA is a complex construct including numerous components (e.g., morpheme type, levels of transparency, and word frequency) that are influential in the developmental progression of MA. Each of these components contributes to the complexity of MA. This complexity needs to be adequately reflected within MA measures, or at least acknowledged during the measure construction process. Understanding prior research and theory surrounding a construct, such as the components of MA, was suggested by Wilson (2005) when creating or evaluating measures because it allows researchers to define all aspects of a construct and to, then, outline the measure's item design. He defines item design as ways to gather evidence of the construct in which each item collects one piece of information under a standard condition (Wilson, 2005). Item design considers both the content (e.g., the construct aspect) and distributional (e.g., the classification or difficulty level) components (Wilson, 2005). Therefore, the previous sections have defined and described the existing literature regarding MA and its various components that will allow researchers to evaluate measurement issues in a more complete way. Prior to describing common measurement issues pertinent to measuring students' MA skills, literature discussing the relationship between MA and reading achievement needs to be understood in order to help researchers and educators the construct of MA and its importance further.

The Connection between MA, PA, and Reading Achievement

Educators are highly concerned with their students' reading achievement. Numerous assessment measures exist and are employed by schools in an effort to analyze skills that impact reading achievement, including, but not limited to, PA, phonics[2], and vocabulary. Educators are particularly concerned with teaching and assessing skills related to PA and phonics. Research suggests that future reading achievement is strongly predicted by PA in early elementary years (Adams, 1990; Kirby, Parrila, & Pfeiffer, 2003; Wagner & Torgesen, 1987; Wagner, Torgesen, Laughon, Simmons, & Rashottte, 1993), but students begin to reach the ceiling on PA and phonics measures as they age. Thus, PA becomes less predictive after third grade when other literacy skills, particularly morphological awareness, become more influential in reading achievement (e.g., word reading and reading comprehension; Carlisle & Nomanbhoy, 1993; Jarmulowicz, Hay, Taran, & Ethington, 2008).

Although PA becomes less predictive as the student ages, research indicates that MA is an essential literacy skill that, if screened or assessed, could potentially aid schools in accurately identifying struggling students in the higher grades (Catts, Petscher, Schatschneider, Bridges, & Mendoza, 2008; Speece et al., 2011). As students age and begin reading more difficult texts, it becomes increasingly necessary for them to employ both PA and MA skills as words become more complex and require both skills to accurately decode complex words (Nagy et al., 2006). There is, however, a debate in regard to the influence of MA on reading achievement, particularly related to the impact of MA on reading comprehension. Whereas some studies (Carlisle & Nomanbhoy, 1993; Shankweiler, Crain, Katz, & Fowler, 1995) suggested that MA does not explain unique variance within reading achievement measures (e.g., word reading and reading comprehension) beyond its shared variance with PA, others indicated that MA does

---

[2] Phonics is the knowledge of letter and sound relationships and how to use those relationships to recognize and decode words; thus, it encompasses the ability to take written language and encode it to speech sounds and vice versa (Mesmer & Griffith, 2005).

explain unique variance (Nagy et al., 2006). The following sections describe each side of this debate.

MA does not explain unique variance. Many researchers believe that MA does not explain variance in reading achievement in early grades because PA is more influential. For example, when Carlisle and Nomanbhoy (1993) investigated the influence of PA and MA on the word reading of first grade students, their results indicated that both MA and PA explained a unique amount of word reading variance, but that PA explained more variance within word reading than MA. In fact, MA, as measured by the Production of Word Forms Test (modeled after Carlisle, 1988), accounted for less than 4% of the variance within word reading, which is a typical reading achievement domain for first grade students. Carlisle and Nomanbhoy concluded that PA is more important for word reading in first graders than MA.

To study the contribution of MA in a more advanced reading achievement area, Nagy et al. (2003) examined fourth grade students' MA to understand its influence on reading comprehension. Results indicated that MA did not explain a significant amount of variance in fourth grade students' reading comprehension (Nagy et al., 2003). Similarly, Shankweiler et al. (1995) examined the MA, PA, and reading comprehension of seven to nine-year old students with reading disabilities. Results indicated that these students had particular difficulty with PA even when the effects of age, listening comprehension, and intelligence were controlled. Students with reading disabilities also demonstrated deficits in MA, as measured by the Test of MA: Base and Derived Forms (modeled after Carlisle, 1988). Regression analyses revealed that MA explained unique variance above and beyond its shared variance with PA in reading comprehension. It is important to note that the researchers specified that the difficulties with MA stemmed from the deficits in PA due to the *strong* association ($r = .67$) between MA and PA

(Shankweiler et al., 1995). Although the researchers stated this, a correlational value of .67 does not indicate a strong relationship, but, rather, a moderate relationship between variables according to the correlation use and interpretation guidelines recommended by Mukaka (2012). A moderate relationship indicates that MA and PA, although sharing some variance, do not completely overlap.

Researchers suggested that the morphological aspect of words can be obscured by the phonological complexity of a word (Cazden, 1968; Kuo & Anderson, 2006) and used this aspect as an explanation for why MA does not explain unique variance when PA is considered. In particular, suffixes that increase phonological complexity of a word seem to be especially difficult for students with reading disabilities to decode (Shankweiler et al., 1995). Studies conducted by Shankweiler et al. (1995) and Mahony et al. (2000) demonstrated that the level of phonological transparency between related words impacted students' accuracy in selecting semantically related word pairs. Additionally, Mahony et al. stated that the close phonological relationship led to false recognition of semantic relatedness. These studies demonstrate that the phonological complexity and predictability of a word and word families may be more important than the semantic or morphological aspects of a word in reading achievement (Mahony et al., 2000; Shankweiler et al., 1995) because phonological predictability allows for easier decoding (Fowler, Feldman, Andjelkovic, & Oney, 2003; Nagy et al., 2006).

MA explains unique variance. Although some have suggested that the effects of MA are explained by its relationship with PA (Carlisle & Nomanbhoy, 1993; Nagy et al., 2003; Shankweiler et al., 1995), others have suggested that MA explains a significant, unique amount of variance within decoding and reading achievement (Deacon, Benere, & Pasquarella, 2013; Deacon & Kirby, 2004; Kieffer & Lesaux, 2008; Kirby et al., 2011; Ku & Anderson, 2003;

Mahony, 1994; Tong, Deacon, Kirby, Cain, & Parrila, 2011; Wolter et al., 2009). MA develops with age, with sharp increases in MA beginning around third grade (Anglin, 1993; Carlisle, 2000, 2003; Deacon & Kirby, 2004). As students progress through school, MA becomes an increasingly important skill as the influence of PA on reading comprehension and other measures of reading achievement decreases (Dietrich & Brady, 2001; Stackhouse, 1990; Stanovich & West, 1989). There is evidence that MA explained a unique and gradually increasing amount of variance in students' decoding abilities (Singson et al., 2000), even when PA was controlled for (Mahony et al., 2000). Students with stronger decoding skills and comprehension abilities were typically more sensitive to the morphological relationships of words (Mahony, 1994; Mahony et al., 2000).

Kirby et al. (2011) tracked students' MA, as measured by a word analogy task, from kindergarten through third grade and compared it to their third grade reading achievement. The students' cognitive ability was assessed in their kindergarten year and their PA was assessed in first grade. In third grade, the students were assessed on five measures of reading achievement: decoding ability, word identification skills, word reading fluency, text reading fluency, and comprehension. After controlling for cognitive ability and PA, MA was found to be a significant predictor of third grade students' reading achievement (Kirby et al., 2011). Specifically, MA scores from second grade significantly predicted third grade decoding ability, word identification skills, text reading fluency, and comprehension. However, MA scores from first grade did not significantly predict any of the third grade reading outcomes. In third grade, MA scores explained a significant amount of variance in all five measures of reading achievement with larger effects than those observed in second grade. These results suggest that MA became an increasingly important predictor in the students' comprehension abilities as they progressed through the elementary grades (Kirby et al., 2011).

Although Kirby at el. (2011) did not observe effects in first grade, Wolter et al. (2009) revealed that first grade students' MA, assessed by a sentence completion task (based on Carlisle, 1995), explained unique variance in their ability to read word lists beyond what was predicted by PA. The contrasting results between Kirby et al. (2011) and Wolter et al. (2009) may be due to the measures used to assess MA. It could be that a word analogy task may be too difficult for first grade students, whereas a sentence completion task may be more appropriate or more sensitive to first graders' MA. Although there were contradictory findings for first grade students, these results still suggested that MA greatly impacts reading achievement in early elementary school and that this impact increased with age.

In a longitudinal study assessing students' MA and reading achievement across time, Deacon and Kirby (2004) compared second grade students' MA, using a sentence analogy task, and PA to two reading achievement measures assessed in second through fifth grade: word reading and comprehension. Results revealed that MA did not explain a significant amount of variance within word reading, but it did approach significance ($p = .08$) in fourth grade (Deacon & Kirby, 2004), thus supporting Carlisle and Nomanbhoy (1993). However, with reading comprehension being the ultimate goal of reading instruction (van den Broek & Espin, 2012), Deacon and Kirby's results regarding reading comprehension are more intriguing. After controlling for PA, intelligence, and prior academic achievement, the researchers found that MA explained significant variance in fourth and fifth grade students' reading comprehension. However, it provided no such prediction for the reading comprehension of third grade students (Deacon & Kirby, 2004). As previously stated, the amount of variance MA explained within comprehension increased over time, but the amount of variance explained by PA did not. Remarkably, results suggested that MA has a greater influence on comprehension than sight word reading (Carlisle & Nomanbhoy, 1993; Deacon & Kirby, 2004), potentially suggesting that students' MA has a greater role in their ability to comprehend text than in decoding individual words. Deacon and Kirby stated that MA aided readers in building text meaning more than in aiding word reading.

To investigate whether MA may explain reading achievement differences in fifth grade students, Tong et al. (2011) calculated predicted comprehension scores using students' age, nonverbal cognitive abilities, and word reading accuracy and speed. These scores were compared to students' actual performance on a comprehension measure in order to create three groups of students: (a) unexpected poor comprehenders (students who scored below the 80% confidence interval of their predicted comprehension score), (b) expected average comprehenders (students who scored within the 80% confidence interval of their predicted comprehension score), and (c) unexpected good comprehenders (student who scored above the 80% confidence interval of their predicted comprehension score; Tong et al., 2011). Students were then assessed on several measures including, but not limited to, MA, PA, naming speed, and orthographic processing (see Table 1 for definition). The three groups varied significantly in terms of their MA, but not in their performance on the other measures administered. Specifically, unexpected poor comprehenders had deficits in MA but not in the other reading skills assessed or in intelligence. Conversely, unexpected good comprehenders had stronger MA than both average and unexpected poor comprehenders (Tong et al., 2011).

An extension of research examining the relationship between MA and reading achievement, Mahony (1994) examined high school and college students' MA and three indicators of their reading achievement: verbal score on the SAT, the Nelson Reading Test, and reading class placement. High school and undergraduate students completed the Morphemes Sensitivity Test which measures MA in four subtests by having students (a) examine syntactic relationships of real words within given context sentences (a subtest modeled after Carlisle, 1988); (b) examine syntactic relationships of nonwords with given context sentences; (c) indicate whether semantic relationships exist between two words; and (d) pronounce phonological changes in morphologically complex words from their root words. Results indicated that MA was significantly related to undergraduates' verbal SAT performance and to high school students' Nelson Reading Test performance (Mahony, 1994). In addition, all four subtests of the MA measure aided in distinguishing proficient high school readers from non-proficient high school readers. Specifically, less proficient readers had difficulty generalizing their knowledge of suffixes to nonwords (Mahony, 1994).

In addition to evidence that indicated that MA is a unique predictor of reading achievement, research also indicated that MA may be a useful tool for improving students' acquisition of vocabulary and defining words within a passage (Ku & Anderson, 2003; Sandra, 1994). In fact, Ku and Anderson (2003) stated that MA allowed students to decipher, construct, and remember unfamiliar word meanings by aiding them in dissecting words into meaningful parts, defining each meaningful part, and creating one unifying definition by merging each meaningful part definition. In their investigation of Ku and Anderson's theory, Wysocki and Jenkins (1987) examined whether students in the fourth, sixth, and eighth grades could produce word definitions by analyzing morphological relationships. Students were asked to define unknown words by using morphological and contextual clues. Results supported Ku and Anderson's (2003) theory; the students, especially older students, better understood words that were morphologically related to previously encountered words. Thus, results supported Ku and Anderson's theory suggesting the students were able to use a morphological strategy to decode and understand word meanings and that this skill increased as students aged (Wysocki & Jenkins, 1987).

Generally, research seems to support that MA influences reading achievement (e.g., Deacon & Kirby, 2004; Kirby et al., 2011; Wolter et al., 2009) with only a few studies not supporting MA explaining variance above and beyond PA (e.g., Carlisle & Nomanbhoy, 1993; Nagy et al., 2003; Shankweiler et al., 1995). Differences in findings between studies that support unique variance explained by MA and those that do not may be due to a number of factors. These factors include the type of item, item presentation, response requirement, and skills being assessed. The type of item (e.g., multiple choice, yes/no, fill in the blank, open-ended, etc.) is usually dictated by the purpose of the measure. For example, if the purpose is to define a given word, an open-ended type of item is typically constructed, whereas assessing whether a word belongs to one category or another would require a yes/no item type. However, the type of item may also be chosen by a researcher to either reduce or increase the item's difficulty, with multiple choice and yes/no style items being theoretically easier than open-ended or fill in blank (Wilson, 2005).

In addition to the type of item used by a researcher, an administrator can present items in a variety of ways. Many measures utilize oral presentation of items whereas others use visual or a combination of oral and visual presentation. Similarly, the task may require students to respond through oral or written responses. Last, the skills being measured could vary by morpheme type (i.e., inflected only, derived only, suffixes only, prefixes only, all types, etc.), by the level of transparency (i.e., no shift, phonological shift, orthographic shift, both shifts, etc.), or by the type of MA knowledge (e.g., relational, distributional, and syntactic) which may constrain the item type and presentation.

Studies examining the importance of MA have employed a variety for assessing MA, with these measures varying in the specific MA skills assessed, the manner in which stimuli/questions were presented, and the types of responses required by the respondent. Such differences likely lead to variation in which components of MA were assessed and, thus, lead to differences in the extent to which the resulting data explained variance in students' reading achievement. Further, it may be that some of the MA measures administered were inappropriate for the age group researchers assessed, which potentially led to null results or results indicating a small amount of unique variance explained by the measure in students' reading achievement. The lack of consistency in measures used and the failure of researchers to ensure that they were employing measures appropriate for the population of individuals whose skills they were assessing has likely contributed to inconsistency in results across studies and, thus, a lack of clarity in the extent to which MA contributes to reading achievement.

What Educators Need to Consider About Measurement and Why Should They Care

Having a well-constructed measure helps educators (e.g., teachers, administrations, and other professionals) understand the level of a student's skills within a given construct and provides information about a student's progress and potential trajectory. The *Standards for Educational and Psychological Testing* (i.e., the *Test Standards*; AERA, APA, & NCME, 2014) were generated to ensure that appropriate and well-constructed measures were created and used in psychological and educational settings. The *Test Standards* guide ethical practices, advise test quality and practices, reflect measurement advances, and, most importantly, provide a definitive set of criteria for all educational and psychological test developers, publishers, and users

("Standards for Educational," 2014). The *Test Standards* are used by the courts as a standard to compare and evaluate tests. Thus, these standards are considered the "gold standard" in test development and evaluation. Within this guide, the *Test Standards* highlight three foundational areas: validity, reliability, and fairness in testing. Each of these contribute significant to the development or evaluation of a measure and deserve to be explored separately and in depth.

Validity. Validity is defined as the extent to which test scores and interpretations of a particular test are supported by evidence and theory regarding the test's construct (AERA, APA, & NCME, 2014). Messick (1995) expanded this definition by stating that theory and empirical evidence need to "support the adequacy and appropriateness of interpretations and actions on the basis of test scores or other modes of assessment" (p. 741). In essence, validity is a summary of evidential support for score interpretations and consequences of those interpretations. Test developers demonstrate validity by integrating several types of evidence, including analyses of the measure content, response processes, internal structure, relation to other constructs, and consequences (AERA, APA, & NCME, 2014).

Arguably, the most difficult aspect of measurement to ensure is demonstrating that a measure is capturing a student's level of construct knowledge. Many test developers run the risk of construct underrepresentation (i.e. when an assessment fails to include important aspects of a construct) as well as adding construct-irrelevant variance (i.e., an assessment being too broad in terms of items having other extraneous aspects making a task too easy or too difficult; Messick, 1995). Although collecting evidence to support that a measure is representing a construct fully is time consuming and difficult to accomplish, it is vitally important in establishing whether a measure is usable. In fact, validity is inextricably tied to the other two foundational areas, reliability and fairness in testing (AERA, APA, & NCME, 2014; Messick, 1995).

Reliability. This second foundational area refers to the precision of a measure and allows generalization of score consistency and replication consideration (Thorndike & Hagen, 1961). The *Test Standards* state that the term has been used in two ways: to refer to (a) the correlation between the scores of two test versions and (b) the consistency of score replications regardless of testing procedure (AERA, APA, & NCME, 2014). But in classical test theory, there are three types of reliability coefficients: alternate-form, test-retest, and internal consistency (AERA, APA, & NCME, 2014). The alternate-form procedure requires a comparison of two versions of the same measure. The test-retest procedure compares two administrations of the same measure. The internal consistency procedure summarizes the relationships or interactions among various subsets of items from one administration of a measure (AERA, APA, & NCME, 2014). The most popular indicator of reliability is Cronbach's alpha, or coefficient alpha (α), and is commonly reported in research when supporting the internal consistency of a given measure. When interpreting coefficient alpha, Cortina (1993) stated that acceptable reliability is an alpha score greater or equal to 0.80 with a higher alpha providing stronger evidence.

Fairness. Fairness in testing bridges the applications and consequences of a measure with the foundational areas of validity and reliability (AERA, APA, & NCME, 2014). Messick (1995) referred to the consequences of a measure as consequential validity. Since Messick (1995), great emphasis has been placed on assessing the potential consequences of a measure as well as combatting its negative aspects and amplifying the positive aspects, therefore culminating in a new foundational area within the *Test Standards* (AERA, APA, & NCME, 2014). Test developers are responsible for minimizing or removing issues that may potentially restrict a student's ability to exhibit his or her knowledge ("Standards for Educational," 2014). Although test developers are responsible for minimizing potential adverse consequences, test users are responsible for knowing the suggested uses of a measure and how to interpret resulting scores properly (AERA, APA, & NCME, 2014). Users are also responsible for properly utilizing the information gained from a measure (AERA, APA, & NCME, 2014). Since educators use test scores to make educational decisions, they need to ensure that the assessments they select are intended to measure the constructs they desire to measure and that they appropriately interpret and utilize the resulting data.

To summarize, the precision of a measure should rely on the theory and support in which the measure is based (AERA, APA, & NCME, 2014). A reliable but invalid measure is essentially pointless. A measure needs to assess the same thing each time it is used; otherwise, it is not useful to make comparisons across time which is important for screening use. Measures need to be based on solid support or theory as well as provide evidence of score reproducibility and replication (AERA, APA, & NCME, 2014). These characteristics allow a measure to be used accurately and fairly.

To date, sufficient analysis of the reliability, validity, or fairness of MA to the extent recommended by the *Test Standards* does not exist (AERA, APA, & NCME, 2014). In order to serve students' needs, MA measures need to be investigated. The purpose of this paper is to begin the investigation of a commonly employed MA measure. Thus far, evidence has been presented to document the definition and components of MA and MA's involvement in reading achievement. However, the measurement limitations within MA must be explored as well as describing a commonly used measure of MA.

The Measurement of MA

The literature demonstrates that researchers do not believe any measure of MA is adequate for assessing MA completely. Measures have often been adapted or modified from one study to the next. Researchers must consider morpheme type, levels of transparency, and word frequency when choosing measures, and they often make adaptations or modification to existing measures to include one or more of these aspects.

The goal of a research study drives a researcher's choice in the measure(s) used in their study. Apel (2014) stated that some MA measures predict early elementary students' reading achievement better than others and other measures discriminate between grades (Apel et al., 2013). Similarly, some MA measures relate to other literacy skills more than other measures due to format and item design (Apel, 2014). For example, a sentence completion task may relate more to comprehension than a task asking students to state whether two words are related to one another. Even choosing an alternate definition of morphological awareness may impact the choice in MA task format (see Apel, 2014 for examples). Furthermore, it is unclear whether the various existing measures of MA are assessing the same underlying construct or if the measures are analyzing different aspects of MA (Apel, 2014), with the latter being more likely. With researchers often using several measures of MA within their studies (e.g., Apel et al., 2013; Ku & Anderson, 2003; Nagy et al., 2006; Wolter et al., 2009), it appears that the MA research community generally agrees that MA has several dimensions that can be measured. This multi-dimensionality of the current nature of MA assessment suggests that one unified measure that adequately assesses all aspects of MA does not currently exist.

Various MA measures. With no single measure of MA across previous studies in which all aspects of MA are assessed, Apel et al. (2013) conducted a study comparing kindergarten, first grade, and second grade students' MA using four different tasks. They hoped to understand whether these measures functioned similarly. These tasks included: (a) the relative task, a production, cloze response task (modeled after Carlisle, 2000); (b) the rehit task requiring students to combine two morphemes, define the new term, and judge the terms used in a context sentence; (c) an affix identification task requiring students to circle prefixes and suffixes in a word list; and (d) a spelling task requiring students to spell morphologically complex words. Results indicated that the relative task had the greatest utility of all four tasks as student performance significantly differed by grade and was associated with reading achievement (Apel et al., 2013). The researchers suggested that the rehit task was useful for analyzing pre-reading skills since the task differentiated kindergarteners from first and second graders. The affix identification task was able to discern the level of emerging literacy skills in students and may help predict spelling ability in first and second graders. Apel et al. (2013) deemed the spelling task too difficult for students in kindergarten through second grade, but found that it uniquely predicted word reading in second graders.

Relationship of MA measures. Apel et al. (2013) demonstrated that various MA measures differed on what they assessed and suggested that examiners use different tasks for different purposes. However, if measures are assessing the same construct, then those measures should be highly related to one another. This validity has not been consistently demonstrated in the literature, if it was analyzed at all.

In their study analyzing MA and reading comprehension, Nagy et al. (2003) asked second and fourth grade students to complete two measures assessing MA: Morphological Relatedness Test (MRT; Mahony, 1994) and the Suffix Choice task (Berninger, Abbott, Billingsley, & Nagy, 2001; Berninger & Nagy, 1999). The MRT, also known as the "Comes From" task, consists of 12 items that require students to state or record whether two provided words are related to one another. The Suffix Choice task consists of 14 items and requires students to choose the best form of a word to complete a context sentence (modeled after Carlisle, 1988) from 4 options.

Once the second and fourth graders had completed both measures, researchers performed correlational analyses for each grade. For second grade students, the correlation was negligible ($r = .24$, $p <.05$). The correlation between the two measures was higher for fourth grade students, but still demonstrated a low relationship between the two measures ($r = .32$, $p <.01$; Nagy et al., 2003).

Deacon et al. (2013) also investigated the relationship between two morphological measures, Sentence Analogy and Word Analogy, completed by second and third grade students. These two tasks are similar in their format (i.e., following the 1:2::3:4 form and focusing on verb tenses), but differ by the number of items. However, the Sentence Analogy includes context sentences whereas Word Analogy does not. The Sentence Analogy task consists of 3 practice items and 8 test items (based on Nunes, Bryant, & Bindman, 1997) and the Word Analogy task consists of 3 practice items and 21 test items. Interestingly, very little was seemingly different between the two tasks. For second grade students, the correlation demonstrated a moderate relationship between the two measures ($r = .53$, $p <.01$). This same result was seen with the third grade students ($r = .56$, $p <.01$; Deacon et al., 2013). With so little difference between these tasks, it might be expected that the relationship between the two would be strong. However, the results demonstrated that the two measures have only a moderate relationship. This may suggest that even slight modifications and adaptations of any measure may change the construct it is measuring.

To investigate whether slight alterations to a measure affects student performance, Singson et al. (2000) examined the MA of students from 3rd through 6th grade on whether item presentation (Written vs. Oral+Written) would affect student performance on the Derivational Suffix Test (adopted from Mahony, 1994 which was modeled after Carlisle, 1988). The

Derivational Suffix Test is a multiple choice sentence completion task assessing MA. The task involves 40 items with 20 presented in Written-only form and the other items presented in Oral+Written form. Students were provided a context sentence and asked to choose the correct word form from four choices in order to complete the sentence. Results revealed that there was a slight disadvantage to the Written-only presentation (Singson et al., 2000). That is, students in the Oral+Written presentation group slightly outperformed the students in the Written-only presentation group. However, the advantage of items presented orally and in writing decreased as grade level increased (Singson et al., 2000).

In their second experiment, Singson et al. (2000) directed another set of students from 3rd through 6th grade to complete the Derivational Suffix Test with 20 items presented in Written-only form (as done previously) and 20 items presented in Oral-only form. Results were similar to those in their first experiment with the two presentations altering performance. However, the Written-only presentation had a slight advantage over the Oral-only presentation which produced the reverse of the previous results (Singson et al., 2000).

Similarly, Mahony et al. (2000) investigated whether item presentation affected another morphological awareness measure, the Morphological Relatedness Test (also known as the "Comes From" task; Mahony, 1994). This study was attempting to replicate the experiments from Singson et al. (2000). The researchers asked students in grades 3 through 6 to complete the 40-item task. The results mirrored those from experiment 1 in Singson et al. with the Oral+Written presentation having a slight advantage with that advantage decreasing with age. However, students performed similarly on the Oral-only and Written-only presentations (Mahony et al., 2000). These results contradicted those in Singson et al. (2000) where Written-only provided an advantage over the Oral-only presentation.

In summary, results from Singson et al. (2000) and Mahony et al. (2000) generally support that slight adaptations to a morphological measure, such as presenting the measure orally versus visually, will alter the performance of the students. But, what would adaptations to morpheme type, word frequency, or levels of transparency do to the results of a study? These types of changes are frequently made to measures, even accepted measures such as Carlisle (1984, 1988, 2000). As suggested by Apel et al. (2013), Mahony et al. (2000), and Singson et al. (2000), different measures and adaptations of the same measure may alter the results of a study and the performance of the students assessed.

Test of Morphological Structure (Carlisle, 1984, 1988, 2000)

One generally accepted measure is Carlisle's (1984, 1988) Test of Morphological Structure as it is used in numerous studies (Berninger, Abbott, Nagy, & Carlisle, 2010 ; Fowler & Liberman, 1995; Kieffer & Lesaux, 2008; McCutchen et al., 2008; Singson et al., 2000; Tong et al., 2011; Wolter et al., 2009). The Test of Morphological Structure, which is a sentence completion task, was created in 1984 by Carlisle to measure derivational morphology. It was published in 1988. The original test contained two subtests, Derived Forms and Base Forms, with each subtest consisting of 40 items with all four transparency levels represented. Item order was determined by creating 10 sets of 4 items with each level of transparency represented and randomly ordered within each set. The 10 sets were then randomized within the original measure. The original version of the measure continued to appear in her research until 1992. Following 1992, Carlisle adapted the measure as part of Carlisle and Nomanbhoy (1993). Specifically, the Derived Form subtest was adapted through the addition of inflectional morphemes and the removal of items with both orthographic and phonological shifts. Carlisle and Nomanbhoy also shortened the subtest to 15 test items and 4 practice items and introduced a new name for the subtest, Production.

Carlisle (2000) employed an altered version of her measure again by shortening both original subtests (e.g., Derived Forms and Base Forms) to 28 items each. She also changed the names of the subtests to Derivation (e.g., Production, Derived Forms) and Decomposition (e.g., Base Forms). This version of the measure appears to be Carlisle's choice in her more recent studies (i.e., Berninger et al., 2010; Carlisle & Fleming, 2003; see Appendix B for summary of Carlisle's measurement modifications and adaptations) as well as being the most cited in other studies investigating MA. Of the 39 studies investigating MA cited in this paper, 48.7% of those studies used one of the versions of Carlisle's (1988, 2000) measure (e.g., Apel et al., 2013; Fracasso, Bangs, & Binder, 2014; Tong et al., 2011). However, 94.7% of those studies adapted or modified the measure or used an already adapted version for their own purposes.

In summary, having measures that are constantly adapted and modified suggests that even established researchers of MA (e.g., Carlisle, Nagy, Deacon, etc.) believed that MA measures are not assessing MA adequately or completely. Measures that are modified or adapted are often unsupported psychometrically. These changes also make it difficult to understand how MA influences reading achievement, including word reading and reading comprehension. In fact, using different or adapted measures does not allow studies to be compared, particularly between different ages. Thus, measures assessing MA need to be analyzed psychometrically with particular focus on their validity and reliability. Hopefully, a "gold standard" measure assessing MA may be achieved eventually. However, it will take several studies to create such a measure. The purpose of this study is to take one step in the process of creating a "gold standard" measure by validating the use of Carlisle's (2000) morphological measure, Test of Morphological Structure.

CHAPTER 3

METHOD AND RESULTS

Purpose

Current research on morphological awareness (MA) yielded mixed results. Some research suggested that MA fails to explains unique variance within reading achievement beyond phonemic awareness (Carlisle & Nomanbhoy, 1993; Nagy et al., 2003; Shankweiler et al., 1995), whereas other research suggested that MA does explain unique variance within reading achievement (Ku & Anderson, 2003; Nagy et al., 2006; Nunes, Bryant, & Olsson, 2003; Singson et al., 2000; Tong et al., 2011). These contradictory findings may be due to the characteristics of the populations sampled, but the differences may also be due to differences in the measures of MA used across studies and the particular components of MA awareness examined within particular studies. Even within the studies suggesting that MA explains unique variance in reading achievement (Deacon & Kirby, 2004; Kirby et al., 2011; Mahony, 1994; Nagy et al., 2006), there is little consistency in the measurement instrument employed. Furthermore, many of the studies which used the same measure actually employed different variations of the same measures. Within the MA literature, researchers have tended to modify measures (e.g., Carlisle & Nomanbhoy, 1993; Kieffer & Lesaux, 2008; McCutchen et al., 2008; Tong et al., 2011) in an effort to either improve upon the measure and/or assess different components of MA. Given the lack of consistency in the instruments employed within the MA literature, it is difficult to either generalize findings or come to any conclusions as to why findings might differ between studies. This line of research is also limited by the fact that researchers have failed to validate the measures they used when attempting to understand the contributions of MA to reading achievement.

The first purpose of the current study is to examine the psychometric quality of a frequently employed MA measure, the Test of Morphological Structure (TMS; Carlisle, 2000). Specifically, this study will, first, examine the impact of three item orders (e.g., original, theoretical, and data-based) on the students' scores. These item orders will not be administered to students in the present study, but rather items were administered in a randomized order to each participant. Resulting data was used to establish item difficulty and determine if previous research utilized an appropriate discontinue rule (Fracasso et al., 2014; Herman, Cote, Reilly, & Binder, 2013; Tighe & Binder, 2013). Analyses were conducted to determine whether scores would differ for students based upon three potential item-orders. Future research will need to be conducted to examine the reliability and validity of data resulting from the items administered in the order determined to best capture student performance.

Second, the current study examined what impact a discontinue rule would have on resulting outcome data when applied to each of the three item orders. Researchers have frequently shortened the TMS (Carlisle, 2000) by deleting items (Apel et al., 2013; Apel & Lawrence, 2011; Apel & Thomas-Tate, 2009; Carlisle & Nomanbhoy, 1993; Deacon, Francis, & Tong, 2015; Kieffer & Lesaux, 2008; McCutchen et al., 2008; Wolter et al., 2009) and applying discontinue rules (Fracasso et al., 2014; Herman et al., 2013; Tighe & Binder, 2013). The purpose of such discontinue rules is to reduce the time and frustration associated with completing the tests, while still allowing resulting data to estimate the construct(s) being assessed reliably and accurately. Changes made to the TMS were typically done without evaluating the potential impact of the deletion of items and without analyses to determine an appropriate discontinue rule that would allow for resulting data to accurately represent students' MA skills. The item order of the TMS (Carlisle, 2000) was not developed from a theoretical model or prior research examining item difficulty. Thus, developing a new or unique discontinue rule is potentially problematic as in order to apply a discontinue rule it is essential that the items of a measure be ordered by difficulty (e.g., easiest to most difficult). Prior research does, however, suggest that the TMS (Carlisle, 2000) remains a fairly consistent predictor and contributor to reading achievement despite being modified or adapted (Ku & Anderson, 2003; Nagy et al., 2006; Nunes et al., 2003; Singson et al., 2000; Tong et al., 2011). Considering researchers' perception of the TMS (Carlisle, 2000) being too time consuming to administer and the need for educational screeners to be shorter in duration (Hosp et al., 2011), research is needed to determine an appropriate discontinue rule. This study will investigate the use of a discontinue rule on students' scores across three item-ordering procedures: original order (i.e., items ordered in the way the test was designed), theoretical order (i.e., items ordered based upon the developmental trajectory of MA), and data-based (i.e., items ordered based upon item analysis, specifically Rasch analysis). Analyses will also be conducted to examine the validity of scores obtained from each of the item-ordering procedures. It is hypothesized that the data-based and theoretical item orders will have stronger psychometric properties (e.g., reliability and validity of scores for intended uses) than the original item order. Specifically, the data-based ordering was the strongest, most accurate measure of the three item-orderings because Rasch analyses provide information on how items range in difficulty within a measure.

The third purpose of this study is to replicate prior research (Deacon & Kirby, 2004; Kirby et al., 2011; Mahony, 1994; Nagy et al., 2006) investigating the variance explained by MA in reading comprehension, an aspect of reading achievement. As discussed in Chapter Two, MA aids individuals in breaking unfamiliar words down into recognizable components to determine meaning. If students are struggling with reading unfamiliar words, then they are likely to be struggling with the overall meaning of the text; therefore, making comprehension difficult. Understanding the potential impact of MA on reading comprehension beyond the influence of phonological awareness (PA) could aid educators in developing more effective lesson plans and interventions for struggling readers. However, few studies have examined the connection between MA and comprehension using psychometrically-validated measures of MA or by examining the differences in variance explained when using similar but slightly modified measures. Therefore, this study will investigate the relationships between MA and reading comprehension using the original item-ordering procedure and the item-ordering procedures demonstrating the best validity. The hypotheses to be examined are (a) the three item-ordering procedures of the TMS (Carlisle, 2000) will have a significant relationship with PA and reading comprehension; and (b) the scores produced by the item-orderings will explain variance within comprehension above the effects of PA with the data-based procedure explaining the most unique variance of the three item-ordering procedures.

Methods

Participants

The present study included 96, fourth grade participants enrolled in a Southeastern elementary school and comprised of 42 females (43.75%) and 54 males (56.25%). The participants ranged in age from 8.03 years to 11.67 years old (M = 10.31, SD = 0.43) and are all native English speakers. The majority of participants reported that they were White (83.33%). Other racial and ethnic backgrounds reported by participants included Black (5.21%), Asian (4.17%), Biracial (3.13%), Multi-racial (2.08%), Hispanic (1.04%), and Other (1.04%). Of the 96 participants, 30 participants are enrolled in programs others than general education. These participants are enrolled in gifted programming (16.67%), special education services (15.63%), and small group interventions (2.08%). Six participants (6.25%) are enrolled in more than one of these programs.

Materials

Morphological tasks. The TMS consists of 56 items equally divided into two tasks: Derivation and Decomposition (Appendices C & D; Carlisle, 2000). The Derivation task assesses participants' knowledge of derivational morphemes by requiring them to build complex words. Similarly, the Decomposition task assesses participants' knowledge of derivational morphemes by requiring them to identify roots words within a complex word. Both tasks provide participants with a target word followed by a context sentence. Participants are required to alter the target word to fit the sentence by either adding or removing a suffix. An example item in the Derivation task is: "Farm. My uncle is a _____." In this example, the correct answer is "farmer." An example item in the Decomposition task includes: "Driver. Children are too young to _____." In this second example, the correct response is "drive." No reliability or validity information for the TMS was provided by Carlisle (1984, 1988, 2000). Tong et al. (2011) reported a Cronbach's alpha of 0.60 for their 40-item, Derivation task completed by 5th grade students. When administering a 25-item, Decomposition task, Kieffer and Lesaux (2008) reported a Cronbach's alpha of 0.78 for 4th grade students and 0.82 for 5th grade students.

Phonological decoding task. To evaluate the participants' phonological decoding ability, the Word Attack Subtest from the Woodcock Johnson Tests of Achievement – Fourth Edition (WJ-IV ACH) was administered to participants according to standardized administration procedures (Schrank, Mather, & McGrew, 2014). This subtest requires that participants read non-words (e.g., tayed, nat, etc.) with resulting data being the number of non-words correctly read. The administrator will test by complete pages until the discontinue rule is met. Administration is discontinued once a participant has mispronounced the six highest-numbered non-words. According to the WJ-IV ACH manual, reliability coefficients ranged from 0.89 to 0.93 on this task for 8 to 10 year old children. To provide validity information, Schrank, Mather, and McGrew (2014) described studies comparing the components of the WJ-IV ACH to other achievement measures. Although, Schrank et al. (2014) did not provide validity information for the Word Attach subtest specifically, the authors described studies that compared reading composites. The Basic Reading Skills composite of the WJ-IV ACH, in which both Word Attack and sight word reading are included, is significantly related to the Reading composite of the Kaufman Test of Educational Achievement, Second Edition (KTEA-II; $r = .93$), and the Basic Reading composite of the Wechsler Individual Achievement Test, Third Edition (WIAT-III; $r = .94$). This demonstrates a strong concurrent validity between reading composites (Schrank, Mather, & McGrew, 2014).

Passage Comprehension. The Passage Comprehension subtest from the WJ-IV ACH was used to evaluate participants' reading comprehension skills. The subtest consists of 85 sentences and passages (Schrank, Mather, & McGrew, 2014). Participants were required to read the passages silently and then fill in the missing word. The administrator will test participants by complete pages and discontinue testing when the six highest-numbered items are incorrect as outlined by standardized administration procedures (Schrank, Mather, & McGrew, 2014). According to the WJ-IV ACH, reliability coefficients ranged from 0.92 to 0.94 on this task for 8 to 10 year olds. To demonstrate concurrent validity, Passage Comprehension was compared to the Reading composite from the KTEA-II (*r* = .85) and the Comprehension and Fluency composite from the WIAT-III (*r* = .81; Schrank, Mather, & McGrew, 2014).

Administration Procedures

Consent forms were sent home to parents by the school. The consent form will ask parents to provide demographic information including date of birth and racial background. After obtaining parental consent, trained researchers will administer each of the measures to the participants in a quiet location. Each testing session will take approximately 30 to 45 minutes to complete. The order in which the 4 tasks were administered was counterbalanced. The items within the MA tasks were randomized for each participant. Trained researchers will record the time it takes each participant to complete each MA task (i.e., two completion times was recorded for the Decomposition and Derivation tasks). After the session is completed, the participants will receive a small reward (e.g., a toy, pencil, sticker, etc.) for their time, concentration, and cooperation.

Scoring Procedures

Morphological tasks. Each participant's data was organized into three item-orderings: 1) the original item order; 2) the theoretical item order; and 3) a data-based item order. The original item order was the items ordered the way the measure was designed (see Appendices C & D). The theoretical item order was determined by the author of this document based upon what research has determined to be the developmental progression of MA. Specifically, the theoretical item order was determined by first considering the levels of transparency (e.g., no shift, orthographic shift, phonological shift, and both orthographic and phonological shifts), and, then, the items were ordered by frequency of the root word (See Appendices E & F for item order and

Appendix G for the building blocks supporting the theoretical order). Last, the data-based item order was based on Rasch analyses.

For each of the item-orderings, a discontinue rule was applied (see Appendix H for an example). The original MA tasks (Carlisle, 2000) did not utilize a discontinue rule. However, many researchers have (e.g., Apel & Thomas-Tate , 2009; Berninger et al., 2010; Carlisle & Fleming, 2003; Deacon, Tong, & Francis, 2015; Spencer et al., 2015; To, Tighe, & Binder, 2014) adapted the measure in order to reduce administration time by either reducing the number of items or instituting a discontinue rule; thus, suggesting that the tasks require too much time to administer. For the current study, a discontinue rule of six consecutive incorrect responses was applied after data collection has been completed to each of the item-ordering procedures. This discontinue rule was selected due to it being the most common rule applied in previous research (e.g., Fracasso et al., 2014; Herman et al., 2013; Tighe & Binder, 2013).

Finally, the participant's score was determined by the number of correct responses for each of the item-orderings after the discontinue rule has been applied. This resulted in three scores for each participant.

Reading tasks. Raw scores were used for both subtests of the WJ-IV ACH (i.e., Word Attack and Passage Comprehension) in all analyses.

Completion time. Each participant's completion time for each MA task was recorded. For each of the two MA task, an estimate time per item was calculated by dividing the total time to complete the time by 28 (the number of items within the task).

Procedural Integrity

Procedural integrity was examined on all measures for 33.33% (*n* = 32) of the participants. A checklist (Appendix I) was created to ensure that task procedures were followed and all tasks were administered correctly. An evaluator examined the audiotaped sessions using the checklist. Procedural integrity was calculated by the number of procedural steps met in the administration of a task being divided by the total number of procedural steps required by the task. This number was multiplied by 100. Procedural integrity was above 99% for all tasks (Word Attack = 100%; Passage Comprehension = 100%; Decomposition = 99.11%; Derivation = 100%).

Results

Rasch Analyses

The Rasch model is equivalent to a one-parameter (e.g., item difficulty), latent trait item response model used in Item Response Theory (IRT; Bond & Fox, 2015; Crocker & Algina, 2008). IRT postulates that a person's expected item response can be represented through a probabilistic, mathematical function (Bond & Fox, 2015). Although the Rasch model is similar to models used in IRT, it differs by focusing on fundamental measurement and by encompassing the idea of unidimensionality, which is that a measure can only examine one construct at a time (Bond & Fox, 2015). Rasch analyses provide fit statistics that describe how an item is fitting a particular construct (i.e., measuring construct validity) and places items in order based on the level of difficulty (Bond & Fox, 2015). This also allows researchers to see where there may be gaps within a line of difficulty or item hierarchy. Overall, Rasch analyses allow for the examination of the internal structure of measures by inspecting (a) unidimensionality, (b) item difficulty, (c) item and person distribution, and (d) model-data fit (Engelhard, 2013).

Within Rasch analysis, variance explained by the Rasch statistic is used to examine unidimensionality. To consider a measure as unidimensional, a minimum of 20% variance is recommended (Linacre, 2006; Reckase, 1979). The amount of variance explained by the Rasch statistic indicates the strength of a measure. Linacre (2006) states that measures with ≥30% explained variance are considered to be moderate measures, whereas ≥ 40% explained variance is considered to be a strong measure. Within the Test of Morphological Structure (TMS) tasks, 43.96% and 42.02% of the variance was explained by the Decomposition and Derivation tasks, respectively. These statistics suggest that both TMS tasks are strong measures of morphological awareness (MA).

Within Rasch analyses, reliability within both TMS tasks was examined in two ways, (a) person separation reliability and (b) item separation reliability. Person separation reliability indicates whether students can be differentiated from one another, whereas item separation reliability indicates whether items can be distinguished from one another. For the Decomposition task, person separation reliability was .71 and item separation reliability equaled .92. The reliability coefficients for the Derivation task were .83 for person separation and .95 for item separation. All reliability coefficients demonstrated strong positive relationships.

The Wright Map is produced when Rasch analyses are run. This map depicts the item and person distribution visually by presenting the location of each item and person using a common ruler, the logit scale (see Figures 3 and 4). The logit scale has a mean of 0 logits. Ideally, the mean logit for the measure's items and persons should be as near to 0 as possible as this would demonstrate overlap between the spread of items and the spread of person performance. When the mean of items and the mean of persons are close to 0, the measure is considered to have adequate overlap and demonstrates that the items are able to differentiate skill levels of the persons tested. For the Decomposition task, the mean of items was -.16 logits whereas the mean of students was 2.18 logits (Figure 3). These means indicate that the items are targeted to differentiate skills as they are not closely aligned; however, the student population performed

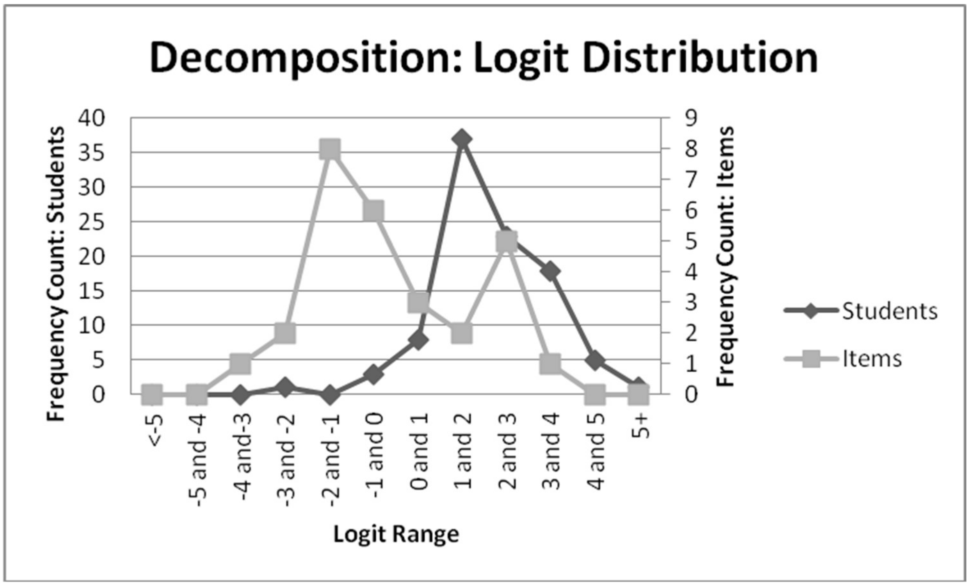

Figure 3. Logit Distribution for Decomposition Task

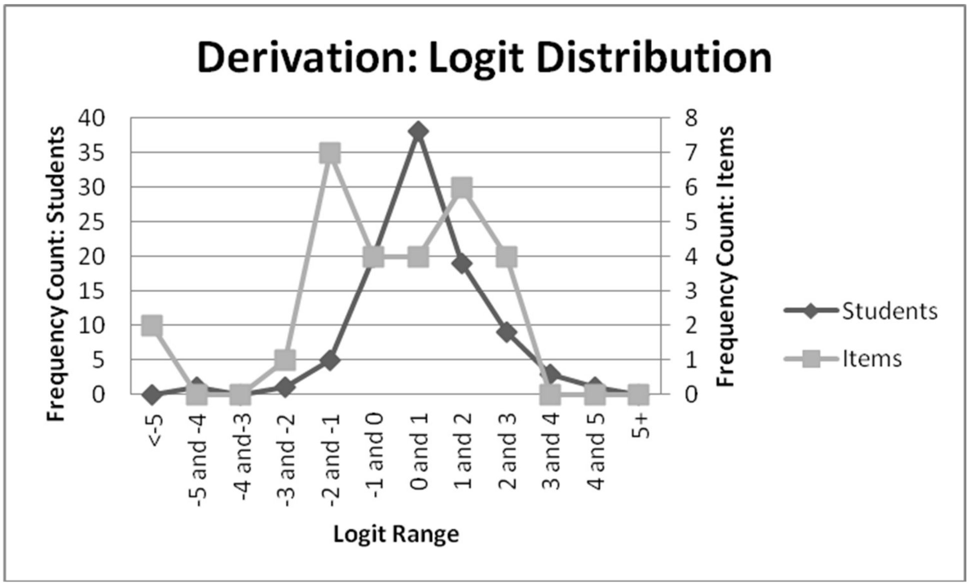

Figure 4. Logit Distribution for Derivation Task

much higher than that for which the Decomposition task is targeted. In contrast, the mean of items (-.23 logits) and the mean of students (.72 logits) for the Derivation task suggest that the items of this task are targeted adequately to the student population tested (Figure 4).

The range of items indicates how large the spread of items covers and the range of the persons' abilities. The Decomposition task items ranged from -4.88 logits to 3.88 logits, demonstrating that the extreme items are almost the same distance from the mean. Thus, the items are not skewed in difficulty. In contrast, the Derivation task demonstrated a skew towards easier items with an item range from -6.43 logits to 2.67 logits. In terms of persons, Derivation continues to be near 0 for the mean logits for persons (.72 logits) whereas Decomposition is further away (2.18 logits).

Last, Rasch analyses examine whether the data fits the expectations of the Rasch analysis (i.e., model-data fit; see Tables 2 and 3). The difficulty hierarchy of the items is inspected by looking at the fit of the data to the analysis expectations (i.e., the model); items that are outside of the fit range may not function as the measure intends in terms of difficulty (Engelhard, 2013). Persons that are outside of the fit range have unexpected responses (Engelhard, 2013). Within Rasch analyses, the data fit is examined by using infit and outfit statistics. To fall within acceptable fit ranges, the items infit and outfit statistics need to be between 0.60 and 1.40; the closer the infit and outfit statistics are to 1, the better the fit (Bond & Fox, 2015). Infit statistics detect whether an item response within the "on-target" items is unexpected (e.g., the consistency in which a person with high ability scores correctly and a person with low ability scores incorrectly) whereas the outfit statistics examined outlying items and responses.

Within the Decomposition task, all items within the task had infit statistics that fell within the acceptable range. However, there were several items that would be considered outliers: items

Table 2

Summary Statistics for Rasch Model: Decomposition Task

	Persons	Items
M	2.18	-0.16
SD	1.26	1.98
N	96	28
Infit *M*	0.98	0.99
Infit *SD*	0.31	0.14
Outfit *M*	0.96	0.99
Outfit *SD*	1.12	0.65
Reliability of Separation	0.71	0.92
Chi-square	322.5	681.2**
df	95	27

Note. * $p \leq 0.05$, ** $p \leq 0.01$

Table 3

Summary Statistics for Rasch Model: Derivation Task

	Persons	Items
M	0.72	-0.23
SD	1.30	2.19
N	96	28
Infit *M*	1.0	1.01
Infit *SD*	0.29	0.14
Outfit *M*	0.93	0.93
Outfit *SD*	0.56	0.21
Reliability of Separation	0.83	0.95
Chi-square	407.8**	704.1**
df	95	27

Note. * $p \leq 0.05$, ** $p \leq 0.01$

Table 4

Item Statistics for the Decomposition Task

Item	Average Rating	Measure (logits)	*SE*	Infit	Outfit
1. Growth. She wanted her plant to _____. *[grow]*	0.98	-2.43	0.75	1.11	0.28
2. Dryer. Put the wash out to _____. *[dry]*	0.96	-1.61	0.54	1.13	0.62
3. Variable. The time of his arrival did not _____. *[vary]*	0.48	2.29	0.23	0.93	0.96
4. Width. The mouth of the river is very _____. *[wide]*	0.67	1.3	0.24	0.99	0.98
5. Density. The smoke in the room was very _____. *[dense]*	0.64	1.47	0.24	1.01	1.23
6. Discussion. The friends have a lot to _____. *[discuss]*	0.94	-1.12	0.45	0.89	1.03
7. Famous. The actor would achieve much _____. *[fame]*	0.73	0.93	0.26	0.84	0.77
8. Description. The picture is hard to _____. *[describe]*	0.82	0.27	0.29	1.11	1.17
9. Fifth. The boy counted from one to _____. *[five]*	0.92	-0.76	0.4	0.99	0.66
10. Election. Which person did they _____? *[elect]*	0.96	-1.61	0.54	0.83	1
11. Strength. The girl was very _____. *[strong]*	0.97	-1.96	0.62	1.27	3.86
12. Decision. The boy found it hard to _____. *[decide]*	0.96	-1.61	0.54	1.17	1.19
13. Popularity. The girl wants to be _____. *[popular]*	0.98	-2.43	0.75	0.69	0.17
14. Runner. How fast can she _____? *[run]*	1	-4.46	1.84	N/A	N/A
15. Publicity. His views were made _____. *[public]*	0.4	2.72	0.24	1.05	1.07
16. Difference. Do their opinions _____? *[differ]*	0.39	2.78	0.24	0.98	1.03
17. Originality. That painting is very _____. *[original]*	0.83	0.18	0.3	0.82	0.78
18. Agreeable. With that statement I could not _____. *[agree]*	0.92	-0.76	0.4	1	0.94
19. Courageous. The man showed great _____. *[courage]*	0.89	-0.34	0.35	0.95	0.71
20. Admission. How many people will they _____? *[admit]*	0.4	2.72	0.24	1.14	1.28
21. Dangerous. Are the children in any	0.95	-1.34	0.49	1.22	0.92

_____? *[danger]*					
22. Reduction. The overweight man was trying to _____. *[reduce]*	0.21	3.88	0.28	0.94	0.96
23. Baker. She put the bread in to _____. *[bake]*	0.49	2.23	0.23	1.03	1.02
24. Division. The cake is hard to _____. *[divide]*	0.89	-0.34	0.35	1.07	1.53
25. Guidance. The map was her _____. *[guide]*	0.9	-0.47	0.36	0.93	0.86
26. Continuous. How long will the storm _____? *[continue]*	0.96	-1.61	0.54	0.84	0.41
27. Reliable. On his friend he could always _____. *[rely]*	0.92	-0.76	0.4	0.78	0.43
28. Acceptance. Is that an offer you can _____? *[accept]*	0.96	-1.61	0.54	0.89	0.86

Note: Infit and outfit are measures of misfit when examining model-data fit. N/A refers to items in which infit and outfit statistics are not calculated.

Table 5

Item Statistics for the Derivation Task

Item	Average Rating	Measure (logits)	*SE*	Infit	Outfit
1. Warm. He chose the jacket for its _____. *[warmth]*	0.54	0.53	0.23	0.89	0.83
2. Teach. He was a very good _____. *[teacher]*	0.99	-5.02	1.12	1.49	0.95
3. Permit. (to allow) Father refused to give _____. *[permission]*	0.51	0.69	0.23	0.97	0.92
4. Profit. Selling lemonade in summer is _____. *[profitable]*	0.38	1.39	0.24	0.95	0.96
5. Appear. He cared about his _____. *[appearance]*	0.86	-1.58	0.33	1	0.71
6. Express. 'OK' is a common _____. *[expression]*	0.72	-0.44	0.25	0.91	0.8
7. Four. The cyclist came in _____. *[fourth]*	0.34	1.57	0.24	1.1	1
8. Remark. The speed of the car was _____. *[remarkable]*	0.82	-1.19	0.3	1	0.86
9. Protect. She wore glasses for _____. *[protection]*	0.9	-1.95	0.37	1.16	1.01
10. Perform. Tonight is the last _____. *[performance]*	0.94	-2.64	0.47	1.01	0.68
11. Expand. The company planned an _____. *[expansion]*	0.36	1.45	0.24	1.06	1.06
12. Revise. This paper is his second _____. *[revision]*	0.2	2.5	0.28	0.96	0.8
13. Reason. Her argument was quite _____. *[reasonable]*	0.85	-1.47	0.32	1	1.27
14. Major. He won the vote by a _____. *[majority]*	0.18	2.67	0.29	0.91	0.74
15. Deep. The lake was well known for its _____. *[depth]*	0.23	2.27	0.27	0.96	0.78
16. Equal. Boys and girls are treated with _____. *[equality]*	0.31	1.74	0.25	1.1	1.08
17. Long. They measured the ladder's _____. *[length]*	0.47	0.9	0.23	1.17	1.54
18. Adventure. The trip sounded _____. *[adventurous]*	0.9	-1.95	0.37	0.86	0.82
19. Absorb. She chose the sponge for its _____. *[absorption or absorbance]*	0.5	0.74	0.23	1	0.97
20. Active. He tired after so much _____. *[activity]*	0.45	1.01	0.23	1.22	1.26
21. Swim. She was a strong _____.	1	-6.43	1.9	N/A	N/A

[swimmer]					
22. Human. The kind man was known for his _____. *[humanity]*	0.27	2	0.26	1.01	0.95
23. Wash. Put the laundry in the _____. *[washer]*	0.77	-0.78	0.27	1.14	1.21
24. Humor. The story was quite _____. *[humorous]*	0.76	-0.71	0.27	0.86	0.72
25. Assist. The teacher will give you _____. *[assistance]*	0.7	-0.31	0.25	1.07	1
26. Mystery. The dark glasses made the man look _____. *[mysterious]*	0.85	-1.47	0.32	0.88	0.67
27. Produce. The play was a grand _____. *[production]*	0.35	1.51	0.24	0.87	0.84
28. Glory. The view from the hill top was _____. *[glorious]*	0.85	-1.47	0.32	0.84	0.67

Note: Infit and outfit are measures of misfit when examining model-data fit. N/A refers to items in which infit and outfit statistics are not calculated.

1, 11, 13, 24, 26, and 27 (see Table 4). In contrast, there is only 1 outlying item (item 17) on the Derivation task, but one item (item 2) that had unexpected responses despite the item and person locations being targeted well (see Table 5).

Looking at the infit statistics for the persons for the Decomposition task, 13 of the 96 values were outside the range of 0.6 to 1.4 (see Table 6). The outfit statistics indicated that the majority of the participants are outliers. Of the 96 participants, 59 of the participants had outfit mean square values outside of the acceptable range. Similarly, the Derivation task had 14 participants with infit statistics outside of the range and 44 participants with outfit statistics outside of the acceptable range (see Table 7).

Item Order and Discontinue Rule Analyses

Correlations between the items from each of the item orders (i.e., original, theoretical, and data-based) were conducted to examine how closely related the item-orders are to one another. Spearman Rho correlations were computed and were found to be insignificant. For the Decomposition task, the original item order had insignificant, negligible correlations with the theoretical (r = .082, p = .678) and data-based (r = .028, p = .888) item orders. The theoretical and data-based item orders for the Decomposition task also had an insignificant, negligible relationship (r = -.005, p = .980). Similarly, the Derivation task original item order had insignificant, negligible relationships with the theoretical (r = -.159, p = .418) and data-based (r = .034, p = .862) item orders and there was an insignificant, negligible correlation between the Derivation task theoretical and data-based item orders (r = .252, p = .196). In summary, results suggested that the original, theoretical, and data-based item orders were not correlated with each other either for the Decomposition or Derivation task. This indicated that the order of items in the three item orders were vastly different from one another.

Table 6

Person Statistics for the Decomposition Task

Person	Average Rating	Measure (logits)	*SE*	Infit	Outfit
1	0.75	1.63	0.58	1.04	0.84
2	0.82	2.34	0.62	0.81	0.36
3	0.57	0.2	0.5	1.97	2.75
4	0.71	1.31	0.56	0.6	0.43
5	0.71	1.31	0.56	0.71	0.47
6	0.93	3.81	0.82	1.26	0.58
7	0.93	3.81	0.82	0.92	0.55
8	0.79	1.97	0.6	0.55	0.28
9	0.82	2.34	0.62	0.94	0.58
10	0.71	1.31	0.56	0.36	0.23
11	0.86	2.74	0.66	1.13	1.21
12	0.54	-0.05	0.5	0.86	0.73
13	0.82	2.34	0.62	0.48	0.22
14	0.21	-2.25	0.55	0.83	0.48
16	0.64	0.73	0.53	0.82	0.78
17	0.75	1.63	0.58	0.93	0.79
18	0.75	1.63	0.58	1.33	1.01
19	0.89	3.22	0.72	1.24	1.16
20	0.89	3.22	0.72	0.87	0.39
21	0.68	1.01	0.54	0.98	1.19
22	0.75	1.63	0.58	1.16	1.46
23	0.75	1.63	0.58	1.16	0.93
24	0.79	1.97	0.6	0.55	0.28
25	0.79	1.97	0.6	0.96	0.71
26	0.82	2.34	0.62	0.99	0.5
27	0.79	1.97	0.6	1.11	0.84
28	0.61	0.46	0.52	1.2	1.02
29	0.75	1.63	0.58	1.22	1.22
30	0.71	1.31	0.56	1.09	0.9
31	0.89	3.22	0.72	0.87	0.39
32	0.61	0.46	0.52	1.15	1.05
33	0.57	0.2	0.5	1.42	1.52
34	0.75	1.63	0.58	0.89	0.63
35	0.71	1.31	0.56	0.52	0.33
36	0.93	3.81	0.82	1.36	9
37	0.79	1.97	0.6	0.77	1.55
38	0.86	2.74	0.66	0.74	0.29
39	0.75	1.63	0.58	1.29	1.08
40	0.89	3.22	0.72	1.25	0.98
41	1	6.02	1.86	N/A	N/A
42	0.89	3.22	0.72	0.75	0.27

43	0.82	2.34	0.62	0.48	0.22
44	0.82	2.34	0.62	0.63	0.29
45	0.63	0.73	0.53	0.88	0.81
46	0.96	4.68	1.08	1.07	0.3
47	0.96	4.68	1.08	1.27	0.96
48	0.71	1.31	0.56	0.87	1.04
49	0.89	3.22	0.72	0.93	0.54
50	0.86	2.74	0.66	1.58	4.02
51	0.54	-0.05	0.5	1.21	1.28
52	0.82	2.34	0.62	1.1	0.65
53	0.93	3.81	0.82	1.14	0.37
54	0.71	1.31	0.56	1.01	0.71
55	0.79	1.97	0.6	0.76	0.38
56	0.71	1.31	0.56	1.06	0.96
57	0.86	2.74	0.66	1.12	0.5
58	0.75	1.63	0.58	0.6	0.35
59	0.93	3.81	0.82	1.36	2.55
60	0.86	2.74	0.66	1.22	0.79
61	0.89	3.22	0.72	0.88	0.42
62	0.71	1.31	0.56	0.84	1.34
63	0.86	2.74	0.66	0.73	0.28
64	0.89	3.22	0.72	0.75	0.27
65	0.71	1.31	0.56	1.61	1.44
66	0.54	-0.05	0.5	0.66	0.48
67	0.79	1.97	0.6	1.13	1.09
68	0.75	1.63	0.58	0.8	0.69
69	0.89	3.22	0.72	0.63	0.22
70	0.75	1.63	0.58	0.83	0.72
71	0.96	4.68	1.08	1.07	0.3
72	0.79	1.97	0.6	0.63	0.32
73	0.71	1.31	0.56	0.92	1.11
74	0.79	1.97	0.6	1.49	2.2
75	0.79	1.97	0.6	0.61	0.32
76	0.82	2.34	0.62	1.02	1.76
77	0.82	2.34	0.62	0.83	1.39
78	0.96	4.68	1.08	1.07	0.3
79	0.68	1.01	0.54	1.19	1.48
80	0.57	0.2	0.5	1.37	1.52
81	0.79	1.97	0.6	0.6	0.32
82	0.64	0.73	0.53	1.74	1.85
83	0.89	3.22	0.72	0.75	0.27
84	0.82	2.34	0.62	1.32	1.43
85	0.89	3.22	0.72	0.87	0.39
86	0.89	3.22	0.72	0.75	0.27
87	0.86	2.74	0.66	0.62	0.25
88	0.86	2.74	0.66	0.89	0.65

89	0.82	2.34	0.62	1.18	1.07
90	0.86	2.74	0.66	1.36	4.86
91	0.86	2.74	0.66	0.94	1.44
92	0.96	4.68	1.08	0.63	0.11
93	0.89	3.22	0.72	1.31	0.67
94	0.86	2.74	0.66	0.62	0.25
95	0.71	1.31	0.56	1.35	1.43
96	0.75	1.63	0.58	0.62	0.51
97	0.86	2.74	0.66	1.46	0.95

Note: Infit and outfit are measures of misfit when examining model-data fit. N/A refers to items in which infit and outfit statistics are not calculated.

Table 7

Person Statistics for the Decomposition Task

Person	Average Rating	Measure (logits)	*SE*	Infit	Outfit
1	0.61	0.7	0.49	0.81	0.59
2	0.5	-0.02	0.49	0.61	0.46
3	0.39	-0.77	0.51	1.33	1.12
4	0.46	-0.27	0.5	0.91	0.8
5	0.5	-0.02	0.49	0.95	0.9
6	0.75	1.69	0.52	1.18	0.9
7	0.79	1.97	0.53	1.2	0.78
8	0.32	-1.29	0.53	1.04	1.06
9	0.96	4.3	1.05	1.04	0.52
10	0.5	-0.02	0.49	1.16	1.05
11	0.79	1.97	0.53	1.17	0.72
12	0.36	-1.03	0.51	1.18	1.06
13	0.64	0.94	0.49	0.8	0.6
14	0.07	-4.82	1.21	0.33	0.05
16	0.32	-1.29	0.53	1.06	0.85
17	0.57	0.46	0.49	1.57	2.25
18	0.46	-0.27	0.5	0.76	0.65
19	0.86	2.6	0.6	0.83	0.47
20	0.75	1.69	0.52	0.78	0.48
21	0.68	1.18	0.5	0.77	0.78
22	0.43	-0.51	0.5	0.59	0.43
23	0.75	1.69	0.52	0.78	0.5
24	0.5	-0.02	0.49	1.06	0.98
25	0.57	0.46	0.49	1.23	1.1
26	0.64	0.94	0.49	1.34	1.19
27	0.54	0.22	0.49	1.64	1.45
28	0.61	0.7	0.49	0.86	0.73
29	0.5	-0.02	0.49	1.11	1.04
30	0.61	0.7	0.49	0.99	0.85
31	0.68	1.18	0.5	0.96	0.84
32	0.18	-2.61	0.65	0.89	1.46
33	0.64	0.94	0.49	0.58	0.41
34	0.68	1.18	0.5	0.96	1.42
35	0.54	0.22	0.49	1.09	0.89
36	0.57	0.46	0.49	0.7	0.55
37	0.57	0.46	0.49	1.22	1.19
38	0.61	0.7	0.49	0.59	0.43
39	0.74	1.63	0.53	0.91	1.78
40	0.61	0.7	0.49	1.42	1.66
41	0.93	3.51	0.77	1.05	0.54
42	0.61	0.7	0.49	1.34	1.36

43	0.68	1.18	0.5	1.07	0.9
44	0.39	-0.77	0.51	1.13	1.09
45	0.39	-0.77	0.51	0.7	0.49
46	0.86	2.6	0.6	1.24	1.61
47	0.75	1.69	0.52	1.2	0.91
48	0.32	-1.29	0.53	0.95	0.77
49	0.61	0.7	0.49	0.77	0.6
50	0.54	0.22	0.49	0.51	0.42
51	0.5	-0.02	0.49	0.69	0.66
52	0.54	0.22	0.49	0.82	0.94
53	0.82	2.27	0.56	0.6	0.33
54	0.64	0.94	0.49	0.81	0.67
55	0.61	0.7	0.49	0.93	0.86
56	0.39	-0.77	0.51	1.52	1.41
57	0.86	2.6	0.6	1.15	0.82
58	0.43	-0.51	0.5	1.42	1.55
59	0.82	2.27	0.56	1.12	0.69
60	0.64	0.94	0.49	1.31	1.14
61	0.5	-0.02	0.49	0.88	0.7
62	0.57	0.46	0.49	1.06	0.82
63	0.75	1.69	0.52	1.25	1.24
64	0.82	2.27	0.56	1.02	0.58
65	0.54	0.22	0.49	1.62	1.89
66	0.43	-0.51	0.5	1.24	4.12
67	0.82	2.27	0.56	0.72	0.39
68	0.64	0.94	0.49	1.31	1.34
69	0.71	1.43	0.5	0.74	0.57
70	0.48	-0.18	0.51	0.98	0.83
71	0.89	3	0.66	0.85	0.39
72	0.64	0.94	0.49	0.75	0.7
73	0.5	-0.02	0.49	0.96	0.81
74	0.75	1.69	0.52	0.98	0.63
75	0.64	0.94	0.49	1	0.8
76	0.75	1.69	0.52	0.94	1.32
77	0.82	2.27	0.56	1.22	2.09
78	0.75	1.69	0.52	1.51	1.17
79	0.57	0.46	0.49	1.53	1.62
80	0.25	-1.88	0.57	0.78	0.43
81	0.54	0.22	0.49	0.69	0.55
82	0.64	0.94	0.49	0.82	0.84
83	0.61	0.7	0.49	0.62	0.45
84	0.71	1.43	0.5	0.87	0.57
85	0.46	-0.27	0.5	1.13	0.96
86	0.82	2.27	0.56	0.98	0.6
87	0.57	0.46	0.49	0.68	0.54
88	0.61	0.7	0.49	0.85	0.81

89	0.5	-0.02	0.49	0.6	0.5
90	0.79	1.97	0.53	1.02	0.64
91	0.64	0.94	0.49	1.25	1.33
92	0.79	1.97	0.53	0.88	0.52
93	0.89	3	0.66	0.84	0.51
94	0.57	0.46	0.49	1.15	1.82
95	0.61	0.7	0.49	1.99	2.39
96	0.64	0.94	0.49	0.75	0.54
97	0.54	0.22	0.49	0.61	0.48

Note: Infit and outfit are measures of misfit when examining model-data fit.

Table 8

Summary of Data for Participants That Met the Discontinue Rule

	# of Participants That Met the Discontinue Rule	Percentage of Participants	Average # of Items Not Administered	Range of Items Not Administered
Decomposition				
Original	2	2.08%	14	8 to 20
Theoretical	2	2.08%	14.5	14 to 15
Data-Based	11	11.46%	2.45	0 to 12
Derivation				
Original	12	12.50%	11.67	1 to 20
Theoretical	14	14.58%	9	0 to 16
Data-Based	41	42.71%	5.07	0 to 20

A discontinue rule of 6 consecutive incorrect responses (as used in Fracasso et al., 2014; Herman et al., 2013; Tighe & Binder, 2013) was applied to produce a score (see Appendix H for an example) for each participant for each of the item orderings. Across item orders, with one exception, few participants met the discontinue rule for either TMS task (see Table 8). The exception was for the Derivation task using the data-based item order with the applied discontinue rule in which 42.71% (n = 41) of the participants met the discontinue rule of 6 consecutive incorrect responses. Using the data-based item order with the applied discontinue rule, the Derivation task would have been shortened an average of 5.07 items during administration.

One-way analyses of variance were used to examine whether the resulting student scores from each item order (i.e., original, theoretical, and data-based) with an applied discontinue rule differed significantly from one another. The scores from each of the Decomposition item orders with the applied discontinue rule were not significantly different from one another, $F(2, 282)$ = .038, p = .962 (see Table 9 for descriptive statistics). Similarly, the scores from the Derivation item orders with the applied discontinue rule employed were not significantly different from one another, $F(2, 279)$ = .194, p = .824. These results suggest that the scores obtained from each of the item orders provide the same information regarding students' MA.

To examine differences between the scores of each item order when applying the discontinue rule and scores without the discontinue rule applied further, dependent t-tests were conducted. For the Decomposition task, the scores produced by all items without an applied discontinue rule were not significantly different than the scores produced by each of the three item orders with the applied discontinue rule (see Table 9). However, the Derivation scores produced by all items without an applied discontinue rule (M = 17.00, SD = 4.67) were

Table 9

Comparison of Total Score and Scores Gained from Each Item Order with an Applied Discontinue Rule

	M	*SD*	*t*	*df*	*p*	*d*
Decomposition						
Original Item Order	22.02	3.713	1.422	94	.158	-0.003
Theoretical Item Order	21.94	3.886	1.469	94	.145	0.018
Data-Based Item Order	22.08	3.420	1.422	94	.158	-0.020
Derivation						
Original Item Order	16.29	5.799	3.352	93	.001	0.135
Theoretical Item Order	16.65	5.232	2.650	93	.009	0.071
Data-Based Item Order	16.74	4.875	3.315	93	.001	0.054

significantly different from the scores produced by the original item order ($M = 16.29$, $SD = 5.80$), $t(94) = 3.352$, $p < .01$, $d = 0.135$, theoretical item order ($M = 16.65$, $SD = 5.23$), $t(94) = 2.65$, $p < .01$, $d = 0.071$, and the data-based item order ($M = 16.74$, $SD = 4.875$), $t(94) = 3.315$, $p < .01$, $d = 0.054$ with the applied discontinue rule.

Rank-order correlations were conducted between the total score of all items and each of the item-orders with the applied discontinue rule. Spearman's rank-order correlations were calculated to determine the variation of the differences between the scores established from the resulting data from each of these scoring procedures and to help determine the appropriateness of the discontinue rule. The scores from the Decomposition and Derivation tasks yielded strong, positive correlations between each of the item orders (see Tables 10 and 11). These results demonstrate that similar information was gained from the scores from each of the TMS tasks regardless of the item order or whether the applied discontinue rule was employed.

Completion Time

Each participant's completion times for each of the TMS tasks were recorded. For each of the two TMS tasks, an estimate time per item was calculated by dividing the total time to complete the task by 28 (the number of items within the task). In order to examine the extent to which employing a discontinue rule impacted test administration time, one-way analyses of variance were used to compare the students' completion times for all items within each of the TMS tasks to the estimated completion times for each of the items orders with an applied discontinue rule. The analyses indicated that the completion times did not vary significantly: Decomposition task $F(3, 380) = .041$, $p = .989$ and Derivation task $F(3, 383) = 1.56$, $p = .199$. However, the concern from prior researchers was that the TMS tasks were too long to

Table 10

Spearman's Rank-Order Correlations Examining Decomposition Task Item Orders

	All Items	Original Item Order	Theoretical Item Order	Data-Based Item Order
All Items	-	-	-	-
Original Item Order	1.00**	-	-	-
Theoretical Item Order	1.00**	1.00**	-	-
Data-Based Item Order	1.00**	1.00**	1.00**	-

Note. * $p \leq 0.05$, ** $p \leq 0.01$

Table 11

Spearman's Rank-Order Correlations Examining Derivation Task Item Orders

	All Items	Original Item Order	Theoretical Item Order	Data-Based Item Order
All Items	-	-	-	-
Original Item Order	.986**	-	-	-
Theoretical Item Order	.994**	.981**	-	-
Data-Based Item Order	.985**	.970**	.977**	-

Note. * $p \leq 0.05$, ** $p \leq 0.01$

Table 12

Comparison of Total Completion Time and Estimated Completion Times for Each Item Order

	M	*SD*	*t*	*df*	*p*	*d*
Decomposition						
Original Item Order	227.73	72.99	1.400	94	0.165	0.037
Theoretical Item Order	228.17	74.63	1.420	94	0.159	0.031
Data-Based Item Order	228.25	72.36	1.995	94	0.049	0.030
Derivation						
Original Item Order	265.65	85.95	3.508	93	0.001	0.218
Theoretical Item Order	270.38	89.32	3.246	93	0.002	0.165
Data-Based Item Order	259.89	84.13	5.172	93	0.000	0.289

administer; thus, dependent *t*-tests were used to compare the students' completion times further (see Table 12).

The entire Decomposition task (i.e., all items without a discontinue rule applied) took an average of 230.42 sec (SD = 72.27) to administer. The completion time to administer the Decomposition task did not significantly differ between all items without an applied discontinue rule and the original item order with the applied discontinue rule (M = 227.73, SD = 72.99), $t(94) = 1.40$, $p = .165$, $d = 0.037$. Similarly, there was no significant difference between the administration times when all items were given without a discontinue rule and the theoretical item order with the applied discontinue rule (M = 228.17, SD = 74.63), $t(94) = 1.42$, $p = .159$, $d = 0.031$). Although a significant difference was found between the Decomposition task using the data-based item order with an applied discontinue rule (M = 228.25, SD = 72.36) and the Decomposition task when administering all items without a discontinue rule, the effect size suggests that the difference is not meaningful, $t(94) = 1.995$, $p = .049$, $d = 0.030$.

In contrast to the results for the completion time for the Decomposition task, the completion time for the Derivation task with the applied discontinue rule, regardless of item order, differed significantly from the administration time of all items without utilizing a discontinue rule. The completion time for all items on the Derivation task (M = 284.47, SD = 86.24) was significantly longer than when the discontinue rule was applied to task following the original item order (M = 265.65, SD = 85.95), $t(93) = 3.508$, $p <.001$, $d = 0.218$, the theoretical item order (M = 269.96, SD = 89.26), $t(93) = 3.246$, $p < .01$, $d = 0.165$, and the data-based item order (M = 259.89, SD = 84.13), $t(93) = 5.172$, $p < .000$, $d = 0.289$.

Reliability

Reliability was examined for the task items to assess whether the data yielded from each of the TMS tasks formed a reliable scale. The most popular reliability indicator is Cronbach's alpha, otherwise known as coefficient alpha (α), which measures internal consistency. Generally, acceptable reliability is indicated by a coefficient alpha of ≥ 0.80 where the closer it approaches 1, the stronger the reliability (Cortina, 1993). Thus, the current study utilized coefficient alpha as the reliability indicator as preliminary evidence. The alpha for the Decomposition task was .71 while the alpha for the Derivation task was .83, which indicates that the items forming the TMS tasks have adequate internal consistency.

Validity

Preliminary construct validity evidence should suggest that a measure is assessing the construct it was developed to assess (Bond & Fox, 2015). Within this study, construct validity was analyzed by examining unidimensionality and convergent relationships. The unidimensionality of the TMS tasks was examined through the Rasch analyses discussed earlier. To recap, the Rasch statistics for both the Decomposition (43.96%) and Derivation (42.02%) tasks suggest that the TMS tasks are strong measures of MA.

Convergent validity (i.e., constructs that should relate to one another do relate) was assessed by comparing participants' phonological awareness and reading comprehension scores to the scores yielded from the TMS tasks for each item-ordering procedure via Pearson correlational analyses (Table 13). These analyses demonstrated low to moderate relationships between PA and the TMS tasks and moderate relationships between reading comprehension and the TMS tasks regardless of item order

Table 13

Pearson Correlations Examining Derivation Task Item Orders

	Word Attack (PA)	Passage Comprehension
Decomposition Task		
All Items	.525**	.584**
Original Item Order	.531**	.567**
Theoretical Item Order	.486**	.575**
Data-Based Item Order	.526**	.583**
Derivation Task		
All Items	.460**	.595**
Original Item Order	.423**	.559**
Theoretical Item Order	.460**	.598**
Data-Based Item Order	.431**	.572**

Note. * $p \leq 0.05$, ** $p \leq 0.01$

Table 14

Hierarchical Multiple Regression Analysis Summary for the Contributions of Phonological Awareness and Morphological Awareness to Reading Comprehension

	B	*SE B*	β	*t*	*p*
Decomposition					
Phonological Awareness	.265	.090	.287	2.947	.004
Original Item Order	.378	.089	.414	4.250	.000
Derivation					
Phonological Awareness	.294	.084	.316	3.495	.001
Original Item Order	.250	.053	.425	4.691	.000

Relationships to Other Literacy Skills

To examine the relationships between MA, phonological awareness (PA), and reading comprehension, several hierarchical multiple linear regression models were run in order to understand the amount of variance explained by PA and MA within reading comprehension. Each model included PA entered into the first block since research states it is acquired first (e.g., Adams, 1990), followed by one of the TMS tasks in the second block (see Table 14 for example). PA significantly explained 25.7% of the variance within reading comprehension. MA significantly explained an additional 13.9% of the variance within reading comprehension when using scores yielded from all Decomposition task items administered, $F(2, 93) = 29.825$, $p < .001$. When the Decomposition task using the original item order with an applied discontinue rule was entered, the amount of variance explained by MA decreased to 12.3%, $F(2, 93) = 27.913$, $p < .001$. The theoretical item ordered Decomposition task explained 14.1% of the variance within reading comprehension, $F(2, 93) = 30.161$, $p < .001$, whereas the data-based item ordered Decomposition task explained 13.8% of the variance, $F(2, 93) = 29.782$, $p < .001$.

The second set of hierarchical linear regression models examined explained variance within reading comprehension when PA and the Derivation task were entered. PA significantly explained 24.6% of the variance within reading comprehension within the second set of models. MA significantly explained an additional 17% of the variance within reading comprehension when using scores yielded from all Derivation task items administered, $F(2, 92) = 32.182$, $p < .001$. Similar to the results seen with the Decomposition task, the amount of variance explained by MA decreased to 14.8% when the Derivation task using the original item order with an applied discontinue rule was included in the model, $F(2, 92) = 29.322$, $p < .001$. MA as gathered by the data-based item ordered Derivation task with an applied discontinue rule significantly contributed to reading comprehension by explaining 15.8% of the variance, $F(2, 92) = 30.559$, $p < .001$, which is also a decrease in the amount of explained variance when compared to the explained variance without the discontinue rule applied. Based on these results, it appears that the theoretical item order with an applied discontinue rule explains the most similar amount of variance (17.4%) within reading comprehension, $F(2, 92) = 32.607$, $p < .001$, when compared to the explained variance when all items are given.

CHAPTER 4

DISCUSSION

Prior research suggests that morphological awareness (MA) is an important skill that contributes to reading and writing abilities. However, there has been inconsistent use of measures examining MA within research (e.g., Carlisle & Nomanbhoy, 1993; Kieffer & Lesaux, 2008; McCutchen et al., 2008; Tong et al., 2011), suggesting that researchers do not necessarily believe an adequate measure of MA exists. In fact, researchers have frequently adapted existing MA measures as opposed to simply using the original versions employed in prior research (e.g., Carlisle & Nomanbhoy, 1993; Fracasso et al., 2014; Herman, Cote, Reilly, & Binder, 2013; Kieffer & Lesaux, 2008; McCutchen et al., 2008; Tighe & Binder, 2013; Tong et al., 2011). One major concern of researchers not employing the same measures but instead modifying them is that one cannot be certain that modifications do not also modify the construct being assessed by the original measures. Modifying measures may impact the reliability and validity of the instrument, which may result in varied and inaccurate interpretations of research across studies (AERA, APA, & NCME, 2014).

The focus of the present study was to investigate the construction and psychometric properties of the Test of Morphological Structure (TMS; Carlisle, 2000), a commonly used but frequently modified, measure of MA. This first step in identifying a potential screener and/or diagnostic measure of students' MA skills was accomplished by utilizing Rasch analyses to examine the Decomposition and Derivation tasks that comprise the TMS. Specifically, unidimensionality, the item hierarchy, and person separation were investigated to understand the construction of the tasks and reliability and validity were used to examine the psychometric properties of the TMS tasks. A second focus was to examine the impact if of the discontinue rule regularly employed by prior researchers on resulting data when applied to each of three item-ordering procedures for each task (i.e., Decomposition and Derivation). The three item orders included the following: (a) original order, which was the order of items employed within the original TMS and prior research studies; (b) theoretical order, which involved ordering items within each task from easiest to most difficult based upon the expected developmental trajectory of MA skills; and (c) data-based order, which involved the ordering of items within each task based upon Rasch analyses of the data collected as part of this study. It is important to note that items were not administered to participants in these three orders. Rather, items were administered to each participant in a randomized order. Each student's data were then ordered according to each of these three orders and scored using the discontinue rule of "six consecutive incorrect items." This discontinue rule is the rule that researchers have regularly employed within the literature (Fracasso et al., 2014; Herman et al., 2013; Tighe & Binder, 2013). The final focus of the current study was to replicate prior research (i.e., Deacon & Kirby, 2004; Kirby et al., 2011; Mahony, 1994; Nagy et al., 2006) by investigating the relationships between MA and other literacy skills.

Construction of the TMS Tasks

Rasch analyses were conducted in an effort to examine unidimensionality and model-data fit. Overall, the TMS tasks were shown to be strong measures of MA based on analyses examining unidimensionality. There were minor concerns regarding the model-data fit of persons for both the Decomposition and Derivation tasks. Rasch analyses were also employed to examine item difficulty, and item and person distribution, specifically whether it may be beneficial to remove or revise TMS items in order to create a stronger measure of MA. These results are discussed in greater detail within the sections below for both TMS tasks.

The Decomposition task. Results from Rasch analyses related to person separation reliability demonstrated that participants scored closer to the ceiling than was expected (i.e., uneven distribution). These results suggest that the Decomposition task may be more appropriate for assessing the MA skills of students with a lower ability level than participants assessed in this study. The observed ceiling effects indicate that either MA begins to develop earlier than previous research suggested (e.g., Anglin, 1993; Ku & Anderson, 2003; Kuo & Anderson, 2006; Mahony, Singson, & Mann, 2000; Nagy, Diakidoy & Anderson 1993; Singson, Mahony, & Mann, 2000; Wolter et al., 2009) or that the students assessed in this study functioned higher than the typical fourth grade population. If it is the former, the Decomposition task may be useful in distinguishing skill levels for struggling fourth grade students but not for highly skilled fourth grade students. The Decomposition task should be further assessed by having both younger and older students take the task, especially if new items are added to address the uneven distribution of items. By including a wider age range of students, a more comprehensive picture of how the measure functions and differentiates student ability and developmental progression will emerge.

Rasch analyses were also employed to examine the spread of item difficulty (i.e., even item distribution) of the items that make up the Decomposition task, with the Wright Map and logit spread indicating that there is no need to either add or remove any items. However, despite good item spread, the distribution overlap for participants and items was not ideal as many of the items were lower on the logit scale than the majority of the participants (as seen on the Wright Map in Figure 5). In fact, almost half of the items were below 0 logits (17 items) while there were only 4 student scores below 0 logits (see Figure 3 for an additional representation).

```
+------------------------------------------------------------------------------------------+
|Measr|+child                                  |-Items                                      |
|-----+----------------------------------------+--------------------------------------------|
|   5 + 41          High Performing Participants +                            Difficult Items |
|     |                                         |                                            |
|     | 46 47 71 78 92                          |                                            |
|     |                                         |                                            |
|     |                                         |                                            |
|     |                                         |                                            |
|   4 +                                         +                                            |
|     | 6  7  36 53 59                          | Reduction                                  |
|     |                                         |                                            |
|     |                                         |                                            |
|     |                                         |                                            |
|     | 19 20 31 40 42 49 61 64 69 83 85 86 93  |                                            |
|   3 +                                         +                                            |
|     |                                         | Difference                                 |
|     | 11 38 50 57 60 63 87 88 90 91 94 97     | Admission    Publicity                     |
|     |                                         |                                            |
|     | 2  9  13 26 43 44 52 76 77 84 89        | Variable                                   |
|     |                                         | Baker                                      |
|   2 + 8  24 25 27 37 55 67 72 74 75 81        +                                            |
|     |                                         |                                            |
|     | 1  17 18 22 23 29 34 39 58 68 70 96     |                                            |
|     |                                         | Density                                    |
|     | 4  5  10 30 35 48 54 56 62 65 73 95     | Width                                      |
|     |                                         |                                            |
|   1 + 21 79                                   + Famous                                     |
|     |                                         |                                            |
|     | 16 45 82                                |                                            |
|     | 28 32                                   |                                            |
|     |                                         | Description                                |
|     | 3  33 80                                | Originality                                |
*   0 * 12 51 66                                *                                            *
|     |                                         |                                            |
|     |                                         | Courageous   Division                      |
|     |                                         | Guidance                                   |
|     |                                         |                                            |
|     |                                         | Agreeable    Fifth       Reliable          |
|  -1 +                                         +                                            |
|     |                                         | Discussion                                 |
|     |                                         | Dangerous                                  |
|     |                                         |                                            |
|     |                                         | Acceptance   Continuous  Decision    Dryer       Election |
|     |                                         |                                            |
|  -2 +                                         + Strength                                   |
|     | 14                                      |                                            |
|     |                                         |                                            |
|     |                                         | Growth       Popularity                    |
|     |                                         |                                            |
|     |                                         |                                            |
|  -3 +             Low Performing Participants + Runner                          Easy Items |
|-----+----------------------------------------+--------------------------------------------|
|Measr|+child                                  |-Items                                      |
+------------------------------------------------------------------------------------------+
```

Figure 5. Wright Map for Decomposition Task

These differences suggest that the Decomposition task was too easy for the participants included in the present study and may be more beneficial for distinguishing the MA skill level of less skilled students.

To further investigate the range of item difficulty and the appropriateness of the Decomposition task items, infit and outfit statistics were examined. Infit statistics from the Rasch analyses examine the consistency with which a person with a strong ability answers an item correctly and a person with low ability answers incorrectly. In contrast, outfit statistics examine whether items and responses fit expectations. Rasch analyses of the Decomposition task indicated 6 of the 28 items were outliers (Items 1, 11, 13, 24, 26, and 27) based on their infit and/or outfit statistics calculated from the current sample. No pattern seemed to exist with in terms of root word frequency or level of transparency for the items (components of MA; for more information see p. 25 & 28). The root word frequency of words within the Decomposition task ranged from 51.9 to 64.1 according to the Standard Frequency Index (Carroll, Davies, Richman, & Davies, 1971). Of the six items identified as outliers, one had no orthographic or phonological shift, one had only a phonological shift, two had orthographic shifts, and two had both an orthographic and phonological shift. Items 24 (-0.34 logits), 26 (-1.61 logits), and 27 (-0.76 logits) were at the same logit level as other items within the task, suggesting that these items could be removed if these items continue to have outlying infit and outfit statistics in future research. Removing these items is unadvisable at this time given that the Decomposition task was too easy for participants of the current study and the current sample was not a representative sample.

The Derivation task. Rasch analyses of the Derivation task items indicated that the items which make up the task are evenly distributed in terms of difficulty. Thus, the Derivation task may be more useful in distinguishing skill levels of fourth grade students than is the Decomposition task. If the data-based item order was administered 42.71% of the participants would have met the discontinue rule, which would have in turn provided varying scores that demonstrate different ability levels. The overlap seen in the Wright Map (Figure 6) between items and persons for the Derivation task demonstrates adequate targeting of items to the participant sample. Despite the adequate targeting, the participant scores on the Derivation task were skewed to the positive logits with 70 of the 96 participants' scores being above 0 logits, whereas the item difficulty range was skewed towards the easier items (i.e., negative logits; see Figure 4 for an additional representation). These results suggest that the Derivation task may be useful in distinguishing the ability levels of younger elementary students or older students who are struggling. However, similar to Decomposition task findings, the participants sampled within the present study performed better than expected for fourth grade students. This suggests that either MA skills develop earlier than suggested by previous research (e.g., Carlisle, 1995; Carlisle & Nomanbhoy, 1993; Carlisle & Stone, 2005; Clark & Cohen, 1984; Jones, 1983; Ku & Anderson, 2003; Mahony et al., 2000; McCutchen et al., 2008; Wolter et al., 2009) or the participants assessed functioned higher than the typical population of fourth grade students. Considering that across both tasks students performed better than expected, it is likely that greater than expected performance was a function of the participants as opposed to prior research underestimating students' abilities. It is important to further assess the construction of the Decomposition task with a sample encompassing a wider age and ability range in order to create a more comprehensive picture.

Rasch analyses were used to investigate the item hierarchy (i.e., the spread of item difficulty). Results of the Rasch analyses suggest that the items which make up the Derivation

```
+------------------------------------------------------------------------------------+
|Measr|+child                               |-Items                                  |
|-----+-------------------------------------+----------------------------------------|
|   5 +          High Performing Participants  +                     Difficult Items |
|     |                                     |                                        |
|     |                                     |                                        |
|     | 9                                   |                                        |
|   4 +                                     +                                        |
|     |                                     |                                        |
|     | 41                                  |                                        |
|     |                                     |                                        |
|   3 + 71 93                               +                                        |
|     |                                     | Major                                  |
|     | 19 46 57                            | Revise                                 |
|     | 53 59 64 67 77 86                   | Deep                                   |
|   2 + 7  11 90 92                         + Human                                  |
|     | 6  20 23 39 47 63 74 76 78          | Equal                                  |
|     | 69 84                               | Expand     Four      Produce    Profit  |
|     | 21 31 34 43                         |                                        |
|   1 + 13 26 33 54 60 68 72 75 82 91 96    + Active     Long                        |
|     | 1  28 30 38 40 42 49 55 83 88 95    | Absorb     Permit                      |
|     | 17 25 36 37 62 79 87 94             | Warm                                   |
|     | 27 35 50 52 65 81 97                |                                        |
*   0 * 2  5  10 24 29 51 61 73 89          *                                        *
|     | 4  18 70 85                         | Assist                                 |
|     | 22 58 66                            | Express                                |
|     | 3  44 45 56                         | Humor      Wash                        |
|  -1 + 12                                  +                                        |
|     | 8  16 48                            | Remark                                 |
|     |                                     | Appear     Glory     Mystery    Reason  |
|     |                                     |                                        |
|  -2 + 80                                  + Adventure  Protect                     |
|     |                                     |                                        |
|     | 32                                  |                                        |
|     |                                     | Perform                                |
|  -3 +                                     +                                        |
|     |                                     |                                        |
|     |                                     |                                        |
|     |                                     |                                        |
|  -4 +                                     +                                        |
|     |                                     |                                        |
|     |                                     |                                        |
|     | 14                                  |                                        |
|  -5 +                                     + Teach                                  |
|     |                                     |                                        |
|     |                                     |                                        |
|     |                                     |                                        |
|  -6 +          Low Performing Participants  + Swim                      Easy Items |
|-----+-------------------------------------+----------------------------------------|
|Measr|+child                               |-Items                                  |
+------------------------------------------------------------------------------------+
```

Figure 6. Wright Map for Derivation Task

task have an even distribution of item difficulty (see figures 4 and 6). Based on the even spread, items do not need to be added or removed.

As with the Decomposition task, the range of item difficulty and the appropriateness of the Derivation items were further examined through the infit and outfit statistics of the Rasch analyses. Based on the infit and outfit statistics, there was only one outlying item: item 17. Item 17 has a root word frequency of 70 and involves both an orthographic and phonological shift. This item is the only item located at 0.9 logits which is close to other items within the task and may allow for this item to be removed after further research with a more comprehensive sample. Additionally, item 2 needs to be further examined as it had inconsistent responses from participants. Similarly to the Decomposition task, it is unadvisable to remove any items at this time until further investigation of the construction of the Derivation task occurs with a more diverse sample.

Impact of the Discontinue Rule

Researchers have frequently modified the TMS in an effort to shorten its administration time (e.g., Apel & Thomas-Tate , 2009; Berninger et al., 2010; Carlisle & Fleming, 2003; Deacon et al., 2015; Spencer et al., 2015; To et al., 2014) with the most common modification being the use of a discontinue rule . The most frequently employed of these discontinue rules is that the administration of the task is discontinued when a student makes six consecutive incorrect responses (e.g., Fracasso et al., 2014; Herman et al., 2013; Tighe & Binder, 2013). In an effort to examine the impact of this discontinue rule, differences between the scores produced by the three item orders examined in this study with the applied discontinue rule and the scores produced by all items was inspected. Analyses were also conducted related to whether applying the discontinue rule would in fact decrease administration time as desired by researchers. Within the remainder of this document, any time the item orders are referred to, the discontinue rule of six consecutive missed items has been applied.

Interestingly, despite wide use of the discontinue rule, results of the current study indicated that prior researchers may not have substantially reduced the number of administered items for the Decomposition task. The large majority of the sample (97.92% = original and theoretical item orders; 88.54% = data-based item order) did not miss enough consecutive items to meet the discontinue rule for any of the three item orders for the Decomposition task. Given students' performance on the items, it is not surprising that for the Decomposition task there were no differences in either scores or calculated administration time between the three item orders or between any item order and all items administered.

For the Derivation task, the score produced with the discontinue rule applied regardless of item order differed from the scores produced by administering all items, but did not differ from one another. Differences between the item order scores and the scores produced from all items suggests that information may be lost when applying the discontinue rule. However, the time to administer the Derivation task was altered. All item order variations of the Derivation task with the discontinue rule applied were significantly shorter than administering all items, with the data-based item order having the shortest administration time. In fact, between 12.5% (original item order) and 42.71% (data-based item order) of participants met the discontinue rule for the Derivation task.

The data-based item order resulted in the greatest decrease in administration time. This finding makes sense given that the data-based item order places items in order of difficulty (least to most). Most measures that utilize a discontinue rule (e.g., standardized norm-references intelligence and achievement assessments) order items from least to most difficulty. Discontinue rules can be applied to these tasks because, if a student is unable to correctly answer a series of consecutive items, it is unlikely that the student will answer the more difficult items that follow. Not needing to administer the more difficult items decreases the time spent assessing a student. If items are not ordered from least to most difficult, a discontinue rule would result in the loss of information about the student's abilities and skills as items would not be administered that the student could potentially answer correctly. Thus, prior researchers that have used the original item order of the TMS tasks may have lost information about their participants' abilities when applying a discontinue rule of six consecutive items (e.g., Carlisle & Nomanbhoy, 1993; Fracasso et al., 2014; Herman et al., 2013; Tighe & Binder, 2013).

Although researchers may have lost important information regarding students by employing a discontinue rule, they unfortunately failed to truly achieve their goal of decreasing administration time. Applying the discontinue rule to any of the item orders for the Decomposition did not result in a significant decrease in time. Although the time was decreased significantly for the Derivation task, an effect size of $d = 0.037$ indicates that the time saved was not meaningful. In fact, it only saved 24.58 sec for the Derivation task. Based upon these findings, it is unadvisable for researchers to employ a discontinue rule to the TMS tasks as it may likely result in a loss of meaningful data and not aid researchers in saving administration time. If information is lost by utilizing a discontinue rule, then it is irresponsible for researchers and educators to use the TMS tasks with an applied discontinue rule.

Reliability

The Rasch analyses inform both task construction and reliability. In addition to examining Rasch analyses of person and item separation reliability (discussed in the Construction of the TMS Task section), internal consistency of the TMS tasks were examined using Cronbach's alpha. Results indicated that both the Decomposition ($\alpha = .92$) and Derivation ($\alpha = .95$) tasks had high internal consistency. Therefore, all items within the TMS tasks formed reliable measures of MA.

Validity

In terms of validity, results indicate that both the Derivation and Decomposition tasks have adequate construct validity with analyses of unidimensionality suggesting that the TMS tasks are strong measures of MA. Similarly, results indicated that the TMS tasks have adequate convergent validity as the observed relationships between MA and other literacy skills were consistent with the majority of previous research using the TMS. Previous literature demonstrated that MA has a low to moderate relationship with phonological awareness (PA; Carlisle, 2000; Carlisle & Nomanbhoy, 1993; Deacon et al., 2013; Kieffer & Lesaux, 2008; Kirby et al., 2011; Nagy et al., 2006) while having a moderate relationship with reading comprehension (Carlisle, 2000; Deacon & Kirby, 2004; Kieffer & Lesaux, 2008; Kirby et al., 2011; Nagy et al., 2003; Shankweiler et al., 1995). Results from the current study replicate these findings with significant relationships observed between TMS scores and measures of PA and reading comprehension. Furthermore, analyses revealed that the predicted relationship between each TMS task with other literacy skills (i.e., phonological awareness and reading comprehension) was significant regardless of item order. Given these results, the constructs measured by the TMS tasks do seem to be meaningful at both tasks have strong unidimensionality and demonstrates a relationship to student reading achievement as seen in prior research. Although there is room to improve upon the TMS tasks by adding other components of MA, the relationships seen between MA, PA and reading comprehension remained significant and comparable to previous research even with the application of a discontinue rule and the ceiling effects observed within the current sample.

Impact on Research

Construction of tasks has a significant impact on research findings and the implications of those findings, particularly when examining relationships between constructs and when comparing populations (i.e., gender differences, differences between racial groups, differences between native language speakers and non-native language speakers, etc.). Thus, it is imperative that researchers first critically examine the measures they plan to use for collecting data before embarking on projects that will employ those measures. The *Test Standards* outline the important components for all measures used in psychological and educational settings (AERA, APA, & NCME, 2014). The standards discuss three foundational areas (i.e., validity, reliability, and fairness) and require that test developers provide evidence that all three areas are assessed fully (AERA, APA, & NCME, 2014). It is evident that the TMS tasks have insufficient evidence of validity, reliability, and fairness. A further concern is that many researchers have adapted or modified the TMS tasks without considering the impact of those changes on the resultant scores. The present study is the first to reanalyze the TMS following the outline provided by the *Test Standards*.

The Decomposition and Derivation tasks explained different amounts of variance within reading comprehension, depending on which item order examined in this study was used. For example, after controlling for PA, MA explained 13.9% of the variance within reading comprehension when all Decomposition items were utilized and 17% when all Deviation items were used. Arguably, including all items in a total score would be the closest method to capturing an individual's MA ability using the TMS rather than utilizing a task with the applied discontinue rule. Although using all items was how the TMS was designed to be used, many researchers adapted it by using a discontinue rule without considering that items were not arranged from least to most difficult; thus, imposing a discontinue rules was unadvisable. When the original item order with the applied discontinue rule was used within the analyses, 12.3% of the variance within reading comprehension was explained by the Decomposition task and 14.8% of the variance was explained by the Derivation task. Within the Decomposition task, the data-based item order produced results most similar to all items administered with 13.8% of the variance explained. Interestingly, the theoretical item order produced results most similar to using all items for the Derivation task with 17.4% of the variance explained. These results do not support the hypothesis that the data-based item order would best approximate the variance explained when all items are used. This may be due to an inappropriate discontinue rule or to the Derivation task's construction.

Results from the current study suggest that prior research (e.g., Carlisle & Nomanbhoy, 1993; Fracasso et al., 2014; Herman et al., 2013; Tighe & Binder, 2013) using the original item order with the applied discontinue rule may not provide as accurate a picture of an individual's MA ability as researchers believe. The present sample is fairly homogeneous in terms of age, racial background, and skill ability and results demonstrated variations in outcomes across item orders. Depending on the samples of prior research (e.g., Carlisle & Nomanbhoy, 1993; Fracasso et al., 2014; Herman et al., 2013; Kieffer & Lesaux, 2008; McCutchen et al., 2008; Tighe & Binder, 2013; Tong et al., 2011), results may have yielded significantly different results than they would have if a discontinue rule had not been applied.

Fairness

Fairness is the last area in which MA measures need to be assessed. However, a comprehensive assessment is beyond the scope of the present study. This study initiates the revisions of a MA assessment in order to make it more comprehensive and fair with continued adherence to strong reliability and validity. Within fairness, reliability and validity are connected to the application and consequences of a measure (AERA, APA, & NCME, 2014). Results from the present study suggest that both the Decomposition and Derivation tasks have adequate internal consistency.

Similarly, both TMS tasks have adequate construct and convergent validity. Even though analyses from the current study suggest that the TMS tasks are measuring MA, researchers must consider whether the construct being measured is MA entirely or only a few of the many components of MA. Ensuring that a measure is representing a construct fully (e.g., when an assessment includes all important aspects of a construct) while balancing construct-irrelevant variance (i.e., a measure being too broad and having extraneous aspects impacting the difficulty; Messick, 1995) is the most difficult aspect of measurement. Collecting this evidence is time intensive and difficult to accomplish, but is vital when creating and examining a measure. This is tied to fairness in testing (AERA, APA, & NCME, 2014; Messick, 1995) as it is irresponsible to use a measure to inform educational decisions without determining whether that measure will yield information that is reliable and valid. As discussed in Chapter 2, MA has several integral aspects such as the semantic and phonological knowledge, and constituents within the construct of MA. These other constituents include morpheme type morpheme type (e.g., inflectional, derivational, etc.), levels of transparency (e.g., phonological transparent, phonologically opaque, etc.), and word frequency.

Once validity concerns are assessed and resolved for the TMS tasks, other aspects of fairness will need to become the primary focus. Extensive examination of the TMS tasks with varying populations is necessary to understand how the TMS tasks function for all groups (i.e., racial and ethnic, age, socio-economic, native language, etc.). It is important to examine how scores vary between groups of individuals as these scores may be used to make diagnoses and educational decisions.

Limitations and Future Directions

One limitation of the current study is the minimal diversity within the participant sample. Participants had a mean age of 10.31 years (range: 8.03 to 11.67 years) and were primarily White/Caucasian (83.3%). Approximately 68.75% of the participants received general education support only and 15.6% received special education support. In addition, 16.67% of the participants qualified for gifted services with 4.17% of participants being twice exceptional (i.e., receiving gifted and special education services). According to data calculated from statistics presented in the 2011 US Census (Davis & Bauman, 2013) and by the National Association for Gifted Children ("Gifted Education in the U.S."), 4.84% of students in 2011 were enrolled in gifted programs. The participant sample has a gifted rate approximately 3.4 times higher than the national rate. Overall, the participant sample in the present study was limited in diversity, age, performed higher than the typical student population, and all received the same instruction (i.e., the students came from one school district). The limited diversity and the instructional backgrounds of participants may have affected results in unknown ways. Therefore, similar analyses need to be conducted with larger samples of increased diversity, age range, and varying educational and instructional backgrounds.

Another limitation of the present study is that the three-item orders were not given to three separate groups of individuals to examine the effect of the item order on administration. In order to determine the data-based item order, testing needed to occur prior to administering the TMS tasks in each of its item order to three separate groups. The present study has served as a determination of a potential data-based item order; however, as previously discussed more testing with more diverse samples, including individuals with varying educational and instructional backgrounds, needs to occur. Once a more concrete data-based item order has been determined, a study examining the effect of item order on administration of the TMS tasks should be conducted in order to determine whether the TMS tasks function as researchers would predict.

Once larger samples are collected, several additional aspects to the measures need to be examined, including the reassessment of any changes or edits to the TMS tasks. First, aspects to test construction need to be reexamined, such as the spread of items, to determine whether there are gaps within the measures that require additional items to be created. It is likely that different demographic factors may impact the spread of items and may highlight problems with aspects of fairness of the TMS tasks, including concerns regarding reliability and validity.

Another aspect of the tasks to examine further is unidimensionality, as it is unclear based upon prior research (i.e., Apel, 2014) whether the TMS tasks and other MA measures are assessing the same construct or different aspects of MA. As the TMS tasks are currently designed, data indicates that there may be two sub-constructs being measured under the umbrella of the MA construct. To examine this, a factor analysis could be conducted. Additionally, it may be that there are sub-constructs of MA that are left out of the TMS. For example, the tasks do not include items with prefixes. Researchers will need to consider whether or not to add tasks to the TMS in an effort to measure aspects of MA not currently measured.

If researchers still believe that the TMS tasks need to utilize a discontinue rule, an appropriate data-based item order should be determined using a larger and more diverse sample. A data-based item order is arguably the most accurate within any sample as it is established using data collected. However, several additional studies are needed before a data-based item order for the TMS tasks can be applied. Researchers cannot use the current study's item order due to the participant sample lacking multiple dimensions of diversity (i.e., age, race, ethnicity, etc.). Understanding the item order based on data will also allow researchers to further understand the developmental progress of MA.

The proposed future research in conjunction with the present study will make a significant contribution to MA research in regard to informing reading and language arts instruction and providing a stronger foundation in understanding the development of MA. Once MA is better understood, researchers and educators may determine that an instructional focus on MA may benefit the development of other literacy skills, particularly if it is determined to be a "building block" skill (e.g., PA or MA). By understanding how MA relates to other literacy skills, educators can design curriculum and lesson plans according to the importance of these various skills.

REFERENCES

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. MIT press.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Anglin, J. M. (1993). Vocabulary development: A morphological analysis. *Monographs of the Society for Research in Child Development, 58*(10), v-165.

Apel, K. (2014). A comprehensive definition of morphological awareness: Implications for assessment. *Topics in Language Disorders*, *34*(3), 197-209.

Apel, K., & Lawrence, J. (2011). Contributions of morphological awareness skills to word-level reading and spelling in first-grade children with and without speech sound disorder. *Journal of Speech, Language, and Hearing Research*, *54*(5), 1312-1327.

Apel, K., & Thomas-Tate, S. (2009). Morphological awareness skills of fourth-grade African American students. *Language, Speech, and Hearing Services in Schools*, *40*(3), 312-324.

Apel, K., Dieham, E., & Apel, L. (2013). Using multiple measures of morphological awareness to assess its relation to reading. *Topics in Language Disorders*, *33*(1), 42-56.

Berko, J. (1958). The child's learning of English morphology. *Word*, *14*(2-3), 150-177.

Berninger, V., Abbott, R. D., Nagy, W., & Carlisle, J. (2010). Growth in phonological, orthographic, and morphological awareness in grades 1 to 6.*Journal of psycholinguistic research*, *39*(2), 141-163.

Berninger, V., Abbott, R., Billingsley, F., & Nagy, W. (2001). Processes underlying timing and frequency: Efficiency, automaticity, coordination, and morphological awareness. In M. Wolf (Ed.), *Dyslexia, fluency, and the brain* (pp. 383-414). Baltimore: York Press.

Berninger, V., & Nagy, W. (1999). *University of Washington Morphological Awareness Battery*. Unpublished experimental test battery, Seattle, WA.

Bond, T.G., & Fox, C.M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences, Third Edition*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Bowers, P. N., & Kirby, J. R. (2010). Effects of morphological instruction on vocabulary acquisition. *Reading and Writing, 23*(5), 515-537. doi:10.1007/s11145-009-9172-z

Bowers, P. N., Kirby, J. R., & Deacon, S. H. (2010). The effects of morphological instruction on literacy skills: A systematic review of the literature. *Review of Educational Research, 80*(2), 144-179. doi: 10.3102/0034654309359353

Carlisle, J. F. (1984). *The relationship between knowledge of derivational morphology and spelling ability in fourth, sixth and eighth graders* (Doctoral dissertation, University of Connecticut).

Carlisle, J. F. (1988). Knowledge of derivational morphology and spelling ability in fourth, sixth, and eighth graders. *Applied Psycholinguistics, 9*, 247-266.

Carlisle, J. F. (1995). Morphological awareness and early reading achievement. In L. B. Feldman (Ed.), *Morphological aspects of language processing.* (pp. 189-209). Hillsdale, NJ England: Lawrence Erlbaum Associates, Inc.

Carlisle, J. F. (2000). Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading and Writing: An Interdisciplinary Journal, 12*(3-4), 169-190.

Carlisle, J. F. (2003). Morphology matters in learning to read: A commentary. *Reading Psychology, 24*(3-4), 291-322. doi:10.1080/02702710390227369

Carlisle, J. F. (2012, July). *Discussant*. Discussion presented at Society for the scientific study of reading, Montreal, Quebec.

Carlisle, J. F., & Fleming, J. (2003). Lexical processing of morphologically complex words in the elementary years. *Scientific Studies of Reading, 7*(3), 239-253. doi: 10.1207/S1532799XSSR0703_3

Carlisle, J. F., & Katz, L. A. (2006). Effects of word and morpheme familiarity on reading of derived words. *Reading and Writing*, *19*(7), 669-693.

Carlisle, J. F., & Nomanbhoy, D. M. (1993). Phonological and morphological awareness in first graders. *Applied Psycholinguistics, 14*(2), 177-195. doi: 10.1017/S0142716400009541

Carlisle, J. F., & Stone, C. (2005). Exploring the role of morphemes in word reading. *Reading Research Quarterly*, *40*(4), 428-449.

Carroll, J. B., Davies, P., Richman, B., & Davies, P. (1971). *The American Heritage word frequency book* (pp. xxi-xl). Boston: Houghton Mifflin.

Catts, H. W., Petscher, Y., Schatschneider, C., Bridges, M. S., & Mendoza, K. (2008). Floor effects associated with universal screening and their impact on the early identification of reading disabilities. *Journal of Learning Disabilities*.

Cazden, C. B. (1968). The acquisition of noun and verb inflections. *Child development*, 433-448.

Clark, E., & Cohen, S. (1984). Productivity and memory for newly formed words. Journal of Child Language, 11, 611–625.

Cook, C. R., Volpe, R. J., & Livanis, A. (2010). Constructing a roadmap for future universal screening research beyond academics. *Assessment for Effective Intervention*, *35*(4), 197-205.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. J*ournal of Applied Psychology, 78*, 98-104.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* New York: Holt, Rinehart and Winston.

Davis, J., & Bauman, K. (2013, September). School Enrollment in the United States: 2011. Retrieved from https://www.census.gov/prod/2013pubs/p20-571.pdf

Deacon, S. H., & Kirby, J. R. (2004). Morphological awareness: Just "more phonological"? The roles of morphological and phonological awareness in reading development. A*pplied Psycholinguistics*, *25*, 223-238. doi: 10.1017.S0124716404001117

Deacon, S. H., Benere, J., & Pasquarella, A. (2013). Reciprocal relationship: Children's morphological awareness and their reading accuracy across grades 2 to 3. *Developmental Psychology*, *49*(6), 1113.

Deacon, S. H., Tong, X., & Francis, K. (2015). The relationship of morphological analysis and morphological decoding to reading comprehension. *Journal of Research in Reading*.

Derwing, B. L., & Baker, W. J. (1979). Recent research on the acquisition of English morphology. *Language acquisition*, 209-223.

Dietrich, J. A., & Brady, S. A. (2001). Phonological representations of adult poor readers: An investigation of specificity and stability. *Applied Psycholinguistics, 22*(3), 383-418. doi: 10.1017/S014271640100306X

Eklund, K., Renshaw, T. L., Dowdy, E., Jimerson, S. R., Hart, S. R., Jones, C. N., & Earhart, J. (2009). Early identification of behavioral and emotional problems in youth: Universal screening versus teacher-referral identification. *The California School Psychologist*, *14*(1), 89-95.

Every Student Succeeds Act (ESSA) of 2015, Pub. L. No. 114–95

Fowler, A., & Liberman, I. Y. (1995). The role of phonology and orthography in morphological awareness. In L. Feldman (Ed.), Morphological aspects of language processing (pp. 189–209). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Fowler, A., Feldman, L. B., Andjelkovic, D.,& Oney, B. (2003).Morphological and phonological analysis by beginning readers: Evidence from Serbian and Turkish.

Fracasso, L. E., Bangs, K. E., & Binder, K. S. (2014). The contributions of phonological and morphological awareness to literacy skills in the adult basic education population. *Journal of learning disabilities*, 0022219414538513.

Gifted Education in the U.S. (n.d.). Retrieved May 12, 2018, from https://www.nagc.org/resources-publications/resources/gifted-education-us

Gordon, P. (1989). Levels of affixation in the acquisition of English morphology. Journal of Memory and Language, 28, 519–530.

Herman, J., Cote, N. G., Reilly, L., & Binder, K. S. (2013). Literacy skill differences between adult native English and native Spanish speakers. *Journal of research and practice for adult literacy, secondary, and basic education*, *2*(3), 147.

Hosp, J. L., Hosp, M. A., & Dole, J. K. (2011). Potential bias in predictive validity of universal screening measures across disaggregation subgroups. *School Psychology Review*, *40*(1), 108.

Individuals with Disabilities Education Improvement Act of 2004, Pub. L. No. 108-446.

Jacob, S., Decker, D. M., & Hartshorne, T. S. (2010). *Ethics and law for school psychologists*. John Wiley & Sons.

Jannedy, S., Poletto, R., & Weldon, T. L. (Eds.). (1994). *Language Files: Materials for an Introduction to Language & Linguistics*. Ohio State University Press.

Jarmulowicz, L., Hay, S. E., Taran, V. L., & Ethington, C. A. (2008). Fitting derivational morphophonology into a developmental model of reading. *Reading and Writing, 21*(3), 275-297. doi: 10.1007/s11145-007-9073-y

Jones, N. K. (1983). Development of morphophonemic segments in children's mental representations of words. *Applied Psycholinguistics*, *12*(02), 217-239.

Joseph, L. M. (2006). *Understanding, assessing, and intervening on reading problems: A guide for school psychologists and other educational consultants*. National Association of School Psychologists.

Kieffer, M. J., & Lesaux, N. K. (2008). The role of derivational morphology in the reading comprehension of Spanish-speaking English language learners. *Reading and Writing*, *21*(8), 783-804.

Kirby, J. R., Deacon, S. H., Bowers, P. N., Izenberg, L., Wade-Woolley, L., & Parrila, R. (2011). Children's morphological awareness and reading ability. *Reading and Writing*, *25*(2), 389-410.

Kirby, J. R., Parrila, R. K., & Pfeiffer, S. L. (2003). Naming speed and phonological awareness as predictors of reading development. *Journal of Educational Psychology*, *95*(3), 453.

Ku, Y., & Anderson, R. C. (2003). Development of morphological awareness in chinese and english. *Reading and Writing, 16*(5), 399-422. doi: 10.1023/A:1024227231216

Kuczaj, S. A. (1977). The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior*, *16*(5), 589-600.

Kuczaj, S. A. (1978). Children's judgments of grammatical and ungrammatical irregular past-tense verbs. *Child Development*, 319-326.

Kuo, L., & Anderson, R. C. (2006). Morphological awareness and learning to read: A cross-language perspective. *Educational Psychologist, 41*(3), 161-180. doi: 10.1207/s15326985ep4103_3

Lieven, E. (2010). Input and first language acquisition: Evaluating the role of frequency. *Lingua*, *120*(11), 2546-2556.

Linacre J.M. (2006). Data Variance Explained by Measures, *Rasch Measurement Transactions, 20:1* p. 1045.

Mahony, D. L. (1994). Using sensitivity to word structure to explain variance in high school and college level reading ability. *Reading and Writing*, *6*(1), 19-44.

Mahony, M., Singson, M., & Mann, V.. (2000). Reading ability and sensitivity to morphological relations. *Reading and Writing: An Interdisciplinary Journal, 12*(3-4), 191-218.

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., Xu, F., & Clahsen, H. (1992). Overregularization in language acquisition. *Monographs of the society for research in child development*, i-178.

McCutchen, D., Green, L., & Abbott, R. D. (2008). Children's morphological knowledge: Links to literacy. *Reading Psychology*, *29*(4), 289-314.

McDonald, M. (2013, October). The Role of the School Psychologist in RTI: Linking Assessment to Intervention. *Communiqué*, 42(2). Retrieved June 22, 2016, from http://www.nasponline.org/publications/periodicals/communique/issues/volume-42-issue-2/the-role-of-the-school-psychologist-in-rti-linking-assessment-to-intervention

McDougal, J., LeBlanc, M., & Hintze, J. (2010). Graphing Student Academic Data for Universal Screening and Progress Monitoring. In A. Canter, L. Paige, & S. Shaw (Eds.), *Helping Children at Home and School: Handouts for Families and Educators III*.

Mesmer, H. A. E., & Griffith, P. L. (2005). Everybody's Selling It: But Just What Is Explicit, Systematic Phonics Instruction?. *The Reading Teacher*, 366-376.

Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741-749.

Model for Comprehensive and Integrated School Psychological Services: NASP Practice Model Overview. (2010, December). *Communiqué*, 39(4). Retrieved from file:///C:/Users/Frank/Downloads/V39N4_NASPPracticeModel.pdf.

Mukaka, M. M. (2012). A guide to appropriate use of Correlation coefficient in medical research. *Malawi Medical Journal*, *24*(3), 69-71.

Nagy, W., Anderson, R. C., Schommer, M., Scott, J. A., & Stallman, A. C. (1989). Morphological families in the internal lexicon. *Reading Research Quarterly*, 262-282.

Nagy, W., Berninger, V. W., & Abbott, R. D. (2006). Contributions of morphology beyond phonology to literacy outcomes of upper elementary and middle-school students. *Journal of Educational Psychology*, *98*(1), 134-147.

Nagy, W., Berninger, V., Abbott, R., Vaughan, K., & Vermeulen, K. (2003). Relationship of morphology and other language skills to literacy skills in at-risk second-grade readers and at-risk fourth-grade writers. *Journal of Educational Psychology, 95*, 730-742.

Nagy, W., Diakidoy, I. A., & Anderson, R. (1993). The acquisition of morphology: Learning the contribution of suffixes to the meanings of derivatives. *Journal of Literacy Research*, *25*(2), 155-170.

No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 115.

Nunes, T., Bryant, P., & Bindman, M. (1997). Morphological spelling strategies: developmental stages and processes. *Developmental psychology*, *33*(4), 637.

Nunes, T., Bryant, P., & Olsson, J. (2003). Learning morphological and phonological spelling rules: An intervention study. *Scientific Studies of Reading*, *7*(3), 289-307.

Petscher, Y., Kim, Y., & Foorman, B. (2011). The Importance of Predictive Power in Early Screening Assessments: Implications for Placement in the Response to Intervention Framework. *Assessment for Effective Intervention*. doi:10.1177/1534508410396698

Rayner, K. & Pollatsek, A. (1989). The psychology of reading. Englewood Cliffs, NJ: Prentice Hall.

Reschly, D.J., & Bergstrom, M.K. (2009). Response to Intervention. In T.B. Gutkin and C.R. Reynolds (Eds). The Handbook of School Psychology, 4th Ed. Wiley

Sandra, D. (1994). The morphology of the mental lexicon: Internal word structure viewed from a psycholinguistic perspective. *Language and cognitive processes*, *9*(3), 227-269.

Sattler, J. M. (2008). *Assessment of children: Cognitive foundations*. San Diego, CA: JM Sattler.

Shankweiler, D., Crain, S., Katz, L., Fowler, A. E., Liberman, A. M., Brady, S. A., Shaywitz, B. A. (1995). Cognitive profiles of reading-disabled children: Comparison of language skills in phonology, morphology, and syntax. *Psychological Science, 6*(3), 149-156. doi: 10.1111/j.1467-9280.1995.tb00324.x

Sharpe, C. (2012, April 9). Elementary Assessments: Universal Screening, Diagnostic, and Progress Monitoring. Retrieved July 28, 2016, from http://www.sde.ct.gov/sde/lib/sde/pdf/curriculum/cali/elementary_assessments_4-9-12.pdf

Singson, M., Mahony, D., & Mann, V. (2000). The relation between reading ability and morphological skills: Evidence from derivational suffixes. *Reading and Writing: An Interdisciplinary Journal, 12*, 219-252.

Speece, D. L., Schatschneider, C., Silverman, R., Case, L. P., Cooper, D. H., & Jacobs, D. M. (2011). Identification of reading problems in first grade within a response-to-intervention framework. *The Elementary school journal*,*111*(4), 585.

Stackhouse, J. (1990). Phonological deficits in developmental reading and spelling disorders. In P. Grunwell (Ed.), *Developmental speech disorders* (pp. 163–182). New York: Churchill Livingstone

Standards for Educational and Psychological Testing: Essential Guidance and Key Developments in a New Era of Testing. (2014, September 12). Retrieved September 12, 2014, from http://www.aera.net/Portals/38/docs/Outreach/Standards_Hill_Briefing_Slides_FINAL.pdf?timestamp=1410876719244

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly, 24*(4), 402-433. doi: 10.2307/747605

The Every Student Succeeds Act: Details of the New Law. (n.d.). Retrieved April 04, 2016, from https://www.nasponline.org/research-and-policy/current-law-and-policy-priorities/policy-priorities/the-every-student-succeeds-act/details-of-essa

Thorndike, R., & Hagen, E. (1961). *Measurement and evaluation in psychology and education*. New York: John Wiley and Sons.

Tong, X., Deacon, S. H., Kirby, J., Cain, K., & Parrila, R. (2011). Morphological awareness: A key to understanding poor reading comprehension in English. *Journal of Educational Psychology, 103*, 523-534. doi: 10.1037/a0023495

Tyler, A., & Nagy, W. (1989). The acquisition of English derivational morphology. *Journal of memory and language*, *28*(6), 649-667.

van den Broek, P., & Espin, C. A. (2012). Connecting cognitive theory and assessment: Measuring individual differences in reading comprehension. *School Psychology Review*, *41*(3), 315-326.

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological bulletin*, *101*(2), 192.

Wagner, R. K., Torgesen, J. K., Laughon, P., Simmons, K., & Rashotte, C. A. (1993). Development of young readers' phonological processing abilities. *Journal of Educational Psychology*, *85*(1), 83.

White, T. G., Sowell, J., & Yanagihara, A. (1989). Teaching elementary students to use word-part clues. *The Reading Teacher*, *42*, 302-309.

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. New York, NY: Psychology Press: Taylor & Francis Group.

Wolter, J. A., & Pike, K. (2015). Dynamic Assessment of Morphological Awareness and Third-Grade Literacy Success. *Language, speech, and hearing services in schools*, *46*(2), 112-126.

Wolter, J. A., Wood, A., & D'zatko, K. W. (2009). The influence of morphological awareness on the literacy development of first-grade children. *Language, Speech, and Hearing Services in Schools*, *40*(3), 286-298.

Wysocki, K., & Jenkins, J. R. (1987). Deriving word meanings through morphological generalization. *Reading Research Quarterly*, 66-81.

Appendix A

Figure from Berko (1958)

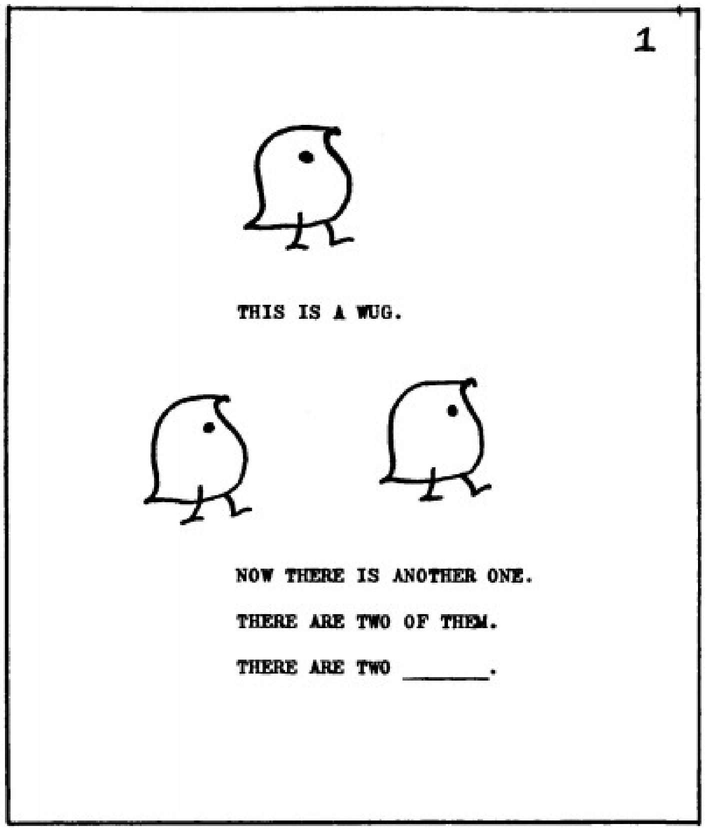

Figure 1. The plural allomorph in /-z/.

Appendix B

Carlisle's Measurement Modifications and Adaptations

Study	**Measure Name**	**# of Subtests**	**# of items**	**Presentation**	**Morpheme Types**	**Transparency Levels**	**Word Frequency**
Carlisle (1984)	Test of Morphological Structure	1) Derived Forms 2) Base Forms	1) 40 items 2) 40 items	Oral	Derivation	None Phonological Orthographic Both	Used words with an SFI of 40 or above
Carlisle (1987)	Test of Morphological Structure	1) Derived Forms 2) Base Forms	1) 40 items 2) 40 items	Oral	Derivation	None Phonological Orthographic Both	Used words with an SFI of 40 or above
Carlisle (1988)	The Morphology Test	1) Derived Forms 2) Base Forms	1) 40 items 2) 40 items	Oral	Derivation	None Phonological Orthographic Both	Used words with an SFI of 40 or above
Carlisle & Nomanbhoy (1993)	Production of Word Forms Test	1) Production	4 practice & 15 test items	Oral	Inflection (5 items) Derivation (10 items)	None Phonological Orthographic	*No information provided*
Carlisle (1995)	The Morphological Production Task	1) Production	*No information provided*	*No information provided*	Inflection Derivation	None Phonological	*No information provided*
Carlisle (2000)	Test of Morphological Structure	1) Decomposition 2) Derivation	1) 28 items 2) 28 items	*No information provided*	Derivation	None Phonological Orthographic Both	Decomposition: Base Forms SFI M = 56.7 Derived Forms SFI M = 50.2 Derivation: Base Forms SFI M =

							55.8 Derived Forms SFI *M* = 51.0
Carlisle & Fleming (2003)	Test of Morphological Structure	1) Decomposition 2) Production	1) 28 items 2) 28 items	*See Carlisle (2000)*	Derivation	None Phonological Orthographic Both	*See Carlisle (2000)*
Berninger, Abbott, Nagy, & Carlisle (2010)	Test of Morphological Structure	1) Decomposition 2) Derivation	1) 28 items 2) 28 items	*See Carlisle (2000)*	Derivation	*See Carlisle (2000)*	*See Carlisle (2000)*

Appendix C

Morphological Awareness Task Directions

Decomposition Task

You must score the participant's answers while they are doing the task. Give them a score of 1 for a correct answer and a 0 for an incorrect answer for each item, i.e.:

0 = incorrect 1 = correct

Discontinue the test when the participant has answered a total of 6 items incorrectly.
Note: You <u>shouldn't</u> read the item numbers.

"This next test is just like the one you just did. I'm going to read you a word, and then I'm going to read you a sentence that is not complete. Think about that word I read to you. Is there a form of that word that would fit in the blank so that the sentence makes sense? If you would like me to repeat the word or sentence, tell me and I can do that for you. Let's try a couple of practice items first."

a. **Driver. Children are too young to _____.** [drive]
b. **Improvement. My teacher wants my spelling to _____.** [improve]

"Ready? Great! Let's get started." Remember to say "blank" where the spaces are.

1. **Growth. She wanted her plant to _____.** [grow]
2. **Dryer. Put the wash out to _____.** [dry]
3. **Variable. The time of his arrival did not _____.** [vary]
4. **Width. The mouth of the river is very _____.** [wide]
5. **Density. The smoke in the room was very _____.** [dense]
6. **Discussion. The friends have a lot to _____.** [discuss]
7. **Famous. The actor would achieve much _____.** [fame]
8. **Description. The picture is hard to _____.** [describe]
9. **Fifth. The boy counted from one to _____.** [five]
10. **Election. Which person did they _____?** [elect]
11. **Strength. The girl was very _____.** [strong]
12. **Decision. The boy found it hard to _____.** [decide]
13. **Popularity. The girl wants to be _____.** [popular]
14. **Runner. How fast can she _____?** [run]
15. **Publicity. His views were made _____.** [public]
16. **Difference. Do their opinions _____?** [differ]
17. **Originality. That painting is very _____.** [original]
18. **Agreeable. With that statement I could not _____.** [agree]
19. **Courageous. The man showed great _____.** [courage]
20. **Admission. How many people will they _____?** [admit]
21. **Dangerous. Are the children in any _____?** [danger]
22. **Reduction. The overweight man was trying to _____.** [reduce]

23. **Baker. She put the bread in to _____.** [bake]
24. **Division. The cake is hard to _____.** [divide]
25. **Guidance. The map was her _____.** [guide]
26. **Continuous. How long will the storm _____?** [continue]
27. **Reliable. On his friend he could always _____.** [rely]
28. **Acceptance. Is that an offer you can _____?** [accept]

Appendix D

Morphological Awareness Task Directions

Derivation Task

You must score the participant's answers while they are doing the task. Give them a score of 1 for a correct answer and a 0 for an incorrect answer for each item, i.e.:

0 = incorrect 1 = correct

Discontinue the test when the participant has answered a total of 6 items incorrectly.
Note: You **shouldn't** read the item numbers.

"I'm going to read you a word, and then I'm going to read you a sentence that is not complete. Think about that word I read to you. Is there a form of that word that would fit in the blank so that the sentence makes sense? If you would like me to repeat the word or sentence, tell me and I can do that for you. Let's try a couple of practice items first."

a. **Farm. My uncle is a __blank___.** [farmer]
b. **Help. My sister is always __blank___.** [helpful or helping].

"Ready? Great! Let's get started." Remember to say "blank" where the spaces are.

1. **Warm. He chose the jacket for its _____.** [warmth]
2. **Teach. He was a very good _____.** [teacher]
3. **Permit.** (to allow) **Father refused to give _____.** [permission]
4. **Profit. Selling lemonade in summer is _____.** [profitable].
5. **Appear. He cared about his _____.** [appearance]
6. **Express. 'OK' is a common _____.** [expression]
7. **Four. The cyclist came in _____.** [fourth]
8. **Remark. The speed of the car was _____.** [remarkable]
9. **Protect. She wore glasses for _____.** [protection]
10. **Perform. Tonight is the last _____.** [performance]
11. **Expand. The company planned an _____.** [expansion]
12. **Revise. This paper is his second _____.** [revision]
13. **Reason. Her argument was quite _____.** [reasonable]
14. **Major. He won the vote by a _____.** [majority]
15. **Deep. The lake was well known for its _____.** [depth]
16. **Equal. Boys and girls are treated with _____.** [equality]
17. **Long. They measured the ladder's _____.** [length]
18. **Adventure. The trip sounded _____.** [adventurous]
19. **Absorb. She chose the sponge for its _____.** [absorption or absorbance]
20. **Active. He tired after so much _____.** [activity]
21. **Swim. She was a strong _____.** [swimmer]
22. **Human. The kind man was known for his _____.** [humanity]
23. **Wash. Put the laundry in the _____.** [washer]

24. **Humor. The story was quite _____.** [humorous]
25. **Assist. The teacher will give you _____.** [assistance]
26. **Mystery. The dark glasses made the man look _____.** [mysterious]
27. **Produce. The play was a grand _____.** [production]
28. **Glory. The view from the hill top was _____.** [glorious]

Appendix E

Decomposition Task Item Order

Original Item Number	Root Word	Complex Word	Root Word Frequency	Level of Transparency
1	Grow	Growth	64.1	None
2	Dry	Dryer	61.9	None
21	Danger	Dangerous	58.4	None
18	Agree	Agreeable	57.6	None
28	Accept	Acceptance	57.1	None
23	Bake	Baker	48.5	None
14	Run	Runner	64.3	Orthographic
26	Continue	Continuous	58.4	Orthographic
25	Guide	Guidance	56.9	Orthographic
3	Vary	Variable	55.7	Orthographic
5	Dense	Density	52.2	Orthographic
27	Rely	Reliable	51.9	Orthographic
7	Fame	Famous	50.4	Orthographic
15	Public	Publicity	62.9	Phonological
13	Popular	Popularity	58.4	Phonological
6	Discuss	Discussion	55.5	Phonological
16	Differ	Difference	54.6	Phonological
19	Courage	Courageous	53.4	Phonological
17	Origin	Originality	53.3	Phonological
10	Elect	Election	47.7	Phonological
9	Five	Fifth	64	Both
11	Strong	Strength	63.6	Both
4	Wide	Width	61.5	Both
12	Decide	Decision	59.9	Both
8	Describe	Description	59.1	Both
22	Reduce	Reduction	55.6	Both
24	Divide	Division	54.8	Both
20	Admit	Admission	53.2	Both

Appendix F

Derivation Task Item Order

Original Item Number	Root Word	Complex Word	Root Word Frequency	Level of Transparency
7	Four	Fourth	65.4	None
1	Warm	Warmth	62.7	None
13	Reason	Reasonable	62.6	None
5	Appear	Appearance	59.4	None
2	Teach	Teacher	56.8	None
21	Swim	Swimmer	56.6	None
10	Perform	Performance	56.2	None
23	Wash	Washer	56	None
4	Profit	Profitable	54	None
25	Assist	Assistance	50.4	None
24	Humor	Humorous	49.7	None
8	Remark	Remarkable	48.2	None
20	Active	Activity	57.5	Orthographic
18	Adventure	Adventurous	53.3	Orthographic
28	Glory	Glorious	50.4	Orthographic
22	Human	Humanity	63.9	Phonological
14	Major	Majority	62.4	Phonological
16	Equal	Equality	59.3	Phonological
9	Protect	Protection	58.8	Phonological
6	Express	Expression	56.5	Phonological
17	Long	Length	70	Both
15	Deep	Depth	62.4	Both
27	Produce	Production	62	Both
26	Mystery	Mysterious	54.8	Both
3	Permit	Permission	54.1	Both
11	Expand	Expansion	53.3	Both
19	Absorb	Absorption	51.5	Both
12	Revise	Revision	46.2	Both

Appendix G

Building Blocks for Variable Construct Map (Wilson, 2005)

Overview:

Building Blocks	**Questions**	**Answers**	**Morphological Awareness Production Task**
Latent Variable	What is the latent variable being measured?	Oral Morphological Awareness	The purpose of this task is to measure an individual's degree of oral morphological awareness.
Observational Design	What is the plan for collecting structured responses or observations from persons?	Items: sentence completion task with various stimulus words	Dichotomous ratings obtained by the administrator from the individual's responses.
Scoring Rules	How are responses or observations categorized to represent person levels on the latent variable?	Items scored as correct (x = 1) or incorrect (x = 0)	Items are marked as correct or incorrect until the individual incorrectly responds to 6 consecutive items
Measurement Model	How are the person and item responses or observations mapped onto the latent variable?	Rasch Model	Dichotomous Rasch Model

Latent Variable (What is the construct?):
The construct is morphological awareness which is the knowledge of and the ability to manipulate prefixes and suffixes (morphemes). It applies to oral and written knowledge but also building and decoding words.

Unidimensionality:
It is a unidimensional variable with various subcomponents (See Content & Distributional Components)

Content & Distributional Components:
Subcomponents of Morphological Awareness

- Transparency
    - No changes
    - Phonological shifts - a change in the sound of the word but not the spelling
    - Orthographic shifts - a change to the spelling of the word but not the sound
    - Both shifts
- Word Frequency
    - Root word v. complex
- Morpheme type
    - Root v. affix

  - Prefix, suffix, or both
  - Inflectional v. derivational
  - Compounds v. affix
- Item presentation
  - Oral, Visual, or Oral & Visual
- Question type
  - Relational, semantic, or distributional
- Response requirement
  - Oral v. written
  - Expressive v. receptive
  - Multiple choice, yes/no, etc.

Wright Map:

Logit Scale	**Persons/Items**	**Items**		
		Affix Type	**Transparency**	**Word Frequency**
5.00	High values on the latent variable/Hard Items	Derivational	Both Shifts	Infrequent
4.00				
3.00				
2.00			Orthographic Shift	
1.00	Midrange values on the latent variable/ Moderately difficult items			
0.00				
-1.00			Phonological Shift	
-2.00				
-3.00	Low values on the latent variable/Easy Items			
-4.00				
-5.00		Inflectional	No Shift	Frequent

Potential Variables That May Affect the Construct:

- SES
- Gender (girls tend to have more social interactions than boys which may influence their language ability)
- Native language
- Age

Appendix H

Scoring for Sample Participant

Participant #1			
Original Item Order		**Theoretical Item Order**	
Item Number	**Response**	**Item Number**	**Response**
1	1	7	0
2	1	1	1
3	1	13	1
4	0	5	1
5	1	2	1
6	1	21	0
7	0	10	1
8	1	23	0
9	1	4	0
10	1	25	1
11	0	24	0
12	1	8	1
13	1	20	0
14	1	18	1
15	1	28	1
16	0	22	0
17	1	14	1
18	1	16	0
19	0	9	1
20	0	6	1
21	0	17	1
22	0	15	1
23	0	27	1
24	0	26	1
25	1	3	1
26	1	11	0
27	1	19	0
28	1	12	1
Total Score	***14***	***Total Score***	***18***

Note: 0 = incorrect; 1 = correct; gray = item not included in the score

Appendix I

Procedural Integrity Checklist

PID: _________

Researcher Conducting IRR: _________

Task Order

Counter Balance Set: _________

Were the tasks administered in the correct order (circle one)?....................Yes No

Passage Comprehension

Researcher read all instructions (circle one)?.. Yes No

Researcher began with the sample item (circle one)?.................................. Yes No

Researcher gave corrective feedback when needed (circle one)?................ Yes No

Researcher established a basal (circle one)?...Yes No

Researcher established a ceiling (circle one)?.. Yes No

Researcher scored the task correctly (circle one)?.......................................Yes No

Researcher recorded all item scores accurately (circle one)?........................Yes No

Word Attack

Researcher read all instructions (circle one)?... Yes No

Researcher began with the sample item (circle one)?.................................. Yes No

Researcher gave corrective feedback when needed (circle one)?................ Yes No

Researcher established a basal (circle one)?...Yes No

Researcher established a ceiling (circle one)?.. Yes No

Researcher scored the task correctly (circle one)?.......................................Yes No

Researcher recorded all item scores accurately (circle one)?........................Yes No

Morphological Awareness Task: Decomposition

Researcher read all instructions (circle one)?... Yes No

Researcher began with the sample items (circle one)?.................................Yes No

Researcher gave corrective feedback when needed (circle one)?................ Yes No

Researcher administered all items (circle one)?... Yes No

Researcher administered items in presented order (circle one)?.................. Yes No

Researcher timed the task correctly (circle one)?.. Yes No

Researcher recorded all item scores accurately (circle one)?........................ Yes No

Morphological Awareness Task: Derivation

Researcher read all instructions (circle one)?... Yes No

Researcher began with the sample items (circle one)?................................. Yes No

Researcher gave corrective feedback when needed (circle one)?................. Yes No

Researcher administered all items (circle one)?... Yes No

Researcher administered items in presented order (circle one)?.................. Yes No

Researcher timed the task correctly (circle one)?.. Yes No

Researcher recorded all item scores accurately (circle one)?........................ Yes No