5-2017

A Dynamic Measure of Morphological Awareness in Young Children

Frances Elizabeth Gibson
Utah State University

A DYNAMIC MEASURE OF MORPHOLOGICAL AWARENESS

IN YOUNG CHILDREN

by

Frances Elizabeth Gibson

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

of

DOCTOR OF PHILOSOPHY

in

Disability Disciplines

Approved:-

Julie Wolter, Ph.D. Major Professor	Timothy Slocum, Ph.D. Committee Co-Chair
Ron Gillam, Ph.D. Committee Member	Kim Corbin-Lewis, Ph.D. Committee Member
Cindy Jones, Ph.D. Committee Member	Mark R. McLellan, Ph.D. Vice President for Research and Dean of the School of Graduate Studies

UTAH STATE UNIVERSITY
Logan, UT

2017

ABSTRACT

A Dynamic Measure of Morphological Awareness in Young Children

by

Frances E. Gibson, Doctor of Philosophy

Utah State University, 2017

Major Professors: Julie A. Wolter, Ph.D., Timothy Slocum, Ph.D.
Department: Special Education and Rehabilitation

This study investigated the validity of a dynamic measure of morphological awareness (DMMA) in young children. During the first semester of first grade, 78 children completed a language and literacy battery of tests focused on morphological awareness, general cognitive ability, general language ability, phonological awareness, vocabulary, word-level reading, and word-level spelling. Morphological awareness was assessed using a standardized static measure and an experimental dynamic measure comprised of two subtasks, receptive discrimination and expressive production.

The validity of the interpretations of morphological awareness performance was explored through sources of evidence based on test content, internal structure and reliability. The performance relationships were explored between all the morphological awareness measures and with the other language and literacy measures. Moderate, significant correlations ($p < 0.01$) were found between the morphological awareness measures for the entire sample. Furthermore, moderate, significant correlations (i.e.,

mostly at the $p < 0.01$ level of significance) were found between the morphological awareness measures and the other language and literacy measures, except general cognitive ability and sight-word reading for the entire sample. However, significant performance differences were found between a typically performing group and an at-risk group of children. The interpretations of DMMA performance demonstrate adequate (i.e., more than 70%) levels of sensitivity and specificity when compared to the classifications of the morphological completion and sentence imitation subtests.

The unique contributions of morphological awareness as assessed by the experimental measure to word level reading and spelling are also explored. Morphological awareness may to contribute variance to word-level reading and spelling; however, whether this is a unique, significant contribution is still unclear at the present time. Further investigation is needed.

The DMMA appears to be a valid measure of the wide range of morphological awareness in young children in the early stages of language and literacy acquisition and development. The DMMA also appears to result in improved outcomes compared to the traditional, static assessments, especially for children who are at-risk for language and literacy difficulties. The DMMA is a promising tool to assess morphological awareness in young children.

(228 pages)

PUBLIC ABSTRACT

A Dynamic Measure of Morphological Awareness in Young Children

Frances E. Gibson

Although sound awareness has been proven critical for skilled literacy development, further investigation is needed to examine additional factors that could also be critical. Awareness of meaning or morphological awareness is an additional factor that could impact literacy development. Although morphological awareness is mastered early in spoken language, little is known in regard to this skill in other language and literacy contexts. This study investigated the validity of a dynamic measure of morphological awareness (DMMA) in young children. Seventy-eight first-grade children completed a language and literacy battery. Morphological awareness was assessed using both a standardized and an experimental measure comprised of two subtasks, comprehension and expression. The dynamic portion of the experimental task used a graduated prompting to support a child's accurate performance.

The validity of the interpretations of morphological awareness performance was explored through multiple sources of evidence. The DMMA content was designed to include both types of meaning units balanced across the subtasks. The stimuli selection, structure, administration and scoring were considered in the design provide consistent presentation and documentation of the children's responses. The performance similarity and dissimilarity were explored and discovered as evidence for the internal structure of the DMMA. The DMMA also appeared to demonstrate consistent measurement of

performance that provides more evidence for validity.

The performance relationships were explored between all the morphological awareness measures and with the other language and literacy measures. Medium-sized, significant relationships were found between the morphological awareness measures individually and with the majority of the other language and literacy measure for the entire sample. However, significant differences were found between the performance subgroupings. The interpretations of DMMA performance appeared to be sensitive to specific classifications predictions when compared to other predictive measures.

The unique contributions of morphological awareness as assessed by the DMMA to literacy skills were also explored. Morphological awareness appears to potentially contribute variance to literacy skills; however, whether this contribution needs to be explored further in young children. These current findings could have been impacted due to the developing emergent nature of the skills targeted in the study population and the significant performance differences between the two subgroupings.

The DMMA appears to be a valid measure of the wide range of morphological awareness in young children in the early stages of language and literacy development. The DMMA also appears to result in improved outcomes compared to the traditional assessments, especially for children who are at-risk for language and literacy difficulties. The DMMA is a promising tool to assess morphological awareness in young children.

ACKNOWLEDGMENTS

"Donde hay gana, hay maña." - Spanish proverb
Literal translation: "Where there is desire, there is ability."

> "If a man empties his purse into his head, no man can take it away from him. An investment in knowledge always pays the best interest."
> —Ben Franklin

I would like to thank my dissertation committee—Dr. K. Corbin-Lewis, Dr. R. Gillam, and Dr. C. Jones—for all of their time, support and guidance throughout this research. Thank you for the constructive, yet supportive feedback that helped to create this research product.

> "I was not sure where I was going, and I could not see what I would do when I got [there]. But you saw further and clearer than I, and you opened the seas before my ship, whose track led me across the waters to a place I had never dreamed of, and which you were even then preparing to be my rescue and my shelter and my home."
> —Thomas Merton

I would also like to thank my primary co-advisors—Dr. Julie Wolter and Dr. Tim Slocum—for their dedication and attention throughout this process. Thank you for helping me to mold the clay of this research proposal and to tame the beast that became the project as a whole. I will be forever grateful to you for all your time and investment in this achievement. You have been integral to my success. I am looking forward to showing you a return on your investment with even more future accomplishments.

> "The goal of education is the advancement of knowledge and the dissemination of truth."
> —John F. Kennedy

I am especially grateful for the participants and their families for their willing involvement in this research. Moreover, I want to express my gratitude to the individual schools, the administrations, and teachers, who also willingly participated in this research—Mr. Johnson, Mrs. Kishbaugh, and Mrs. Moore at the Edith Bowen Laboratory School; Mrs. Sorenson, Mrs. Bill, Mrs. Lindstrom, and Mrs. Jensen at Wilson Elementary School; and Mrs. Ware, Mrs. Ermer, and Mrs. Wright at Adams Elementary School—amidst the myriad roles and responsibilities of the academic year. Thank you for your collaboration in helping to advance language and literacy research.

> "Education is simply the soul of a society as it passes from one generation to another."
>
> —G. K. Chesterton

I would like to thank the graduate students who assisted me with the data collection process—Payton Allen, Jaci Curth, Mackenzie Coombs, and Cheska Dietsch. You were an invaluable addition to this research.

> "The language of friendship is not words but meanings."
>
> —Henry David Thoreau

I also would like to acknowledge and express my gratitude for my "village" of people that have supported me throughout this leg of my academic journey: the individuals of Utah State University (USU)—COMD and SPED departments, my colleagues in the USU-Speech-Language Hearing Clinic (USU-SHLC; Anne Elsweiler, Kristin Mosman, Arron Preston, Debbie Amundson, Debbie Evans, Jill Andrus, Jackie Littledike, Jamie Meecham, Leigh Stott) and my supervisees. Your comradery and friendship have been invaluable to me throughout my program. I will cherish these bonds as I move onto the next steps in my journey.

> "The only rock I know that stays steady, the only institution I know that works, is the family."
>
> —Lee Iacocca

I want to thank my family, especially my parents. Thank you for all of your love, support, encouragement, and guidance throughout my entire life. None of this would have been possible without you. I am so very lucky and blessed to have you as parents. I also want to thank my extended family for all of their support and love throughout this process. I want to thank my "friend family" that also has supported me through this process. I am looking forward to coming home.

"Came but for friendship, and took away love." —Thomas Moore

I also want to thank my "Logan" community for providing a second home when I was so far away from my first home. Your hospitality and friendship assisted my acclimation to this leg of my journey. I would like to thank the National Institutes of Health (R03 grant PAR-10-055) for their essential funding that provided financial support during my doctoral studies

> "Learning is not attained by chance, it must be sought for with ardor and attended to with diligence."
>
> —Abigail Adams

Thank you to everyone who has played a part in my overall academic journey and ultimate success!

"Trabayo feito ben parez." - Asturian proverb—
Literal translation: "A finished work looks good."

Frances Elizabeth Gibson

CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Language and Literacy

Oral language dates back to prehistoric times as a means to communicate wants and needs and predates literacy (i.e. written language) in the history of human civilization. Oral language and literacy historically served different purposes. Oral language provided a mode of communication and literacy provided a means to record the communication within a society. Thus, literacy development is rooted in oral language and involves a process of communicating written language, whether this language is being decoded (i.e., reading) or encoded (i.e., writing and spelling; Paul, 1995). Children begin to develop literacy from a foundation of spoken language, and it is built on what is already known and mastered, especially in the English sound system. The ability to understand and express spoken language can directly impact literacy development (Carlisle, 2003). On a basic level, literacy integrates multiple linguistic codes also used in spoken language—not only sound (phonology) but also meaning (morphology and semantics), structure (morphology and syntax), and, potentially, use (pragmatics). Successful, skilled literacy requires rapid and accurate decoding of these various linguistic codes as well as comprehension on multiple levels from the word- to passage-level written text.

Literacy as a Metalinguistic Construct

Metalinguistic abilities are built on all facets of language development. Similarly, language development is built off of multiple aspects that function both independently and in combination for functional literacy ability. Phonemes, the smallest linguistic units of sound, are first learned in spoken words and the awareness of how phonemes are segmented and blended to make words is important in learning to read and spell. Similarly, morphemes, the smallest units of meaning, and semantics, the wider word meaning and message, are strongly related through their connection to word meaning. For example, semantic organization of words into associational networks relies on children's metalinguistic abilities to determine relationships between words beyond just knowing their meaning (e.g., morphologically related words like mark, remark, remarkable, remarkably, unremarkably; Nagy, Anderson, Schommer, Scott, & Stallman, 1989; Owens, 1996).

Metalinguistic skill is a critical impetus to the development of language/ literacy in the school-age years and beyond (Apel, Wilson-Fowler, Brimo, & Perrin, 2012; Nippold, 2007). Young children usually shift toward more of a metalinguistic focus around the time they enter school depending on their past language exposure and this development continues well through elementary school (Bowey & Francis, 1991; Justice & Ezell, 2004; Webster & Plante, 1992). By about third or fourth grade, students are increasingly required to shift from learning language to read to reading and using language to learn in order to meet progressively complex academic demands (Anderson & Nagy, 1989; Nagy, Berninger, & Abbott, 2006; Nagy, Carlisle, & Goodwin, 2014).

Some researchers hypothesized that this shift in focus may be correlated with successful development of metalinguistic skills since literacy itself can be thought of as a metalinguistic skill (Carlisle, 1995; Owens, 1996; Pence-Turnbull & Justice, 2011).

Researchers are still debating the relationship between language and metalinguistic development. Some postulate that language and metalinguistic abilities develop in stages and that this development is connected to both cognitive development and beneficial experiences like formal literacy instruction in school (Carlisle, 1995; Chesnick et al., 1992; Tunmer et al., 1988; Valtin, 1984). One proposed model posits a three-stage model of language awareness from: (1) unconscious awareness (automatic use of language), (2) actual awareness, and (3) conscious awareness (Valtin, 1984). More specifically, during the second stage of actual awareness, children increase their ability to abstract language from the contexts, but still rely on implicit knowledge of language. At the conscious level, the ability to abstract language is fully developed from context with explicit knowledge of language. Ultimately, it is this final stage of conscious awareness that is supported and enhanced by formal instruction and literacy skills (Valtin, 1984). The transition between the second and third stages of this model from actual awareness to conscious awareness appears to be of the most interest to researchers and educators because of the overlap with literacy development and formal schooling.

Literacy, Phonological Awareness, and Beyond

Overall, metalinguistic language abilities are some of the best indicators of ultimate literacy success (Katz, Shankweiler, & Liberman, 1981; Kemper, 1985; Mann, Shankweiler, & Smith, 1984; Shankweiler, Liberman, Mark, Fowler, & Fischer, 1979), and much of the research on the relationship between metalinguistic skill and literacy development has focused on phonological/ phonemic awareness (Goswami, 2000; Kuo & Anderson, 2006). Phonemic awareness, the ability to consciously reflect on and use speech sounds, has been established as integral to skilled literacy development (Apel et al., 2012; Carlisle & Nomanbhoy, 1993; Deacon & Kirby, 2004; Ehri et al., 2001; Mahony, Singson, & Mann, 2000; National Reading Panel, National Institute of Child Health, & Human Development, 2000). The evidence indicates that phonological awareness accounts for 28% to 43% of the variance in the performance of children's word-level reading and spelling (Apel et al., 2012; Cunningham, Perry, & Stanovich, 2001; Manis, Doi, & Bhadha, 2000; Swanson, Trainin, Necoechea, & Hammill, 2003). Moreover, the pairing of phonemic awareness and vocabulary comprehension is also highly predictive of literacy success according to the National Reading Panel (2000). Although the evidence supports the contribution of phonemic awareness and vocabulary to literacy development, more recent research indicates the need for an expanded focus to increase children's metalinguistic ability beyond phonological awareness to additional metalinguistic factors such as morphological awareness and orthographic awareness, that also account for unique contributions to literacy development. For example, Wolter, Wood, and Dzatko (2009) found that morphological awareness contributed 10% and 8% unique, significant variance in reading and spelling, respectively, beyond phonological awareness. Likewise, Walker and Hauerwas (2006) found that orthographic awareness accounted for 39% to 60% unique, significant variance to spelling depending on the targeted task.

To become skilled readers and writers, children must develop effective decoding strategies for many written words such as the alphabetic principle, or linking phoneme (sounds) to grapheme (letter) correspondences as well as an understanding of morphological meaning units (Mahony et al., 2000; Nunes, Bryant, & Barros, 2012; Wolter et al., 2009). A well-established evidence base indicates that as children develop skilled literacy they progress from an initial reliance on the basic phoneme-to-grapheme correspondences that involve individual sounds and letters (i.e., small units) to a more efficient and effective dependence on the expanded larger units that involve words or phrases for successful reading and writing (Deacon & Kirby, 2004; Ehri, 2000; Ehri & Wilce, 1985; Mahony et al., 2000; Nunes et al., 2012; Wolter et al., 2009). Written English is not only alphabetic, but also morphophonemic, which transmits sound and meaning information within the larger symbolic units like the syllable, morpheme (e.g., basic meaning units), and related linguistic boundaries such as words and sentences. The majority of departures from a strictly alphabetic perspective in the spelling conventions in written English preserve the consistency of morpheme unit spelling (Nagy et al., 2006). Thus, written word recognition involves the processing of words at a level beyond the phoneme-grapheme relationship which can include the awareness of syllable units or morphemic units (Nunes et al., 2012).

Morphology provides useful cues to help decipher semantic, syntactic, phonological, and orthographic codes. Unlike the processing of phoneme-grapheme or larger units such as syllables, consideration of the morpheme(s) of a word can be a more useful linguistic component to children because additional semantic and/or syntactic meaning can be inferred from this linguistic processing. That is, the morphological code can be used to infer meaning and grammatical information and thus overall comprehension of a word. Indeed, evidence indicates that children with stronger morphological awareness demonstrate increased skill at deciphering unfamiliar words in both spoken or printed language (Nagy et al., 2006; Nunes & Bryant, 2006; Nunes et al., 2012). In addition, morphemes also have a consistent and predictable orthographic structure, (i.e., the letter sequence) and pronunciation (e.g., *–ed* indicates past tense with a predictable pronunciation of /*d*/ after the letter *n* in a word like *rained*) and as such morphemes can be useful in inferring the decoded pronunciation and spelling of a word (Carlisle, 1995; Carlisle & Nomanbhoy, 1993; Deacon & Kirby, 2004; Nagy et al., 2006). In conclusion, morphological awareness is an important metalinguistic skill that can directly impact successful literacy development, and thus further examination of this skill is needed in order to discover how best to assess and harness this skill to support children's literacy success.

CHAPTER II

LITERATURE REVIEW

Morphological Awareness

Morphological Awareness Defined

Morphological awareness involves the metalinguistic capability to explicitly think about and manipulate the morphological code (i.e., word structure conventions; Carlisle, 2003; Kuo & Anderson, 2006). Compared to other related constructs, morphological awareness involves the application of tacit morphological conventions in a definitive manner (Kuo & Anderson, 2006). The construct of morphological awareness develops over time and refers to, "the explicit metalinguistic ability to actively reflect on how base words (roots) and affixes (i.e., prefixes and suffixes) create and/or change meaning" (Wolter, 2014, p. 229). Even the though they can easily produce morphemes (e.g., plural nouns) in oral language, young children are not explicitly aware of morphemes until early elementary school and knowledge of the morphological rules in a language is often implicit (Berko, 1958). Beyond the basic level of this knowledge, the use of morphological knowledge requires some metalinguistic skill (e.g., cognizant attention to the morphological units and word structure; Kuo & Anderson, 2006; Nagy, Berninger, Abbott, Vaughan, & Vermeulen, 2003; Rubin, Patterson, & Kantor, 1991). Thus, morphological awareness is an explicit ability and contrasts with the construct of morphological production that naturally happens in conversation and generally involves unconscious expression of morphemes (Apel & Lawrence, 2011).

Theoretical Foundations

Early literacy involves a shift from primarily spoken or external language to more of a mental or internal language system (Tunmer et al., 1988). This subsequent processing involves metalinguistic abilities that include a deeper understanding and awareness of language (Kintsch & Kozminsky, 1977). General linguistic theory states that young children must have some means of relating language to the events of the world it represents. They also must be able to ultimately identify the relevant units of the language they hear without explicit help from adults, but often do require some assistance in the early stages of language development in that they initially rely on adult modeling, scaffolding, and prompting.

Any model of reading acquisition informs not only how students are instructed, but also has important implications for both students who are typically-developing and those who struggle. Since phonological processing is so critical to the development of skilled reading, weak decoding decreases a student's chances to develop printed word knowledge, seen in a weaker correspondence between phonological and deeper, linguistic aspects (i.e., morphological characteristics) in these students (Share, 1995). However, in the face of this phonological challenge, readers who struggle appear to rely more on their metalinguistic skill. This relative shift in focus provides another illustration of a natural impulse within readers who struggle to utilize nonphonological factors as a compensatory strategy for questionable phonological skills, highlighting the use of metalinguistic skill as a compensatory tool. However, proficient literacy requires both adept phonological and metalinguistic processing (Share, 1995). At the present time, there is a need to examine and document the typical development of these essential metalinguistic skills to inform future investigation of these skills, particularly in students who struggle.

The self-teaching hypothesis. Share (1995, 1999, 2004) proposed the *self-teaching* hypothesis to explain printed word learning as a dual-factor process. According to the self-teaching hypothesis, phonological processing is the primary factor critical for word identification and learning. However, phonological processing is not the singular factor integral to printed word learning, but it is only a beginning to the recognition and comprehension of a word by providing initial access for additional, deeper word learning (Share, 1995).

Share (1995) posits that phonological recoding simply provides the opportunity for self-teaching of supplementary features such as the awareness of other deeper, higher-order linguistic characteristics like morphological features that factor into how words are ultimately recognized and learned. Even though these higher-order linguistic characteristics are relegated to a secondary position, according to the self-teaching hypothesis, their inclusion in word-learning should explain more variance in literacy outcomes beyond that explained by the phonological characteristics since these skills probably originate from slightly different cognitive processes (Share, 1995). This point highlights the fact that the printed word contains more linguistic, higher-order morphological information rather than just phonological information. In early reading development, the process of understanding, decoding/translating and expressing does appear to begin with the initial “cracking” of the phonological code; however, as children become more adept readers they begin to incorporate the additional consideration of the morphological, semantic, and syntactic codes. For example, Cunningham's (2006) findings suggest that semantic and syntactic structure within connected text reliably impacts accurate target word reading and supports printed word knowledge.

After beginning readers demonstrate a firm mastery of the phoneme-grapheme correspondences, their focus shifts to utilizing other deeper, linguistic characteristics further in order to build essential skill for proficient literacy (Share, 1995, 2004;Share, Brady, Braze, & Fowler, 2011). The results of Share (1999) first supported this hypothesis by finding that children needed minimal exposure to successfully learn novel words with learning, "attributable to *saying* rather than *seeing* the novel letter strings" (Share, 1999, p. 111—with original emphasis). Share (1995) also supports this idea through his statement that, "there is evidence that rudimentary, yet functional self-teaching may develop at the very outset of learning to read…well before a child has acquired 'conventional' decoding skill" (p. 163). Even children, as young as first grade, not only demonstrated the acquisition of higher-order, word knowledge but also that this acquisition occurs in a relatively rapid manner (Cunningham, 2006).

Young children bring what they know and have developed in their oral/spoken language to the language-based literacy learning process. For example, many young children exhibit consistent mastery of morphological inflections (i.e., *-s*, *-ing*, and *–ed*) in spoken language as well as some emergent skill with early morphological derivations (i.e., *-y* and *–er*). The transition from implicit knowledge of language structure especially morphological knowledge in spoken language to explicit knowledge and use initially occurs through self-teaching from multiple models and exposure to appropriate language use from mature language users and/or the environment around them. However, the beginning of formal schooling not only increases the probability of additional models and exposure, but also includes direct instruction that highlights the targeted forms further facilitating morphological learning. Now, the children have the additional task of understanding and using these morphological structures appropriately when words are encountered during reading and when words are needed during writing. Within literacy activities, morphology connects with the deeper, higher-order linguistic characteristics of semantics, syntax, and orthography. Ultimately, morphology plays an integral part in the form and meaning of written words. Knowledge of morphological units within words supports reading and spelling through "chunking" the base word and the morphological unit separately and decreasing the retrieval demand. Children, especially young ones, find it easier to remember the larger units that compose the whole word rather than all the individual phonemes at once. Within connected text, morphological units add to the overall meaning and message of the text. Weak knowledge of these units can lead to misinterpretation and/or lack of understanding of the text, which helps to explain the correlation between morphological awareness and reading comprehension. The automatic recognition of words exhibited by proficient readers is acquired from, "word specific…representations linked to phonological, semantic, morphological, and syntactic information" (Share, 2004; p. 267). The evidence continues to support a case for the need for additional deeper, linguistic units beyond single phonemes in reading and literacy acquisition (Share et al., 2011; Share, 1995).

Lexicalization. Early decoding skill is "based on simple one-to-one correspondences that are relatively insensitive to orthographic and morphemic context" (Share, 1995, p 163). However, once this basic decoding skill is established, the connection between phonology and the other linguistic, deeper characteristics of a word becomes more apparent and "increasingly context-sensitive or 'lexicalized'" (Share, 1995, pp. 163-164). In addition to the self-teaching hypothesis, the lexicalization hypothesis attempts to explain the shift from basic phonological to deeper, metalinguistic processing of words and their components. As their skill develops, readers who are typically-developing are compelled to move beyond the basic phonological skill level to examine "higher-order regularities" that are often specific to the individual word or "lexicalized" (Share, 1995). Many words, especially multimorphemic words, contain specific characteristics (i.e., higher-order regularities) beyond the phoneme-grapheme correspondences that facilitate the word recognition process to further build and refine individual word knowledge within children's mental lexicon. For example, the word *walked* can be processed beyond the initial phonological decoding to include the bound morpheme *–ed* that adds to the base verb *walk* to indicate that it occurred in the past (i.e., regular past tense morphemic unit). The learning of additional base words (i.e., through initial phonological recoding) like regular past tense verbs (e.g., *looked*, *talked*, and *worked*) also supports the deeper acquisition of the past tense morpheme, which is the higher-order regularity common to all three words. These higher-order regularities support more specific differentiation among the words that expand a child's lexicon (e.g., the *–ed* on *walked* denotes a specific, deeper difference as compared to *walk*; signifies added the addition of "past tense"). Using a similar explanation, Katz and Frost's (1992, as cited in Share 2004) "orthographic depth hypothesis" posits that, "deep orthographies [, like English,] encourage a reader to process printed words by referring to their morphology…" (p. 291), which is also supported by the findings of Cunningham (2006) and Seymour, Aro, and Erskine (2003).

The lexicalization hypothesis is also supported by additional research evidence. Fledgling readers find elements involving more consistent, context-free connections easier to read (Share, 1995). The natural decrease in over-generalization of applying regular correspondences to irregular words, also known as "regularization," by these fledging readers is seen as support for the lexicalization hypothesis (Share, 1995). For example, young children who are developing their knowledge of past tense often apply the regular past tense ending of *–ed* to both regular and irregular verbs (e.g., *eated*) until they develop mastery of the distinction between regular and irregular verbs in the past tense. Self-teaching supports the learning of linguistic conventions through repeated exposure and opportunities to apply these conventions accurately. Finally, the lexicalization hypothesis is further supported by early spelling development evidence that parallels the developmental pattern of reading and moves sequentially from dependence on the direct sound-to-symbol correspondences before shifting to use of higher-order regularities, such as bound morphemes and letter sequences (Share, 1995).

Overall, the self-teaching hypothesis describes this unique process by presenting "a theory about how children teach themselves to read" (Share, 1995, p. 201). A strong evidence base exists supporting a typical progression from a beginning and overgeneralized sound-symbol system within young children maturing into a more cultivated, refined knowledge of the interconnection between phonology and higher-order regularities, including morphology and morphemic patterns, that is essential for accurate, competent skilled literacy (Share, 1995). Additionally, because of the complexity of the knowledge and skills required for competent reading from a basic to "lexicalized" level, it is only possible to directly teach the necessary building blocks (i.e., linguistic units like morphemes) so that children can then ultimately construct and modify their own understanding as they develop their morphological awareness skill.

Morphological Development

There are two main types of morphemes that children develop, free and bound. Free morphemes can stand on their own to fully convey their meaning, like root words (e.g., *book* and *talk*). Two free morphemes can be combined to create a new blended meaning in a compound word, like the words *sunshine* and *cowgirl*, but each morpheme or word segment does have a stand-alone meaning. In contrast to a free morpheme, a bound morpheme needs to be attached to another morpheme in order to fully convey it's meaning, like *-s* to signal plurality or *-ed* to signal regular past tense. Bound morphemes can also be subcategorized into two types, inflectional/grammatical and derivational. Inflectional or grammatical morphemes add basic grammatical information, like plurality (*-s*), possession (*-'s*), or tense (*-ed*), and typically develop before derivational morphemes (R. Brown, 1973; Miller & Chapman, 1981; Moats & Smith, 1992). The later developing derivational morphemes, on the other hand, are added to root words to change word class (e.g., the verb *teach* to noun *teacher*) and expand meaning and create word families (e.g., *schools*, *preschool*, *scholar*, *scholarship* that all share the same common root meaning of *school)*.

Inflections. There is general agreement that children demonstrate morphological awareness of inflections prior to derivations during metalinguistic activities and connected speech (Adams, 1990; Carlisle, 2003; Kirby et al., 2012; Kuo & Anderson, 2006). Most of the English inflections are mastered by the beginning of formal schooling (Anglin, 1993; R. Brown, 1973) although estimates of exact age ranges for acquisition differ markedly. Children and adults who have difficulty learning language and/or acquiring literacy also have similar difficulty mastering inflectional endings (Carlisle, 2003; Leonard, 1998; Liberman, Rubin, Duques, & Carlisle, 1985; Rubin et al., 1991; Vogel, 1983; Wiig, Semel, & Crouse, 1973). A striking finding in Rubin et al.'s (1991) study was that the performance of adults with a history of language/literacy learning difficulties did not differ significantly from that of second graders with this same history, which suggests that morphological development does not simply evolve due to language experience or growth as cited by Carlisle (2003). The complexity of children's language can provide a strong estimate of his level of linguistic development (Anglin, 1993; R. Brown, 1973).

According to developmental evidence, the earliest words acquired by young children contain only one morpheme (i.e., monomorphemic; free morphemes without any bound attachments), but the complexity within this basic lexicon increases exponentially as children encounter the contextual demand for inflections, even during the initial years of language acquisition (Anglin, 1980, 1993; R. Brown, 1973; Cazden, 1968; De Villiers

& De Villiers, 1973).

Derivations. Derivations are the latest morphological form to develop; however, they may develop earlier than what was originally believed. Some derivational endings appear to be learned relatively earlier than originally expected in the latter preschool years rather than during the school-aged years (Clark & Cohen, 1984; Clark & Hecht, 1982). Clark and Hecht explored the comprehension and production of the two types of the *-er* suffix (e.g., the agentive and the instrumental. They found that 36 of their 48 children from 3- to 5-years-old demonstrated comprehension of both of these types, but they appeared to have a better understanding of the agentive form as compared to the instrumental form. The three-year-olds produced the agentive form (e.g., *-er*) about 55% of the time verses 91% production by the five-year-olds. Moreover, the 3-year-olds produced the instrumental form about 42% of the time versus 72% production by the 5-year-olds (Anglin, 1993; Clark & Hecht, 1982). Bowerman (1982) provided evidence that preschool children may have some understanding and use of the *un-* reversative prefix although they may inaccurately apply or overregularize this understanding, until it is fully mastered in the latter preschool years (Anglin, 1993).

There is evidence for a substantial growth in the understanding of derivations from grade one to grade five. A large portion of this growth is due to an increased acquisition and understanding of morphologically complex words that contain three or more morphemes, called multimorphemic. Furthermore, there also is evidence that children increasingly use morphological problem-solving to add to their mental lexicons, which is consistent not only of word and word parts directly learned as unique wholes but also of words that can be figured out based on their connection to words and word parts already learned (Anglin, 1993).

Overall, the evidence indicates that the morphemic categories are not fully mastered individually before subsequent categories (e.g., inflections before derivations; Carlisle, 2003). Even though children entering kindergarten may have a strong knowledge of inflections they do not appear to have fully acquired these (e.g., the *-es* plural allophone) until after some formal schooling (Carlisle, 2003; Gleason, 2012). Young children definitely exhibit some knowledge of the morphological endings, but have not fully mastered how to appropriately apply the rules that govern the construction of words (e.g., *flyable* in Clark, 1982; *unstraighting* in Bowerman, 1982; also cited by Carlisle, 2003).

Additional factors affecting morphological awareness development. Beyond consideration of inflectional and derivational morphology, the three additional factors of word frequency, transparency, and imageability appear to impact children's morphological awareness development (Anglin, 1993; Freyd & Baron, 1982) and will be discussed separately.

Word frequency. When reflecting on the development of morphological awareness, the frequency of written multi-morphemic word exposure, or word frequency, is a factor influencing development of this metalinguistic skill. Word frequency can vary depending on what texts are being used to gather these occurrence counts. *The Educator's Word Frequency Guide* (Zeno, Ivens, Millard, & Duvvuri, 1995) provides word frequencies based on more than 17 million words within the written texts of school-aged children. Words with a frequency of 45-50 or higher can be considered high-frequency while those with 40 or lower can be considered low-frequency (Larsen & Nippold, 2007; Wolter, 2014; Wolter & Pike, 2015). Knowledge of word building is increasingly important to literacy skills as children's experience builds and shifts from primarily the high frequency words that support initial reading and spelling skill to low frequency, more morphologically complex words, especially in the later elementary years and beyond (Berninger et al., 2010; Nagy & Anderson, 1984). Words with higher frequency morphemes are likely to be attended to and acquired before those with lower frequency (Carlisle & Katz, 2006; Deacon, Whalen, & Kirby, 2011; Nippold & Sun, 2008). Throughout formal schooling, children experience an estimated 88,000 unique words in their academic materials up to the end of middle school (Berninger et al., 2010; Nagy & Anderson, 1984). Furthermore, up to three word relatives can be comprehended from each unique word that has been acquired (Berninger et al., 2010).

Word transparency. Morphologically related words can have transparent relationships with little to minimal phonological (sound) or orthographic (spelling) changes from the base root word to the morphologically complex form (e.g., pronunciation and spelling; *six* and *sixth* or *box* and *boxing*). Morphologically related word pairs may be less transparent and thus have more opaque relationships where there are phonological and or orthographic changes from the base root word to the morphologically complex form (e.g., *five* and *fifth* or *nature* and *natural*).

Children's attention to and acquisition of morphologically complex words is impacted by this level of transparency of the relationship within the word pair (e.g., the base root word and the morphologically complex word form; Carlisle, Stone, & Katz, 2001; Clin, Wade-Woolley, & Heggie, 2009; Deacon & Bryant, 2005; Windsor, 2000; Wolter, 2014). On morphological awareness tasks, stimulus items containing transparent relationships resulted in more accurate young school-age performance compared to those items with opaque relationships (e.g., Carlisle & Nomanbhoy, 1993; Carlisle et al., 2001; Deacon & Bryant, 2005; Gonnerman, Seidenberg, & Andersen, 2007; Libben, Gibson, Yoon, & Sandra, 2003; Nagy & Anderson, 1984; Windsor, 2000; Wolter, 2014). Thus, those suffixes that can be considered more transparent (e.g., requiring fewer adjustments to attach to the base root word) impact the development of derivational morphemes (e.g., Carlisle, 1988, 2000; Champion, 1997; Tyler & Nagy, 1989). When derivational errors are made, younger children tend to apply these more transparent and frequently occurring suffixes probably since they need minimal changes (e.g., responding with *producement* rather than the targeted response *production* when provided with the base word of 'produce' and a sentence context; Carlisle, 1988, 2000).

Word imageability. Finally, similar to transparency, imageability factors into the ease with which children explicitly use morphological rules during word recognition. The imageability of a word refers to how readily a word evokes a cognitive mental visual representation (Paivio, 1991). Words that are more readily visually represented or have high imageability (e.g., *explode*) may be acquired more easily than those that are more abstract and less readily visualized or have low imageability (e.g., *decide*; de Groot, 1989; Masterson, Druks, & Gallienne, 2008; Prado & Ullman, 2009; Strain, Patterson, & Seidenberg, 1995). Wolter (2014) explored this imageability in morphologically complex words and found that morphological derivatives with high imageability (e.g., *explosion*) were more accurately produced in a morphological awareness task than those of low imageability (e.g., *decision*) in typically developing third-grade children without a confound of word transparency. In addition, Dye, Walenski, Prado, Mostofsky, and Ullman (2013) found that as young children develop their morphological awareness skills imageability has a stronger impact than word frequency, but this impact gradually shifts as morphological awareness skill is developed.

Assessment Research

To date, only a handful of standardized measures of morphological awareness have been developed. Most of these standardized measures are subtests that are part of wider assessments (Apel, 2014). The majority of morphological awareness assessments are nonstandardized. Although, several researchers have developed their own measures to assess morphological awareness, these measures vary widely in the main targeted tasks including: production, discrimination, manipulation (segmenting and blending), and analogy. Also, the majority of these non-standardized assessments target spoken responses only, which may or may not also reflect the children's written response. Moreover, variation is possible within the stimuli items (inflections and/or derivations) and their characteristics (frequency, transparency and/or imageability; Apel, 2014). Finally, most of the current studies have focused on older children rather than younger children who are still developing their metalinguistic skills like morphological awareness. Despite these limitations, multiple efforts using these measures have been made to assess children's use of morphological awareness to determine whether children use this knowledge early in their development and whether it is related to their early literacy success. The following sections will review the relevant research and note the specific tasks used because they directly inform the current study.

Early elementary morphological awareness tasks and literacy success. As one of the first studies that focused on morphological awareness assessment, Carlisle (1995) followed longitudinally 85 typically-developing young children from kindergarten through second grade. She explored whether a) morphological awareness uniquely contributed to later reading performance over language knowledge; b) there was a significant amount of morphological awareness growth between the early grades of kindergarten and first grade; and c) whether phonological and morphological awareness performance accounted for significant variance in second grade reading achievement as measured by word analysis and comprehension.

To assess morphological awareness, Carlisle (1995) designed two tasks - one focusing on production-expressive and another judgment-receptive, to administer to both kindergarten and first grade children. The production (expressive) task required that children respond with a morphologically complex word form in response to a stimuli sentence (e.g., "*Farm. My uncle is a* ___."—with the expected response of *farmer*). The types of stimuli responses were distributed between several varied forms requiring: inflectional morphemic changes (e.g., *toy* and *toys*), derivational, transparent morphemic changes (e.g., *drive* and *driver*), and derivational, opaque morphemic changes with phonological changes as well (e.g., *explode* and *explosion*)—see the following sections for more information on transparent and opaque stimuli word pairs. Additionally, the researcher attempted to select stimuli targets that were familiar to young children. The initial judgment (receptive) task was revised due to strong floor effects with high error rates and apparent guessing. The revised task required the children to judge the relationship between a targeted word pair of a base root and a morphologically complex word form on a novel "silliness" scale whether the pairing *made sense* or *is silly* in response to a stimuli sentence that was either accurate (e.g., "*A person who teaches is a teacher*") or a foil (e.g., "*A person who makes dolls is a dollar*").

The results of Carlisle's (1995) study indicate a significant relationship between the two morphological awareness tasks (i.e., production and judgment-revised; $r = 0.55$), but the production task was more related to both reading outcomes (i.e., phonetic/word analysis and reading comprehension) than the revised judgment task, which was consistent with other evidence (Carlisle, 1995; Carlisle & Nomanbhoy, 1993; Fowler, 1988). Researchers have hypothesized that children put more effort and attention to a task requiring them to produce a word rather than just a dichotomous yes/no judgment as required by the researcher-developed silliness scale—*makes sense* or *is silly*. Also, the production task involved manipulation and analysis of the sound structure of words as well as the syntactical and meaning contexts that was a closer approximation of what needs to be done during reading comprehension tasks. Morphological production tasks have potential to target the skills and knowledge necessary for morphological awareness.

Second, the results indicated that a portion of the variance within both reading outcomes was explained by the metalinguistic nature of the tasks. Moderate, positive task correlations were found between the phonological awareness task and the morphological production task (*r* = .52) as well as between this same task and the morphological judgment task-revised (*r* = .30); although this latter relationship was weaker, it still highlighted a connection between the morphological and phonological tasks. While the judgment task required some latent, innate awareness, it also plausibly involved some explicit awareness as well. Even though the results of the judgment task performance could not be compared due to the necessary revision, the significant jump in performance on the morphological production task from kindergarten to first grade could have certainly signaled a developmental progression from an innate, implicit awareness to a more definitive, explicit awareness of morphology in these young children. According to the results of the regression analyses, the judgment task may not have effectively measured the children's morphological performance in relation to their reading comprehension outcome. It is important to note that issues with the judgment task were present from the beginning of the study as seen in the significant floor effects precipitating a revision during the study itself, which could also have influenced and negatively skewed these findings. Additionally, the use of only a phonetic/word analysis and a reading comprehension measure resulted in a limited measurement of reading abilities as these tests could only assessed the two poles on a spectrum of ability from basic phonological decoding to the more complex text integration for understanding. As an intermediate task between phonetic analysis and reading comprehension, the inclusion of a word-level reading measure such as including sight-word reading and decoding as well as the range from simple to difficult morphologically complex words, especially from grade level materials, could have provided an important bridge skill between levels of complexity. This could have created a more balanced estimate of language/literacy performance and development as a whole.

Carlisle's (1995) results supported the evidence that young children can demonstrate adequate performance on the tasks selected for this study, including the morphological production and judgment tasks (Carlisle & Nomanbhoy, 1993; Clark, 1978; Van Kleeck, 1982). Children as young as kindergarteners appeared to demonstrate emerging morphological awareness as evidenced by their performance on these tasks. Furthermore, not only did the performance of the first-grade children demonstrate a significant increase but this task also accounted for part of the variance in second grade reading outcome tasks along with phonological awareness performance. The best predictor of reading comprehension performance was the children's performance on the morphological production task. Therefore, morphological awareness appeared to also significantly contribute to early reading outcomes in young children. Carlisle (1995) concluded that the significant, complex relationship between phonological awareness and morphological awareness warranted further investigation in young children. These investigations should include variation in stimuli (i.e., inflectional and derivational morphemes) and tasks (i.e., production and judgment). Finally, task appropriateness for young children should be explored in order to maximize focus on the language skill being targeted by the assessment. These influences were taken into consideration in the current study. Carlisle's (1995) morphological awareness production and judgment tasks were adapted accordingly in the current study (see methods section for details) since the tasks were not originally designed for dynamic assessment.

Despite Carlisle's (1995) seminal study noting the importance of morphological awareness early in grade school literacy development, few studies were conducted to investigate morphological awareness in early school age children until more recent years. The majority of research directly after 1995 was focused on the development and importance of morphological awareness in later school-age years (e.g., Carlisle, 2000; Kuo & Anderson, 2006; Mahony et al., 2000; Schwiebert, Green, & McCutchen, 2002; Singson, Mahony, & Mann, 2000). It was only within the last approximately 10-15 years that researchers have begun to revisit and strategically investigate the development and importance of morphological awareness in early primary grades and have since found early morphological awareness to be significantly related to reading and spelling abilities even in typically developing children.

Contributions to reading performance. Morphological awareness contributes unique, significant variance in word-level reading (i.e., sight-word reading and decoding) beyond phonological awareness (Apel, Diehm, & Apel, 2013; Wolter et al., 2009) and other related factors (cognition, age, and letter knowledge; Apel & Lawrence, 2011) that appears to grow throughout the early elementary grades. Apel et al. (2009) have found that the unique, contributions of morphological awareness to sight-word reading ability increased from 11% in Kindergarten to 21% in third grade beyond the variance accounted for by phonological awareness (Apel et al., 2012, 2013; Apel & Lawrence, 2011). Similarly, the findings of multiple researchers also provide evidence for the unique contributions of morphological awareness to word-level reading beyond other related factors like phonological awareness, cognition, age, and intelligence (Apel & Lawrence, 2011; Deacon, 2012; Deacon & Kirby, 2004; Kirby et al., 2012) that also increase during the early elementary years.

In addition to sight-word reading, morphological awareness contributes to text-level reading speed or fluency after accounting for intelligence and phonological awareness and word-level reading speed (Kirby et al., 2012). Morphological awareness facilitates word-level reading through the processing of morphemic units and more accurate, efficient recognition of words (Carlisle & Stone, 2005; Kirby et al., 2012). Carlisle and Stone (2005) found that second- and third-grade children read morphologically complex words (i.e., two-morpheme, accurate words; *hilly*) faster and with more accuracy than foils words that mimicked the morphological complexity (i.e., one-morpheme, inaccurate words; *silly*). They also found a significant relationship between reading skill and the word-level reading measures (i.e., sight-word reading and decoding). Morphological awareness facilitates word-level reading through a connection to other linguistic factors like semantics even when only focused on inflections (Wolter et al., 2009).

Contributions to spelling performance. In addition to the receptive written-language skill of reading, morphological awareness also appears to significantly contribute to the expressive written-language skill of spelling beyond that of other established predictors in early primary grades. Apel and Lawrence (2011) studied the predictors of spelling in a mixed sample of first-grade children who were typically developing and those who were diagnosed with a speech sound disorder (SSD). After controlling for cognition, age, phonological awareness, and letter knowledge, they found that morphological awareness uniquely contributed variance to the spelling performance of both participant groups. However, morphological awareness contributed more variance to the spelling of the typically-developing children (18%) rather than the children with speech sound disorder (4%; Apel & Lawrence, 2011). Likewise, Wolter et al. (2009) found that morphological awareness uniquely contributed to the spelling of typically developing first-grade children after phonological awareness was accounted for, with both variables relating significantly to spelling. Finally, Apel et al. (2012) explored the contributions of a composite of linguistic variables that included morphological awareness to spelling after controlling for age in a sample of second- and third-grade children. The addition of this linguistic composite (morphological awareness, orthographic awareness, vocabulary, rapid automatic naming) accounted for an additional 36% unique variance in the spelling beyond the 19% accounted for by age with only morphological awareness significantly contributed to this variance.

In a related study, Walker and Hauerwas (2006) found that phonological awareness, morphological awareness and orthographic awareness all increased and developed in conjunction with that of spelling skill in grades 1 through 3 with differing levels of significance. Although morphological awareness significantly related to the spelling outcomes for the entire sample, only phonological awareness and orthographic awareness significantly related to the spelling of first graders. In second grade, only orthographic awareness remained significant. Finally, morphological awareness was the only significant predictor in the spelling of the third graders. These mixed findings highlight the connections that the various types of linguistic awareness, including morphological awareness have to spelling outcomes in the early elementary years; however, further research is needed to more fully explore this development of those relations.

Multi-linguistic aspects of literacy development. Morphological awareness and phonological awareness exhibit similar performance in reading acquisition (Deacon & Kirby, 2004). There appears to be a significant interaction between morphological awareness and phonological awareness in word-level reading (i.e., sight-word reading and decoding) and this interaction relationship may reflect an increased effect of morphological awareness when phonological awareness is weak versus when phonological awareness is strong (Bryant, Nunes, & Bindman, 1998; Deacon, 2012; Elbro & Arnbak, 1996). Morphological awareness appears to be impacted by phonological awareness with stronger phonological awareness being related to stronger metalinguistic skill that also supports stronger morphological awareness. For example, children with stronger phonological awareness appear to be more sensitive to the morphological structure of words compared to their peers (Carlisle & Nomanbhoy, 1993). In sum, other linguistic characteristics like morphological awareness appear to support the word-reading process when phonological awareness skill is weak, and this interactive relationship appears to be present even in younger elementary school children (Carlisle & Nomanbhoy, 1993; Deacon, 2012).

Morphological awareness uniquely contributes to reading skill beyond phonological awareness and this contribution appears to increase throughout the early grades (i.e., from Kindergarten to second grade; Apel et al., 2013; Deacon & Kirby, 2004). For example, morphological awareness was less developed in third graders compared to fifth graders but it was still significantly related to literacy development for both grades (Carlisle, 2000). Furthermore, the ability to spell increases with grade level, and this spelling ability follows a developmental pattern that improves as knowledge and awareness of the language and relevant components increases (Walker & Hauerwas, 2006). In conclusion, these results provide evidence for the literacy development theories based on the acquisition and simultaneous growth of multiple aspects of linguistic awareness (Apel & Lawrence, 2011; Apel & Masterson, 2001; Deacon, 2012). Moreover, these aforementioned results indicate that morphological awareness should be included in early school-age models of literacy development, language-literacy assessment batteries and literacy interventions (Kirby et al., 2012; Wolter et al., 2009).

Tasks and limitations of current early school-age morphological awareness. Despite the research noting the importance of morphological awareness in early school-age literacy success, the limitations of early school-age morphological awareness tasks continue to challenge researchers. The wide variation across researcher-developed morphological awareness tasks and prominent floor effects of current early tasks result in considerable discrepancy about what is known about early development and what early tasks may best predict literacy success.

Apel et al. (2012) adapted a morphological production task to include printed stimuli both with and without a base root word that elicited a written response rather than an oral response. The task stimuli focused solely on derivational suffixes only. Morphological awareness was the only linguistic variable examined that strongly related to all three literacy outcomes and was the focus of the researchers' investigation. They examined the contributions of a composite of linguistic performance including morphological awareness to literacy including reading and spelling beyond the contributions of age because both second- and third-graders were included in the sample. Consistent, unique contributions of age were found across all three literacy tasks (i.e., word-reading, spelling, and reading comprehension). Although both morphological awareness and orthographic awareness uniquely predicted the children's word reading performance, only morphological awareness was found to uniquely predict the children's spelling performance.

Apel et al. (2013) utilized four separate morphological tasks (blending with definition, production, spelling, and identification) to examine their contributions to the literacy outcomes of word-level reading, decoding, and reading comprehension. The task stimuli included both transparent inflections and derivations, but stimulus type was not equivalent between these two types. The findings differed between the multiple grade levels included in the sample. For the kindergarten children, only the morphological blending task neared significant prediction of word-reading and only the morphological relatives task predicted decoding. However, morphological awareness did not add any unique, additional variance into word-level reading, decoding nor reading comprehension. Although both the morphological awareness tasks exhibited small to medium sized relationships with the literacy outcomes, but phonemic awareness appeared to subsume all of this variance in the regression equations. Finally, for the second-grade children, the morphological production and spelling tasks uniquely predicted word-level reading while the blending and spelling uniquely predicted decoding but the production task approached significant prediction also. Only the production task uniquely predicted reading comprehension in the performance of second graders.

Carlisle and Nomanbhoy (1993) studied the predictors of picture vocabulary and phonological awareness in first grade children's morphological awareness (e.g., two separate tasks—judgment and production). The stimuli focused on the *–er* derivational morpheme for the judgment task and both inflections and derivations on the production task that were equivalently divided between transparent and opaque word pairs. They found that picture vocabulary and phonological awareness accounted for 19% of the variance in the judgment task. Vocabulary uniquely contributed 10% of this variance and phonological awareness contributed only 3%. Word meaning appeared to make a bigger contribution to sentence judgment than phonological awareness. On the other hand, they found that picture vocabulary and phonological awareness accounted for 37% in the production task. Vocabulary uniquely contributed 11% of this variance and phonological awareness contributed only 13%. Both predictors were important to performance on the production task. Overall, vocabulary and phonological awareness were important related variables to the morphological awareness tasks, whether focused on judgment or production.

Carlisle (2000) developed a test of morphological structure that included both a derivation task and a decomposition task. The stimuli only focused on derivations and included both transparent and opaque word pairs. She found that morphological awareness within a composite focused on morphological structure and meaning along with sight-word reading accounted for 43% variance in reading comprehension of third grade children, but only sight-word reading was significant. However, for fifth grade children, this morphological awareness composite and sight-word reading accounted for 55% with only the morphological structure measure contributing significantly to this outcome. Similarly, the morphological composite accounted for 41% of the variance in third grade vocabulary with only the sight-word reading performance accounting for significant variance. Finally, the morphological composite accounted for 53% of the variance in fifth grade with the sight-word reading remaining as the only significant variable in this equation. It is also important to note that a ceiling effect was found in the 5^{th} grade morphological awareness performance. In sum, the fifth grade children demonstrated more morphological awareness compared to third grade children. Furthermore, the morphological awareness of both third and fifth graders was related to reading achievement.

Deacon and Kirby (2004) utilized a sentence analogy task that was focused on simple present and past verbs (i.e., inflections only). They found that morphological awareness contributed less than one percent variance to sight-word reading in third grade children after phonological awareness, prior sight-word reading (i.e., assessed in second grade), and intelligence, but when morphological awareness was entered before phonological awareness, it's contribution increased to 1% variance. However, when prior sight-word reading performance was not added into the equation, current morphological awareness and phonological awareness performance both contributed 8% variance in third-grade sight-word reading. Similarly, morphological awareness contributed one percent to decoding after the other predictors including prior decoding performance, but this contribution increased to 2% when entered before phonological awareness in the performance of third grade children. When prior decoding ability was not entered into the equation, morphological awareness contributed 9% of the variance compared to the 10% of the variance contributed by phonological awareness. Finally, morphological awareness contributed 4% variance to third-grade reading comprehension beyond the other predictors including second-grade reading comprehension yet this variance was only 1% when morphological awareness was entered before phonological awareness. However, when second-grade reading comprehension was not added to the equation, morphological awareness contributed 8% and phonological awareness contributed 8% to reading comprehension. Deacon (2012) in a subsequent study also utilized a sentence analogy task focused on simple present and past verbs (i.e., inflections only) to assess morphological awareness in early elementary children. She found that morphological awareness made a relatively small and significant (e.g., 1%) contribution to sight-word reading compared to the other predictor variables phonological awareness and orthographic awareness with the entire equation accounting for 72% of the variance. Likewise, morphological awareness contributed 2% variance to decoding with the entire equation accounting for 75% of the variance.

Kirby et al (2012) also utilized a word analogy task that focused on 10 inflection and 10 derivation stimuli items with a mix of both transparent and opaque word pair relationships. Their findings indicate that morphological awareness in second and third grade added 4% and 5% variance, respectively, to sight-word reading after controlling for intelligence and phonological awareness. Similarly, morphological awareness performance in second and third grade contributed 3% and 4% variance in decoding, respectively, after the other predictors were controlled while morphological awareness performance in first grade also did not contribute any significant variance. Finally, second- and third-grade morphological awareness performance contributed 4% and 6% variance in reading comprehension, respectively after the other predictors were controlled. However, first-grade morphological awareness performance did not contribute significantly to any of these outcomes. In conclusion, the morphological awareness performance outcomes appeared to be impacted by the type of task used to assess them.

Assessment research: Summary. Given the aforementioned research summary, several common themes emerged from the evidence-base for the assessment of morphological awareness that informed the current project. First, there was wide variation in the (a) tasks and level of morphological processing assessed (i.e., production, judgment or analogy; Apel et al., 2013; Carlisle, 2000; Carlisle & Nomanbhoy, 1993), (b) grade level of development (Apel et al., 2013; Deacon, 2012; Deacon & Kirby, 2004; Nagy et al., 2006; Walker & Hauerwas, 2006), and/or (c) the abilities of the children studied (Apel & Lawrence, 2011; Peterson, Pennington, Shriberg, & Boada, 2009).

Second, although many researchers believed that morphological awareness began to play an important role in literacy in the later elementary and early middle school grades (Carlisle, 2000; Kuo & Anderson, 2006; Mahony et al., 2000; Nagy et al., 2006; Schweibert, Green, & McCutchen, 2002; Singson et al., 2000), the evidence now suggested that morphological awareness begins to play an important role in literacy in the early elementary school grades (Apel et al., 2013; Carlisle, 1995, 1996; Carlisle & Nomanbhoy, 1993; Nagy et al., 2003; Nunes, Bryant, & Bindman, 2006; Walker & Hauerwas, 2006; Wolter et al., 2009).

Third, morphological awareness appeared to work synergistically with other linguistic characteristics to support literacy performance (Carlisle, 2000; McCutchen, Green, & Abbott, 2008; Singson et al., 2000). Morphological awareness interacted with other forms of linguistic awareness including phonological awareness (Carlisle & Nomanbhoy, 1993; Deacon, 2012; Deacon & Kirby, 2004; Mahony et al., 2000; Schweibert, Green, & McCutchen, 2002; Singson et al., 2000) and orthographic awareness (McCutchen et al., 2008; Walker & Hauerwas, 2006). For example, children with stronger phonological awareness skill appeared to have some morphological understanding as well while morphological awareness may have supported the literacy skills in children with weaker phonological awareness (Carlisle & Nomanbhoy, 1993; Bryant et al., 1998; Deacon, 2012; Elbro & Arnbak, 1996; Mahony et al., 2000; Singson et al., 2000). Morphological awareness played a comparable role to phonological awareness in literacy development (i.e., reading; Deacon & Kirby, 2004); however, morphological awareness should not be considered as a substitute for phonological awareness because morphological awareness connects a word form and meaning beyond just the sound units (McCutchen et al., 2008). Although phonological awareness plays a crucial role in early literacy development, this importance shifted to other linguistic characteristics that support skill literacy (i.e., morphological awareness and orthographic awareness) as children encountered increasingly more demanding contexts (Schweibert et al., 2002; Singson et al., 2000)

Fourth, morphological awareness made a significant, unique contribution to literacy outcomes including: reading (Apel & Lawrence, 2011; Apel et al., 2012, 2013;; Carlisle, 1995, 2000; Carlisle & Nomanbhoy, 1993; Deacon, 2012; Deacon & Kirby, 2004; Kirby et al., 2012; Mahony et al., 2000; McCutchen et al., 2008; Nagy et al., 2006; Nunes et al., 2006; Roman, Kirby, Parrila, Wade-Woolley, & Deacon, 2009; Schweibert et al., 2002; Siegel et al., 2008; Singson et al., 2000; Walker & Hauerwas, 2006; Wolter et al., 2009), vocabulary (Carlisle, 2000; Nagy, Berninger, & Abbott, 2006; Schweibert, Green, & McCutchen, 2002), spelling (Apel & Lawrence, 2011; Carlisle, 2000; Nagy et al., 2006; Schweibert et al., 2002; Siegel, 2008; Singson et al., 2000; Walker & Hauerwas, 2006; Wolter et al., 2009), and comprehension (Carlisle, 2000; McCutchen et al., 2008; Nagy et al., 2006; Schweibert et al., 2002), beyond the significant contributions of other related variables (i.e., age, cognition, intelligence, vocabulary, phonological awareness, and orthographic awareness). More specifically, morphological awareness contributed to multiple levels of reading from word-level (i.e., real - sight-word reading and pseudoword - decoding) to text-level (i.e., reading comprehension) but appeared to have the strongest effect on reading comprehension (Apel et al., 2013; Carlisle, 2003; Deacon, 2012; Deacon & Kirby, 2004; Kirby et al., 2012; Kuo & Anderson, 2006). This evidence indicated that morphological awareness through the ability to recognize and segment morphemic units supported accurate, efficient word reading, which facilitates both faster pronunciation and text understanding (Carlisle, 2000; Carlisle & Stone, 2005; Elbro & Arnbak, 1996; Kirby et al., 2012; Nagy et al., 2006). The application of a morphemic strategy appeared effective in decoding both real and pseudowords, especially pseudowords that contain accurate morphemes like *meaningful* (Deacon, 2012; Deacon & Kirby, 2004; Kirby et al., 2012; Roman et al., 2000). In sum, morphological awareness was another important factor that should be included in the current models of literacy development as well as in early assessment and intervention (Apel & Masterson, 2001; Deacon, 2012; Kirby et al., 2012; McCutchen et al., 2008; Wolter et al., 2009). Therefore, additional research and evidence is needed to further explore the role of morphological awareness in literacy development.

In sum, morphological awareness appeared to be significantly related to reading at the word in sight-word reading and decoding and text levels, reading comprehension, spelling, and vocabulary as early as the elementary school years. Moreover, morphological awareness contributed unique variance to these outcomes even when other predictors were controlled (e.g., intelligence, age, phonological awareness, vocabulary). An important caveat to note was that much of this research used a composite of morphological awareness tasks that could be incorporating a wide variation of task demand and could be directly impact these results. Therefore, further research is needed to explore the impact of the individual tasks and stimuli on performance outcomes to create a clearer picture of the significant relationship between morphological awareness and literacy outcomes.

There were several limitations to note in the current evidence for morphological assessment. Most of the current assessment measures were researcher developed and have not been standardized. There was wide variability in the tasks required (e.g., how morphological awareness is elicited) and stimuli included (e.g., types of items included). Additionally, many of the current morphological awareness assessments required a spoken response only rather than including a written component. Also, most of the studies have focused on older school-aged children who have received some formal literacy instruction. When researchers have focused on younger children, significant floor effects have been found. These results indicated that the task was too difficult for these younger participants and was not sensitive enough to assess emerging skills. Even though there was evidence for the importance of linguistic units in the early development of word learning and literacy, there was a dearth of measures to assess these skills in young children. Moreover, using a dynamic approach to assessment of morphological awareness is necessary considering: 1) the metalinguistic nature of this skill and 2) this alternative approach may more adequately target this emerging skill in young children, and 3) this approach may provide an opportunity to explore the theories of word learning and literacy development further in the future.

Dynamic Assessment

Dynamic Assessment Defined

Dynamic assessment provides an alternative to traditional standardized testing and may provide an insight into emerging abilities of children developing a new skill such as morphological awareness in early elementary school. Dynamic assessment refers to a family of approaches that focus on children's receptivity to instruction or prompting rather than whether they can only produce an accurate answer or not. Dynamic assessment pairs performance and instruction by adapting to children's answer pattern to determine how this pattern may change when further instruction or support is provided. The purpose of dynamic assessment is to compare children's independent performance with their supported performance, which also assesses learning potential rather than just the learning products or what the children have learned to that point in time (Grigorenko & Sternberg, 1998). A supported performance can result in both a better performance and provide additional information regarding what supports are beneficial for children to maximize their educational performance. This assessment approach enables the examiner to observe the process of learning, which can supply more useful information about the children and how they learn and help to then guide further education and intervention. Because of a relative flexible nature, dynamic assessment may be more sensitive to emerging skill, and thus can benefit young children who often encounter floor effects on literacy assessments because of their emergent knowledge and performance. Finally, dynamic assessment has potential for early screening measures of emergent literacy so that children who are experiencing difficulties can be supported as early as possible to increase their ultimate success.

Theoretical Foundations

The theoretical foundation for dynamic assessment is deeply grounded in Vygotsky's (1978) theories on human development (Grigorenko & Sternberg, 1998). Originally, Vygotsky's dynamic approach to assessment proved useful in identifying struggling learners in the Soviet Union. At the core of the theoretical foundations of dynamic assessment is Vygotsky's concept of children's *zone of proximal development*

(ZPD). The ZPD can be operationalized as the range between the children's independent performance at one point in time and what can be achieved with the support of an adult model at a subsequent point in time. His approach started with a measure of initial performance followed by some teaching in a targeted domain that was then followed by a posttest measure of performance (i.e., a test-teach-test method). Children with a high degree of readiness, or a "broad" ZPD, demonstrated a strong benefit from the brief instruction provided while those with "narrow" ZPDs demonstrated a lesser degree of change in response to the instruction. Two children may appear to perform equivalently when assessed traditionally, but these same two children may demonstrate two different levels of responsiveness to intervention when adult guidance was provided producing an underlying contrast between the children's individual performances (Rothman, Semmel, & Gaylord-Ross, 1990). Ultimately, it was the measure of change in performance that appeared to be more predictive of future achievement than the initial, independent (i.e., pretest) performance (Campione & Brown, 1987). Grigorenko and Sternberg (1998) succinctly summarized Vygotsky's concept of the ZPD as, "…reflect[ing] development itself. It is not what one is but what one can become. It is not what has developed but what is developing" (p. 78).

Vygotsky's conclusions derived from the ZPD concept can be grouped into several pertinent categories: (a) maturation of cognitive functions, (b) development and learning, and (c) the independent abilities of children compared to what they can do with support (dependent on help from others). First, on the topic of the maturation of cognitive functions, Vygotsky concluded that cognitive abilities did not mature all at once but developed concurrently over time. Throughout development, children have some cognitive abilities that have matured while others are still in the maturing process. Grigorenko and Sternberg (1998) directly quote Vygotsky's original explanation of this as:

> …the state of development is never defined only by what has matured. If a gardener decides only to evaluate the mature or harvested fruits of the apple tree, he cannot determine the state of his orchard. Maturing trees must also be taken into consideration. The psychologist must not limit his analysis to functions that have matured. He must consider those that are in the process of maturing. If he is to evaluate fully the state of the child's development, the psychologist must consider not only the actual level of development but the zone of proximal development…. (p. 78)

In sum, this quote illustrates the need to not only assess what children have learned and then use to respond accurately to a static test stimulus, but also how children are learning in order to more fully and effectively assess the children's overall expression and understanding. Examining children's learning process can target skills that may still be progressing towards mastery or are just emerging. Both the products and the process of learning are important. The traditional approach to assessment only focuses on the products of learning, which does not produce a full estimate of a children's learning. A dynamic approach to assessment can look at both product and process to produce a more adequate approximation of children's learning and performance, which is particularly important for children who do not perform well in response to the traditional approach to assessment. Next, Vygotsky's points focused on development and learning as well as the comparison of the children's independent and dependent abilities both relate to the supposition that children can positively develop in response to direction and learning within this ZPD, (e.g., beyond what they know already and what they can learn with some assistance; Grigorenko & Sternberg, 1998). This topic stems from the construct of the ZPD, the zone between what children can perform and learn independently to what children can perform and learn with maximal assistance. According to Vygotsky, children's performance and learning can be maximized when instruction and experiences fall within the children's ZPD.

Types of Dynamic Assessment Approaches

Test—teach—test approach. One of the earlier-developed and most recognized approaches to dynamic assessment is the test-teach-test approach. In the traditional approach to assessment, the examiner is indirectly involved (i.e., recording responses) through strict guidelines to ensure a standardized delivery so that performance can be compared to the normative data. However, in the dynamic test-teach-test approach to assessment, the examiner has more direct involvement in the process through providing standardized instruction (Elliott, 2003; Missiuna & Samuels, 1989; Petersen & Gillam, 2015). In the test-teach-test approach, an initial test is followed by systematic instruction involving a targeted task then a final test is administered in order to explore potential learning effects. Based on Budoff's (1974) learning model, the children's posttest performance is seen as the most accurate representation of their performance.

As part of the test-teach-test approach, the goal of the brief systematic teaching is not to create immediate accuracy on the task targeted because the focus is not directly on performance, but on learning potential or performance change (i.e., the children's response to the instruction/ intervention; Embretson, 1987; Pena, 2000; Petersen, Allen, & Spencer, 2016; Spencer, Petersen, & Adams, 2015). The test-teach-test approach was originally developed for children who do not perform well on traditional assessment to provide a fairer testing situation for these children whose performance may be decreased due to limited experience based on their young age—because of emerging skill, cultural/linguistic background, and/or the presence of impairment (Bridges, 2009; Budoff, 1987; Cho et al., 2017; Daniel, 1997; Feuerstein, Rand, Hoffman, & Miller, 1980; Gutierrez-Clellen & Pena, 2001; Hasson & Joffe, 2007; Haywood & Tzuriel, 1992; Missiuna & Samuels, 1989; Peña & Gillam, 2000; Peña, Iglesias, & Lidz, 2001; Pena, Quinn, & Iglesias, 1992; Spencer et al., 2015; Sternberg, 1996). Moreover, this dynamic approach to assessment helps to address the failure of traditional assessment by also providing remedial information for children who struggle (Laughon, 1990). In conclusion, the current evidence supports the effectiveness of this dynamic test-teach-test approach to assessment; however, more exploration is still needed to fully establish the reliability and validity of this approach.

Graduated prompting approach. Given the aforementioned stated promise and benefit, it is important to explore beyond the general construct of dynamic assessment to more specifically determine the best approaches to use for conducting such assessment. One of the main approaches to dynamic assessment involves a graduated prompting method. The graduated prompting method of dynamic assessment is, "greatly influenced by Vygotsky's theory about learning and development and his notion of a zone of proximal development…[by utilizing a] gradual transfer of control of learning from the adult to the child" (Jitendra & Kameenui, 1993, p. 5; see also Campione & Brown, 1987). By targeting children's zone of proximal development, a better prediction of the children's readiness to learn or the benefits accrued from instruction can be made to inform educational planning and intervention (Jitendra & Kameenui, 1993).

In contrast to other dynamic assessment approaches, the graduated prompting approach seeks to measure the amount of assistance needed by children to assess their level of learning and/or learning potential according to Jitendra and Kameenui (1993). To this end, Campione, Brown, and Ferrara (1982) originally developed a dynamic measure designed with a sequential set of prompts to provide a range from minimal support to most supportive. The graduated prompting approach assesses the amount of assistance needed by children to accurately respond to the stimuli as cited by Jitendra and Kameenui. The graduated prompts administered depend on the children's response(s). If inaccurate responses are given, the subsequent prompt is administered using a branching method. If the children provide an accurate response, then the subsequent prompts are not needed. The children's level of learning and learning potential is seen as inversely related to the amount of prompting they need in order to respond accurately. This prompting system also shows the least amount of help that a learner needs in order to accurately respond, which is a relatively unique feature of this approach (Jitendra & Kameenui, 1993). Overall, "the remarkable achievement of [the graduated prompting approach] is in its creative quantification and standardization of the intervention and transfer stages, an achievement that has not been equally accomplished by [most of the other main dynamic assessment approaches]" (Grigorenko & Sternberg, 1998, p. 95).

The research involving the graduated prompting approach has focused on the learning and transfer of skills in varied ability groups and the comparison of static verses dynamic measures. Grigorenko and Sternberg conducted a systematic review examining the breadth of this research that resulted in two main findings. First, they developed and evaluated the psychometric properties of measures of learning and transfer in order to investigate the performance of different ability groups. Studies utilizing these measures indicated that children in lower achieving ability groups needed more support, more prompts, and did not transfer the skills learned in dynamic assessment as readily as their higher performing peers (Campione, Brown, Ferrara, Jones, & Steinberg, 1985; Day & Zajakowski, 1991; Ferrara, Brown, & Campione, 1986; Grigorenko & Sternberg, 1998). Grigorenko and Sternberg also cite Ferrara and colleagues' (1986) finding that younger children also appear to perform in a similar fashion when compared to older children.

Next, Grigorenko and Sternberg (1998) compared the outcomes of traditional static assessments to dynamic assessment using a graduated prompting approach with a specific focus on what does dynamic assessment provide above and beyond that of static assessment. Confirming their original hypothesis, "that the learning and transfer scores would provide information beyond that obtainable from static tests," A. Brown, Bryant, and Campione (1983) as cited by Campione (1989) found that the best, primary indicators of increased performance were the guided learning and transfer scores ($r \geq .60$) followed by the static ability scores that remained predictive but of lesser importance ($r \geq .45$), which is similar to the findings of other researchers (Day, Engelhardt, Maxwell, & Bolig, 1997; Resing, 1993; Speece, Cooper, & Kibler, 1990) as cited by Grigorenko and Sternberg (1998). Resing's (1997) findings showed that children with learning difficulties like those with diagnosed learning disabilities and those labeled as "slow learners" required double to triple the amount of hints, respectively, to achieve an accurate answer when compared to their mainstream peers. She found that the learning potential demonstrated on the dynamic assessment yielded additional qualitative information in regard to children's level of cognition (e.g., hints and strategies that led to success) over that provided by the traditional, static measures (as cited by Tzuriel, 2000). Additionally, the effects of dynamic assessment appear to reach beyond the immediate assessment to more long-term maintenance to several months following the initial assessment (Resing, 1993). Speece et al. (1990) found that the young children in their study could only be distinguished by their posttest dynamic assessment performance, not on the static measures as cited by Grigorenko and Sternberg (1998).

Dynamic Assessment of Language

Most of the evidence in support of dynamic assessment has been conducted in the field of psychology with a focus on the cognitive domain. However, research involving dynamic assessment approaches is emerging as a focus for research in other fields like speech-language pathology to gain a clearer picture of the language abilities of children, especially those who experience difficulties. Just as dynamic assessment was found to be more efficacious for children from minority backgrounds, with some adaptation to more language-based domains, dynamic assessment can also be used to identify children with language-based difficulties, including both language impairment and reading difficulties (Kester, Peña, & Gillam, 2001; Lidz & Pena, 1996; Olswang & Bain, 1996; Sharoni & Greenfeld, 1999; Tzuriel, 2000).

Some language and literacy researchers also have experimented with eliciting linguistic performance in children using a graduated prompting approach of dynamic assessment as a means to scaffold an accurate response. Most notably, Bridges and Catts (2011) developed a dynamic screening of phonological awareness in order to assess the phonemic awareness of kindergarteners to inform further educational programming. They assessed two samples of about 90 kindergarteners from three small schools in the Midwestern U.S. About half of these kindergarteners were typically-developing and the other half were designated as “at-risk” as indicated by their performance on the Initial Sound Fluency (ISF) and the Letter Sound Fluency (LSF) subtests of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good, Kaminski, Smith, Laimon, & Dill, 2003). The curriculum in all three schools contained a daily focus on phonological awareness, phonics, and/or word recognition instruction. The at-risk students were also provided a supplemental focus, either in a small group or individually, on the alphabetic principle in order to provide added support for their language/literacy performance. The researchers compared the Dynamic Screening of Phonological Awareness (DSPA; Bridges & Catts, 2010) to two separate static measures (i.e., the Static Deletion Task (SDT) administered to the first sample and the Initial Sound Fluency subtest (ISF) of the DIBELS to the second sample at the beginning of the kindergarten year).

The researchers found that the DSPA resulted in the most normal distribution of performance compared to the other measures (i.e., least amount of skew as demonstrated by the smallest quantity and the value closest to zero). This finding supported the idea that no floor effects were exhibited in the children’s performances on the DSPA; whereas, floor effects did appear to be present on the other measures administered in this study, including the largest floor effect exhibited in the children's Word Identification performance. These results were not surprising considering that the kindergarteners just started receiving formal instruction in an educational setting and were only beginning to develop their literacy skills. They also found that the DSPA predicted about 5-10% additional variance in the subsequent reading outcomes beyond that predicted by a static assessment targeting the same domain. Additionally, the researchers explored how accurately the DSPA classified students of varying reading ability. The DSPA identified a quarter of the sample compared to the original stated half of the sample that had been identified as at risk for literacy difficulties.

Both individually and in conjunction with the SDT, the DSPA was a statistically significant predictor of the Word Attack outcome measure. The finding that the DSPA was more predictive of the Word Attack performance outcome rather than the Word Identification was not surprising considering the closer correlation between the tasks targeted (i.e., reading of pseudowords) and the instruction provided in the curriculum (i.e., focused on phonological and/or phonemic awareness). They found that the children's performance on the dynamic screener significantly correlated with their level of reading achievement at the end of kindergarten indicating that this assessment could be beneficial (Bridges & Catts, 2011). The DSPA appeared to be helpful as a supplemental screening measure, but exhibited potentially similar results if used as a primary measure alone. In conclusion, the graduated prompting approach to assessment used by Bridges and Catts (2011) appeared to be an effective tool to measure early literacy skills in young children. The findings from the DSPA were promising and warranted further study.

Subsequently, researchers have begun to apply this graduated prompting approach to develop a dynamic assessment of morphological awareness rather than phonological awareness that will be further discussed in a subsequent section.

Dynamic assessment of morphology. Given the promise of dynamic assessment as a means to determine a range of skill, it is important to determine how it has been used and/or whether this alternative technique can also be used in assessing morphological awareness. Larsen and Nippold (2007) developed a dynamic assessment of morphological awareness in order to gain a deeper understanding of school-aged children's ability to provide an explanation of word meaning based on their analysis of a word's morphological structure (i.e., morphological analysis skill). Using a sample size of 50 sixth graders (19 boys and 31 girls) with an average age of 12;2 (ages ranged from 10;9 to 12;10), the researchers administered a hearing screening to confirm adequate hearing, a standardized measure of receptive vocabulary, and the morphological analysis measure developed by the researchers that was always administered by the first author. Additionally, the students' fifth grade State of Oregon standardized assessment (OSA; Oregon Department of Education, 2005) Reading and Literature scores were obtained and included a focus on the children's explicit and implicit comprehension, word knowledge, and awareness of figurative language (e.g., metaphors). The researchers systematically developed their Dynamic Assessment Task of Morphological Analysis (DATMA) from the work of Anglin (1993) and designed the assessment in four stages: (a) selected the words, (b) generated the assessment prompts, (c) wrote the administration script, and (d) devised a scoring system. The task included a set of fifteen stimuli items of derived words that were presented in a random order. Children were initially asked to define uncommon words made up of common base words and bound morphemes (e.g., *yellowing*). Based on their responses, it was determined whether they used morphological knowledge to provide the definitions of the uncommon derived words. If the children did not demonstrate the morphological ability to define the uncommon morphologically derived word, a corresponding graduated prompting schema was provided to facilitate a correct definition using morphological knowledge.

Based on three main research questions, the researchers tested the effectiveness of this newly designed measure (i.e., DATMA). First, they explored whether the DATMA obtained a range of morphological awareness student performance. Performance variability on the DATMA was significant with students' raw scores ranging from 23 (31% accurate) to 68 (91% accurate) out of 75 total points possible. Therefore, some of the students required multiple levels of supportive prompts while others relied more on their morphological skill with little need for the prompting. Secondly, the researchers explored the relationship between a student's word knowledge and reading comprehension through calculating correlations among the various task performances (i.e., DATMA, PPVT-III, and the OSA). All the calculations yielded moderately positive and statistically significant correlational coefficients ranging from $r = 0.65$ to $r = 0.36$. Overall, these results suggested that a better performance on the DATMA corresponded with higher word knowledge and reading comprehension performance in the students.

Following these findings, the researchers performed a couple post-hoc analyses to explore whether the students' morphological skill level (i.e., obtained by the DATMA) also had a general connection to their overall levels of literacy. The researchers combined the students' individual scores into a composite of overall literacy skill and then divided the sample into three separate groups (i.e., low, average, and high) based on this composite score. The results of a subsequent ANOVA with a post-hoc Tukey procedure revealed a significant main effect for group indicating statistically significant differences in performance (i.e., overall literacy levels) among the three groups, $F(2,47) = 131.79$, $p < .0001$, $\eta = .92$. The high literacy level group demonstrated the strongest literacy skills followed by the average group that outperformed the lowest achieving group. A second ANOVA with a post-hoc Tukey procedure was performed to explore the relationship between the children's overall literacy level and their performance on the DATMA resulted in another statistically significant group main effect, $F(2,47) = 9.30$, $p = .0004$, $\eta = .53$. These results confirmed that the children in the low literacy level group relied on more prompting to correctly define morphologically complex words when compared to the other literacy level groups. An additional analysis of the prompting use by literacy level group revealed, that although most students used some prompting, no ceiling effects were found as evidenced by the scoring means below the five-point total threshold (i.e., low—$M = 2.90$; average—$M = 3.40$; high—$M = 3.68$). Therefore, the DATMA resulted in a range of performance and DATMA performance correlated with overall literacy skill in this population.

Overall, Larsen and Nippold's (2007) study demonstrates that a morphological assessment can be dynamically designed to determine the level of support school-aged children (i.e., in sixth-grade) needed to develop morphological analysis skills related to word meaning. This study added to the evidence of a wide variation in the language development and task performance of typically-developing children. The range of performance demonstrated on morphological awareness tasks appears to be related to overall literacy success in this population of children (Larsen & Nippold, 2007; Nippold, Allen, & Kirsch, 2001; Nippold, Moran, & Schwarz, 2001a, 2001b). This evidence provided support for use of a dynamic assessment task to measure metalinguistic skill that is related to language and literacy development in children (Nippold, 2007).

Dynamic Assessment with Young Children

Wolter and Pike (2015) investigated whether a dynamic assessment of morphological awareness was an effective, sensitive measure of morphological performance in third graders and whether this performance would relate to the third graders' overall literacy skills. More specifically, the researchers modified and applied Larsen and Nippold's (2007) model to further investigate whether a dynamic assessment utilizing a graduated prompting approach obtained a range of performance that would parallel the emergent quality of morphological awareness expected in the early school-aged years. Additionally, the researchers were particularly interested in whether the students' performance on this dynamic measure correlated with subsequent literacy achievement (i.e., reading and spelling measures). Using a sample size of 54 typically developing third graders (28 boys and 26 girls), the researchers administered a battery of language and literacy measures including those focused on: phonemic awareness, receptive vocabulary, word-level reading, spelling, text-level reading comprehension, and the experimental dynamic prompting of morphological awareness. The researchers also systematically developed their Dynamic Assessment of Primary Morphological Analysis (DAPMA) from the work of Larsen and Nippold (2007). The task was designed to be administered as a set of 16 stimuli items of derived morphologically complex words in a random order that subsequently the children were asked to define followed by brief questions to ascertain whether their definition was based on the morphological information contained in the word. A corresponding graduated prompting schema was also developed to facilitate a correct definition and was administered until a correct definition was provided or all the prompts had been used. This prompting schema incorporated prompts focused on both morphological awareness (1-4) and vocabulary (5-6) strategies in order to facilitate an accurate word meaning response.

Like Larsen and Nippold's (2007) DATMA, the main aim of the DAPMA was focused on the children's abilities to discover the meaning of a morphologically complex word based on their knowledge of its components (i.e., base root word and affix—prefix or suffix). In selecting the stimuli, Wolter and Pike (2015) targeted low frequency derived words (i.e., to control for any possible familiarity with the word prior to the study), containing high frequency base root words (i.e., to help ensure some basic knowledge of the word meaning), as a way to avoid possible ceiling effects and emphasize the use of morphological ability overall. The researchers also controlled for the transparency of the change from base word to morphologically complex form making sure that no phonetic change had to be made when transitioning from the former to the latter. Based on expected performance, the researchers decided to only include derivative stimuli since third graders should demonstrate emerging knowledge and use of morphological derivations after having mastered inflections. After including 6 stimuli items from the original DATMA (Larsen & Nippold, 2007), 10 new stimuli items were created by the authors based on the criteria mentioned above (Wolter & Pike, 2015). During the task administration, the stimuli presentation utilized two formats of the target word (i.e., visual), written form and spoken, oral form, in individualized sessions of about 90 minutes' total over two to three instances. The children were asked to simply define the target word. When the children either answered incorrectly and/or the examiner was unsure of their response, the prompting schema was used to further probe the children's performance. The researchers posited that the graduated prompts were organized in order of assistance from minimal to maximal support.

Furthermore, the researchers developed specific scoring guidelines similar to those implemented by Larsen and Nippold (2007). Larsen and Nippold based their scoring system on whether the children mentioned both the meaning of the base root word and the change in meaning from the addition of the accompanying affix (e.g., prefix or suffix). Wolter and Pike (2015) followed this guideline with some important modification including a requirement that the definitions provided needed to be definitively appropriate and demonstratively different from the meaning of the base root word individually yet could still be correct when the children used more informal language expected of third graders. Unlike Larsen and Nippold, Wolter and Pike awarded an additional half-point bonus when the children could both clearly define the word and affix (e.g., prefix or suffix), change as well as use a direct reference to a component morpheme rather than using the direct reference to both morphemes in order to obtain the half-point bonus in Larsen and Nippold. The remaining point values, based on whether the children provided a subsequent definition, were awarded as follows: (a) implicit inquiry about the target's morphological structure (i.e., asking if the target contained separate parts followed by the original question of what the target word means for four points), (b) the prior prompt was revised to include what the separate word parts were and then asked what the definition was for three points, (c) the target word was presented in a sentence followed by a request for the definition for two points, (d) the target definition was presented with a multiple choice format for the children to select the correct word for one point, and (e) zero points for an incorrect definition. Although these prompts were designed and organized into a sequence of increasing word-learning support, the first set of prompts (prompts one through four) all focused on more of a specific morphological approach as compared to the latter prompts that focused on more of a vocabulary (i.e., semantic) approach.

The initial results indicated that all the children performed within the typical range on the standardized measures, which confirmed the participant inclusion criteria. Also, no significant gender difference in the DAPMA performance was found $t(52) = -0.81$, $p = 0.42$, allowing the researchers to analyze the results as a whole rather than in two separate subgroups. The children's performance on the DAPMA exhibited a range of performance. On average, the children obtained a raw score of approximately 38 points out of 88 total with a corresponding accuracy of 34% (*SD* = 11; *Range* = 3 points to 58 points). The skewness and kurtosis values indicated that the data was normally distributed with the majority of scores falling in the range of 40% to 60% accuracy for 30 students. Overall, these findings support the hypothesis that third graders demonstrate some morphological awareness skill and that a dynamic assessment measure allowed for the measurement of this emerging skill without evidence of a floor effect.

The researchers also explored the effectiveness of the graduated prompts and whether some prompts were more helpful compared to the others. As mentioned above, the first prompts (i.e., one to four) incorporated a morphological-orientation verses the vocabulary/semantic-orientation of the latter prompts (i.e., five and six). More specifically, Wolter and Pike (2015) analyzed the frequency of prompt administration and the children's benefit from each prompt. Variation in response to the prompts was noted. However, prompts three to six appeared to be the most effective while the first two prompts (i.e., the least supportive), the least effective. In general, the children provided correct target answers without prompting about 3% of the time. When the children struggled to independently define the targets, providing the morphologically oriented prompts focused on word parts resulted in an accurate response about 27% of the time. However, the final prompt level involving a multiple-choice format resulted in an accurate response about 80% of the time. A comparison of the prompting types revealed that the latter vocabulary-oriented prompts were significantly more helpful and resulted in 52% accuracy (*SD* = 20.68) than the morphologically oriented prompts that resulted in 35% accuracy (*SD* = 21.69), $t(53) = -3.25$, $p < .05$; $d = -.89$. In conclusion, the graduated prompts clearly provided important support for the children's performance on this emergent skill.

Using multivariate regressions, Wolter and Pike (2015) analyzed these correlations further to examine the predictive validity of the DAPMA in relation to the other literacy outcomes. For word-level reading focused on sight-words, phonemic awareness did not account for any significant variance while receptive vocabulary contributed 13% significant, individual variance. Moreover, the DAPMA significantly contributed to the individual variance in sight-word reading, but this contribution was quite small compared to the other factors, $r^2 = .01$; $F(1,51) = 4.29$, $p < .05$. This finding was not very surprising since the primary aim of the DAPMA focused on word meaning rather than decoding/reading, but interestingly it still significantly correlated. For reading/text comprehension, both phonemic awareness and receptive vocabulary contributed an adjusted significant unique variance (i.e., 16% and 18%, respectively). When the DAPMA was entered into the regression equation, the total variance rose to 54% with the DAPMA contributing 20% added variance beyond the former factors, $r^2 = .56$; $F(1,50) = 21.44$, $p < .01$. This finding signaled a much stronger relationship between the DAPMA and reading comprehension, which parallels the greater reliance on word meaning in order to effectively comprehend text. Likewise, both phonemic awareness and receptive vocabulary contributed significant, unique variance in the spelling outcome (i.e., 8% and 4% adjusted variance, respectively). Finally, the DAPMA contributed to this significant, individual variance, but this contribution was quite small compared to the other factors, $r^2 = .16$; $F(1,50) = 3.21$, $p < .05$. Similar to the word-level regression equation, this finding was not very surprising since the primary aim of the DAPMA focused on word meaning rather than reading/decoding, but interestingly it still correlated significantly.

The DAPMA proved effective in assessing early developing morphological awareness skills and how to support and enhance children's use of these skills for word acquisition. The relationship between the children's performance on the DAPMA and reading comprehension abilities was strong and significant after other important language skills like phonemic awareness and receptive vocabulary have been factored into the equation. No floor effects were seen in the children's DAPMA performance indicating that the stimuli items and task targeted emerging and developing skill rather than beyond the current skill level of the children. Additionally, the prompting reliance demonstrated by the children further supports the idea that morphological awareness was still emerging and developing in these children versus being readily applied as a mastered skill. Considering that the DAPMA significantly correlated strongly with reading comprehension as well as relatively less with word-level reading and spelling, the results of DAPMA have the potential to inform educational programming that maximizes literacy skill development, especially morphological awareness. In conclusion, even though these researchers have made great strides to assess and investigate the constructs of morphological awareness and dynamic assessment, much more still remains to be done to provide clarity of these constructs.

Dynamic assessment with young children: Task development summary. Taken together the aforementioned research informed the development of a dynamic morphological awareness task for young children, and the unique combination of the elements will be considered as follows (see Table 1). First, the graduated prompting schema of a developed task for young children should be based on the effective

Table 1

Sources for Components to be Included as a Dynamic Assessment Task for Young Children

	Component				Morpheme type			Change		Frequency	
Study	MA	GP	Sample	Task target	CP	Inf	Der	T	O	Hi	Low
Bridges & Catts (2011)[a]		X	Kindergarten	a. Segmentation - Word & phoneme	X						
Carlisle (1995)	X		Kindergarten to second grade	a. Production b. Judgment		X	X	X	X		
Carlisle (2000)	X		Third and fifth grade	a. Reading of MCWs; b. Morph structure - decomposition & production; c. Word meaning	X		X	X	X	X	X
Larsen & Nippold (2007)	X	X	Sixth grade	a. Word Meaning from Morph Structure			X	X		Root	MC
Wolter & Pike (2015)	X	X	Third grade	a. Word Meaning from Morph Structure			X	X		Root	MC

[a] Focused on phonological awareness.
MA = Morphological awareness.
GP = Graduated prompting approach.
CP = Compound words.
Inf = Inflections.
Der = Derivations.
T = Transparent change from base root to complex form.
O = Opaque change from base root to complex form.
MC = Morphologically complex word form.

prompting schema for phonological awareness developed by Bridges and Catts (2011) that was used with the same-aged children. Moreover, the individual subtasks may be modeled after the revised judgment and production tasks of Carlisle (1995) that were administered a sample of same-aged children. These subtasks, as noted, also were found to be effective with older children (Carlisle, 2000), but used a static, traditional method of assessment and thus dynamic assessment would provide a way to assess emerging knowledge of younger children. In addition, a dynamic assessment of morphological awareness should incorporate the work of Larsen and Nippold (2007) and Wolter and Pike (2015) who directly assessed morphological awareness using a dynamic assessment, albeit in older children. Larsen and Nippold used such an assessment approach to target knowledge of word meaning based on morphological structure but with older children (i.e., sixth graders). Similarly, Wolter and Pike extended these findings to another targeted population (i.e., third grade) who were still older than the population of the current study. Also, Wolter and Pike adapted the graduated prompting schema to include both prompts focused on a mix of morphological awareness (1-4) and vocabulary (5-6) strategies to support the morphological awareness in their sample that informed the prompt selection for the DMMA. Therefore, the current study was designed to combine elements of these studies to focus on the dynamic assessment of morphological awareness in young children.

Dynamic Assessment Evidence

Despite the promise and benefit of a dynamic approach to assessment, currently there are some limitations to note when implementing such approaches. First, a relative dearth of empirical evidence exists supporting the psychometric properties of dynamic assessment. Additionally, oftentimes a lack of sufficient methodological detail for replication also exists for the various approaches beyond the initial evidence from the researchers who developed these approaches (Grigorenko & Sternberg, 1998). Moreover, the amount of variation in the approaches within the over-arching construct of dynamic assessment is also paralleled in their outcomes. Finally, a dynamic approach to assessment also presents multiple challenges that directly impact the determination of efficacy and effectiveness of this promising assessment technique. In an attempt to address some of the concerns raised by Grigorenko and Sternberg, Swanson and Lussier (2001) systematically reviewed the current evidence with a particular focus on the origins of the performance changes evidenced through dynamic assessment and the performance variation from dynamic assessment. Ultimately, Swanson and Lussier (2001) sought to provide general empirical support of dynamic assessment and compare this approach to the traditional, static approach to assessment. In general, the researchers noted a wide variation in the details provided and studies within the pool of articles included in their meta-analysis. Most of the effect sizes reported within the studies were impacted by a multitude of factors as indicated by significant homogeneity "Q" statistics. Several significant main effects involving age, sample size, general type of dynamic assessment, and ability group were found (Swanson & Lussier, 2001).

The review of the evidence regarding dynamic assessment indicates several important findings in the areas of: type of scaffolding/feedback, comparison of a dynamic versus a static approach, and ability groups (Swanson & Lussier, 2001). First, Swanson and Lussier found that the dynamic assessment approaches involving scaffolds and strategy training both produced a moderately strong impact on performance (i.e., larger effect sizes; d = .48 and d = .65, respectively) when compared to the weaker impact of coaching studies (d = .21). These performance effect sizes were not directly impacted by the other study variables like participant ability group, number of assessment sessions, and/or the participants' ages. In addressing a critical challenge to dynamic assessment, the results of Swanson and Lussier's review also suggests that the changes in performance are connected to the assessment conditions rather than just the retesting situation (i.e., providing the same or similar stimuli repeatedly). Another meta-analysis conducted by Caffrey, Fuchs, and Fuchs (2008) also compared the results from studies involving contingent feedback (i.e., individualized instruction in response to student failure), and noncontingent feedback (i.e., standardized instruction in response to student failure). Although feedback correlated significantly and moderately to achievement in general, the authors concluded that non-contingent feedback (e.g., stronger correlation; r = .56), had a greater impact on academic performance than contingent feedback (r = .39; Caffrey et al., 2008). In summary, providing systematic, standardized feedback to participants seemed to be more effective in increasing performance rather than individualized feedback. This finding highlighted the potential benefit of a systematic, graduated prompting approach to dynamic assessment beyond other approaches to dynamic assessment.

Second, Swanson and Lussier (2001) found that when dynamic assessment was used to inform pre- and post-intervention, larger performance effects were obtained after children receive a corresponding treatment. More specifically, stronger effects appear to be produced from studies that compared static and dynamic performance of independent samples, i.e., separate samples ($d = 0.85$), rather than studies that compared the pretest and posttest performance of a single sample of participants ($d = .45$; Swanson & Lussier, 2001). The increase in the magnitude or size of the effects related to dynamic assessment was partially due to the added information it generates that can be utilized in treatment and planning as well as the better match to children's learning, both what has been learned and how it has been learned. However, both of these effect sizes indicated that dynamic assessment positively impacted performance even when confounding variables have been factored into the statistical analyses. Children who demonstrated the lowest performance on traditional, static measures often obtain higher learning potential scores (e.g., utilized scaffolds and support to yield accurate responses) that can provide insight into how they could potentially improve and how treatment could support their learning, which has implications as a measure of the learning process rather than just the product of the process (Hessels, 1997; Resing & Van Wijk, 1996; Tzuriel & Kaufman, 1999). Additionally, dynamic assessment appeared to more strongly correlate with academic achievement than traditional static measures (Caffrey et al., 2008). Thus, the overall evidence indicated that academic achievement was better predicted by learning potential scores than static scores (Tzuriel, 2000). Ultimately, dynamic assessment did appear to assess performance beyond that of the traditional, static approach, including under varied conditions and when effect sizes are corrected for statistical concerns (e.g., pretest sensitivity and upward biases; Grigorenko & Sternberg, 1998; Swanson & Lussier, 2001).

As such, a dynamic approach to assessment did provide added value beyond what can be provided by traditional, static assessment individually.

Finally, Swanson and Lussier (2001) found that dynamic assessment appeared to be more sensitive (i.e., typically resulting in an increase) to the performance of children from differing ability groups when compared to traditional, static assessment (Tzuriel, 2000). Ordinarily, students from varied ability groups (i.e., from culturally-linguistically diverse backgrounds (CLD), differing from the mainstream, majority culture, from varied socioeconomic (SES) levels based on family income and resources, with language/ learning difficulties, and those who are deaf) were at a disadvantage when administered traditional, static assessments (Tzuriel, 2000) as a result of inherent test bias and insensitivity to variation stemming from the standardization process. Additionally, younger children appeared to benefit more from dynamic assessment than older children yielding larger effect sizes in response to assessments (Swanson & Lussier, 2001). Dynamic assessment could potentially provide a more efficacious, effective option to measuring the performance these ability groups rather than static assessment. It is also important to note that although these ability groups seemed clearly delineated conceptually, oftentimes the practical distinction among these ability groups was not often viable because there is frequent overlap among them. Moreover, there was also a need to account for the synergistic impact on the children's performance based on this overlap. However, these groupings helped to give some general direction to a review of the evidence and support student learning.

Additionally, the evidence showed that young children from minority backgrounds, like CLD and low-SES, performed at a much lower level on a pre-test compared to children from the majority, mainstream backgrounds; however, using dynamic assessment, the children from these backgrounds appeared to “narrow the gap” and appeared to “catch-up” to their majority peers, which was evidence of a cultural difference not a disorder or impairment (Tzuriel & Caspi, 1992; Tzuriel & Kaufman, 1999; Tzuriel & Klein, 1985, 1987). Moreover, not only did children with developmental difficulties appear to capitalize on their skills in response to dynamic assessment by demonstrating stronger growth from pre- to posttest, but also when assessed longitudinally appeared to continue expanding on these skills at a two-week follow-up, when compared to a control group of age and cognitive-level matched peers (Reinharth, 1989). After finding that learning potential scores had significant additional predictive value for school performance, Resing (1993) concluded, “that the learning potential tests are of most importance when there are doubts about the children’s real intelligence level because of cultural background or disadvantaged educational history” as quoted by Tzuriel (2000, p. 404).

Furthermore, the evidence also suggested that dynamic assessment has the highest correspondence to the academic achievement of students with disabilities (Caffrey et al., 2008). Children with differing levels of ability appeared to demonstrate statistically significant differences in performance in response to dynamic assessment ($p < .01$); however, these performance differences are directly impacted by the domain (i.e., what is being assessed) and type (i.e., how this domain is being assessed). However, when both the assessment domain and type were incorporated into a regression model, the performance differences among ability groups were no longer statistically significant (Swanson & Lussier, 2001). Therefore, when the differences in performance related to both what is being assessed and how it is being assessed were accounted for, these differences among the varied ability groups were no longer statistically significant. Subsequently, the individual performances, regardless of group, became more comparable that cannot be done with performance on traditional, static assessments because of these aforementioned differences. In sum, the resultant equivalence of the varied ability groups' performance suggested that new abilities were being targeted (i.e., beyond the abilities targeted by traditional static assessments; Swanson & Lussier, 2001). Overall, dynamic assessment appeared to be more beneficial for children from a variety of ability groups than the traditional conventional assessments, especially for those who struggle with language/literacy.

Research Questions

In conclusion, the evidence suggested several main points. First, literacy is a metalinguistic process. Even though conclusive evidence exists to support phonological awareness as critical for literacy, other critical literacy skills beyond phonological skill need to be explored to support skilled development. From the perspective of literacy as a language-based process, the next logical language aspect to examine beyond phonology is morphology. Preliminary evidence supports that morphological awareness also is important to literacy development and success. Morphological awareness needs to be considered in the early literacy process because it helps to build and refine word learning.

Developmental norms of morphological awareness (i.e., in spoken language; inflections mastered by four years, two months old) indicate that this skill develops early in spoken language, yet little is known about the connection between this development and literacy. Also, it is important to establish how morphological awareness and literacy develop within young children who are typically performing to inform further exploration into this development in children who are at-risk or struggle so that the learning and achievement of all students can be supported and enhanced. Therefore, the investigation of the connection between morphological awareness and literacy should begin as early as possible in order to follow the growth of this skill as it becomes integrated into the literacy acquisition process. It was hypothesized that morphological awareness will correlate and contribute unique variance in literacy skills.

Second, most of the preliminary evidence has focused on older school-aged children and when studies have focused on young children significant floor effects were found. Therefore, evidence is still needed to support the connection between morphological awareness and literacy in young children. This participant population is targeted to examine literacy development as early as possible with little to no direct instruction that could confound the targeted skills and abilities. It is important to establish a baseline for typical performance in order to ultimately identify young children who struggle as early as possible so that support and intervention can be provided to increase their potential success. It was hypothesized that young children will exhibit morphological awareness and this awareness will correlate with other related literacy skill development.

Finally, the current method of traditional, static assessment limits adequate measurement of emerging morphological awareness and early literacy skills. Traditional, static assessments have resulted in significant floor effects when assessing morphological awareness in young children due to the emerging nature of morphological awareness and knowledge in this population. An alternative assessment approach needs to be considered in order to address these current limitations. A dynamic graduated prompting assessment approach has potential to be more sensitive to emerging skill in young children, especially including morphological awareness. The graduated prompting approach provides a dual benefit of dynamically assessing the young children's skill through gradually increased support within a standard protocol that can facilitate some comparison among individual performance. It was hypothesized that a dynamic measure of morphological awareness will be a feasible, valid and reliable instrument to assess this emergent skill in young children.

The main goal of the current study was to investigate the validity and reliability of a dynamic screening measure of morphological awareness (DMMA) for assessing morphological performance in young children. Furthermore, the current study was focused on also documenting the contributions morphological awareness makes to literacy skills. Therefore, the current study was focused on addressing the following research questions:

1. To what degree can performance on the DMMA be interpreted as a valid measure of morphological awareness in young children (i.e., 6- to 7-years-old)?
    a. To what extent does the DMMA represent the construct of MA?
    b. To what extent does the interpretation of the children's performance on

the DMMA relate to the performance on other language and language skills in young children?

c. Is the DMMA sensitive to performance differences within the emergent morphological awareness skill in young children?

2. What is the evidence that morphological awareness as assessed with the DMMA significantly contributes to early literacy skills (e.g., reading and spelling) above and beyond phonological awareness as was indicated in the literature?

   a. To what extent do phonological awareness and morphological awareness correlate with the early literacy skills (i.e., reading & spelling)?

   b. How do the unique, significant contributions made by phonological awareness and morphological awareness compare to each other?

The Pilot Project

Methods

To inform the wider implementation of the current dissertation study, a pilot project of the study procedures focused on an investigation of the feasibility of the methodology and procedures to assess the language and literacy skill of young children. Approval of the study and associated procedures by USU's Institutional Review Board as well as the appropriate informed consent paperwork was completed. The participants were recruited from The Bear River Charter School. A pilot sample of 14 children participated in the pilot. The mean age of the sample was 6 years, 2 months (*SD* = 4 months) with a range from 5;10 to 6;11. The gender distribution was 50% male and 50% female. The primary investigator collected and scored all the data for the pilot project. The methods implemented in the pilot project followed the methods of the wider study outlined below (refer to the Methods section of this dissertation study for further detail).

Results and Findings

Correlations were calculated between all the measures in the language/literacy battery. However, these correlations should be interpreted as potential trends and possible patterns in the data since the sample size was small. Nonetheless, a handful of significant correlations were found at both the .05 and .01 levels. The following correlations were significant at the $p < .05$ level: the vocabulary and general language ability ($r = .56$) as well as the DMMA-Total with the DMMA-RD ($r = .59$), spelling ($r = .56$), and sight-word reading ($r = .59$). The following correlations were significant at the $p < .01$ level: phonological/phonemic awareness and sight-word reading ($r = .71$), vocabulary and spelling ($r = .67$), and sight-word reading and decoding ($r = .71$). Moreover, two additional correlations approached significance: phonological/phonemic awareness with decoding ($r = .51$) and DMMA-EP with sight-word reading ($r = .56$). It was also interesting to note that while only a handful of these correlations reached significance those that did not still ranged from very small to very large—in quantity—and were most probably impacted by the lack of power directly effecting nonsignificance outcomes.

More specifically, prompt frequencies for each of the DMMA tasks were also calculated to investigate the patterns of use. On the expressive production task (DMMA-EP), the mean score obtained was 187 points out of 224 possible (*SD* = 21.06), which corresponds with a mean points percentage of 83% (*SD* = 9.40%). The mode of the scores received was a “4” that was obtained about 65% (*SD* = 13.17%) of the time with a mean of 37 out of 56 questions. The remaining score frequencies achieved were as follows (i.e., in order from largest to smallest): “3” with a mean of about 22% (*SD* =

7.76%) of the time, “0” with a mean of about 5% (*SD* = 4.57%) of the time, “2” with a mean of about 2% (*SD* = 2.66%) of the time, and “1” with a mean of about 2% (*SD* = 2.15%) of the time. On the receptive discrimination task (DMMA-RD), the mean score obtained was 198 points out of 228 possible (*SD* = 29.24), which corresponds with a mean points percentage of 87% (*SD* =12.82%). Similar to the expressive production task results, the mode of the scores received was a “4” that was obtained about 42% (*SD* = 13.42%) of the time (e.g., with a mean of 24 out of 57 questions). The remaining score frequencies achieved were as follows (i.e., in order from largest to smallest): “3” with a mean of about 22% of the time (*SD* = 10.63%), “0” with a mean of about 21% of the time (*SD* = 17.32%), “2” with a mean of about 10% of the time (*SD* = 5.86%), and “1” with a mean of about 5% of the time (*SD* = 4.50%). Although the DMMA-RD resulted in a slightly more variable performance, the findings indicated that the children demonstrated morphological awareness on both DMMA tasks. Overall, these findings suggested that the tasks were relatively achievable by this participant population with no floor effects.

In sum, as alluded in prior sections, the main purpose of the pilot project was to establish the feasibility of the wider study and related procedures. First, the pilot study confirmed that the targeted administration time and protocol was feasible with a sample similar to the targeted sample. However, during the pilot project it became apparent that the administration of the entire experimental tasks within one session could become unreasonable when multiple prompting was needed. Therefore, the decision was made to split these tasks into two separate, yet equivalent formats for the wider study. The same practice items were used on both formats; one format included questions 1–30 and the other with questions 31-60 for both the experimental tasks**.** Second, the pilot study provided an opportunity to develop and refine the answer sheets for the experimental measures to facilitate efficient, accurate recording of the participant's responses. Third, it also became apparent during the pilot study that although the measures of the language/literacy battery were randomized, care needed to be taken to off-set the administration of the standardized morphological awareness task using the Morphological Completion measure of the TOLD-P4 and the experimental expressive production task, the DMMA-EP, because of the similarity and potential confound of performance between these two tasks. Therefore, the examiner made sure that these measures were administered in different sessions in the current study, which also continued after the DMMA-EP was split into the two formats. Fourth, the pilot study also provided an opportunity to determine the feasibility of a modified response system for the experimental receptive discrimination tasks using a 5-point smiley Likert scale for the DMMA-RD. Although most of the children in the pilot sample did not have difficulty with this response scale, some difficulty was noted with the youngest participants. Following the pilot, the scale was revised to increase the distinct variation between the response choices that included more differentiation and correspondence with the answer choices—in both face design and corresponding background color provided as outlined below. In sum, the pilot study afforded an opportunity to determine feasibility and fine-tune the protocol and procedures in preparation for the wider-scale, current study.

Limitations

There were several limitations of note that directly impact the results of the pilot study: sample size and prompting administration. The sample size was limited due to the recruitment pool and scheduling challenges. Furthermore, some participants were difficult to test due to planned vacations and variations in attendance. Finally, the administration of each DMMA task in a single session was laborious when multiple prompting was required causing some potential fatigue, but all children adequately persisted through these administrations.

CHAPTER III

METHODS

Participants

A total of 78 first-grade children with a mean age of 6 years, 7 months with a range from 6;1 to 7;3 participated in the study from three elementary schools including the Edith Bowen Laboratory School, Wilson Elementary and Adams Elementary. A recruitment packet provided to the cooperating teachers was sent home with each of the first-graders at these locations. The packet included a letter providing an introduction and description of the study as well as two copies of the informed consent form, with one to sign and one to retain for parent/guardian records. The classroom teachers initially distributed the packets and gathered the returned forms that were then retrieved by the research team. Utah State University's Institutional Review Board approved the study and all associated procedures prior to initiation.

Demographics

Eighty children were initially recruited for the study with an attrition of two children, which left a final sample of 78 children that ranged in age from 6;1 to 7;3 (see Tables 2 and 3 for detailed demographics). This sample was almost equivalently split between 51% male (n = 40) and 49% female (n = 38). A total of 75 questionnaires were returned with variable completeness. Based on 61 questionnaire responses, the ethnic distribution of sample was comprised of: 87% White (n = 53), 11% Hispanic (n = 7), 7% Asian (n = 4), 5% Native Hawaiian/Other Pacific Islander (n = 3), 2% American Indian/

Table 2

Participant Characteristics: Demographics

Characteristic	n^a	Percent
Gender	78	
Male	40	51
Female	38	49
Ethnicity	61	
American Indian or Native Alaskan	1	2
Asian	4	7
Black or African American	1	2
Hispanic	7	11
Native Hawaiian or Other Pacific Islander	3	5
White	53	87
Primary Language Spoken	73	
English	72	99
Other	1	1
Highest level of education achieved by parent(s)	137	
Graduate/professional learning	33	24
Standard college or university graduation	47	34
Partial collegeb	26	19
High school graduate	21	15
Partial high school graduate—10th or 11th grade	9	7
Junior high school—including 9th grade	0	0
Less than 7th grade	1	1

[a]Varied due to response rate.
[b]At least 1 year of specialized training.

Native Alaskan and Black/African American (n = 1, for both). Based on 73 questionnaire responses, English was reported to be the primary/native language for all of the children (99%; n = 72), but one child (1%; n = 1). These demographics parallel those of the wider Intermountain West community where the study was conducted (U.S. Census Bureau, 2015a). The highest level of education reported by the parents (i.e., based on 137 parents) showed a definite skew towards more education than the general American population (U.S. Census Bureau, 2015b) with the majority of the parents

Table 3

Participant Characteristics: Educational and Literacy Experience

Characteristics	n^a	Percent
Preschool experience	71	
Yes	64	90
No	10	14
Length of preschool experience	61	
Less than 1 year	4	7
1 year	33	54
2 years	18	30
3 to 5 years	6	10
Literacy resources at home	72	
Very inadequate	1	1
Inadequate	1	1
Fairly adequate	5	7
Adequate	14	19
Very adequate	51	71
Independently literacy engagement at home	74	
Less than 1 a week	9	12
1 to 3 a week	25	34
3 to 6 a week	29	39
More than 6 a week	11	15

[a]Varied due to response rate.

having completed college followed by those who had completed graduate or professional learning. This skew reflected the fact that the study was conducted in a location with a major university. Based on 71 questionnaire responses, a majority of the children had participated in some preschool experience: 33 for 1-year (*n* = 54%), 18 for 2 years (*n* = 30%), 6 for 3 to 5 years (*n* = 10%) and 4 for less than 1 year (*n* =7%). Furthermore, based on 72 questionnaire responses, most of the children came from homes with literacy resources at the following levels: very adequate (*n* = 51), adequate (*n* = 14), fairly adequate (*n* = 5), inadequate (*n* = 1) and very inadequate (*n* = 1). Finally, based on 74 questionnaire responses, most of the children independently engaged in literacy activities at home: occasionally (i.e., 3 to 6 times a week; 39%), rarely (i.e., 1 to 3 times a week; 34%), frequently (i.e., more than 6 weeks; 15%), and almost never (i.e., less than 1 time a week; 12%). In summary, most of the children were acclimated and primed for literacy development based on several factors including their reported experience and access to resources.

Procedures

All measures were administered in a quiet location, one-on-one with a participant and an examiner, either the primary investigator or a trained research assistant with the exception of the Spelling Dictation Task that was administered in small groups of no more than four children. The trained research assistants were all second-year graduate students in a speech-language pathology program who participated in a training for test administration and had direct experience administering assessments to children in the targeted population. The measures were administered in a counterbalanced order to control for potential fatigue effect on performance. Specific care was taken not to administer the standardized morphological awareness measure, the Morphological Completion subtest of the TOLD-P4 (Hammill & Newcomer, 2008), in the same session as the experimental expressive production task, the DMMA-EP, because of the similarity between the tasks in order to avoid any potential confusion that could confound performance.

The original time frame of two to three separate sessions of approximately 30-45 minutes in length differed from the actual administration because the schedule had to accommodate the instructional and curricular demands required by the classroom teachers. The classroom teachers preferred that the children only be removed from the classroom for 20-minute segments. Therefore, the number of sessions for administration was recalculated to three to seven separate sessions of approximately 20 minutes, which was equivalent to the original time frame. The language and literacy battery was individually administered over and average of approximately five sessions for all children with a range from three to seven sessions. More specifically, the entire battery was administered to 15 children over three sessions, 29 children over four sessions, 16 children over five sessions, 8 children over both 6 and 7 weeks exclusively. The parents of one child made arrangements to administer the language/literacy outside of the classroom in two sessions of 60 minutes each. The administration sessions were conducted as consecutively as possible. Overall, the 5 sessions average was conducted over two consecutive weeks with some exceptions due to child absence and special event scheduling (e.g., a convention and family or holiday vacation days). The full research battery task administration was completed within one week for 13 children, 2 weeks for 40 children, 3 weeks for 17 children, 4 weeks for 5 children, and 5 weeks for one child. Scheduling disruptions like the holiday break when the schools were closed impacted the administrations of 3 weeks and more. When students were absent, these sessions were made-up at the next available time following the absence. Participants were provided a reward, including a small toy and/or pencil following each testing session. The primary investigator and trained research assistant(s) collected and scored all the data.

Language/Literacy Battery Measures

All participating children were administered a battery of language and literacy measures. The following were the areas targeted with the measures used: (1) general language ability—Sentence Imitation subtest (TOLD-P:4; Hammill & Newcomer, 2008); (2) general cognitive (nonverbal) ability—the Matrices subtest (K-BIT-2; Kaufman & Kaufman, 2004); (3) phonological/phonemic awareness—Elision subtest (CTOPP-2; Wagner, Torgesen, Rashotte, & Pearson, 2013); (4) word-level reading—Word Identification and Word Attack—(Woodcock Reading Mastery Test, 3rd ed.; Woodcock, 2011); (5) word-level spelling—Primary Spelling Inventory (PSI; Bear, Invernizzi, Templeton, & Johnston, 2008); (6) receptive vocabulary (PPVT-4; Dunn & Dunn, 2007); and (7) morphological awareness—the Morphological Completion subtest (TOLD-P4; Hammill & Newcomer, 2008). The Dynamic Measure of Morphological Awareness (DMMA; experimental) comprised of two separate tasks—receptive-discrimination and expressive-production. Apart from the Dynamic Measurement of Morphological Awareness tasks, these measures were selected because they have been established in the literature to measure the targeted literacy skills in young children. Additional detail is provided in the succeeding sections and in Tables 4, 5, and 6. A summary of the reliability and the administration characteristics of the measures in the language and literacy battery are provided in Tables 5 and 6, respectively.

Table 4

Language/Literacy Battery Measures

Area assessed	Subtest or skill	Instrument
General language ability	Sentence imitation	Test of Language Development—Primary, 4th Edition (TOLD-P:4)
General cognitive ability	Matrices	Kaufman Brief Intelligence Test, 2nd Edition (K-Bit-2)
Phonological awareness	Elision	Comprehensive Test of Phonological Processes, 2nd Edition (CTOPP-2)
Reading (word-level)	Word identification	Woodcock Reading Mastery Test, 4th Edition (WRMT-3)
	Word attack	Woodcock Reading Mastery Test, 4th Edition (WRMT-3)
Spelling (word-level)	Spelling dictation	Primary Spelling Inventory (Bear, Invernizzi, Templeton, & Johnston, 2008)
Vocabulary	Receptive vocabulary	Peabody Picture Vocabulary Test, 4th Edition (PPVT-4)
Morphological awareness	Morphological completion	Test of Language Development—Primary, 4th Edition (TOLD-P:4)
	Receptive discrimination	Dynamic Measure of Morphological Awareness (DMMA; Experimental)
	Expressive production	Dynamic Measure of Morphological Awareness (DMMA; Experimental)

General Language Ability

The Sentence Imitation subtest (TOLD-P:4; Newcomer & Hammill, 2008) was included in the language/literacy battery as a measure of general language ability. Sentence Imitation was a core subtest of the Test of Language Development—Primary, 4th Edition (TOLD-P:4; Hammill & Newcomer, 2008), that assessed how well children can accurately repeat sentences with increasing complexity, which were orally presented by the examiner. The core subtests of this instrument were normed for ages 4 years old to 8 years, 11 months old. The internal consistency measure for this instrument was

Table 5

Language/Literacy Battery: Measure Characteristics

Measure	Age range	Internal consistency/coefficient alpha	Test-retest reliability	Conclusions[a]
Sentence Imitation (TOLD-P:4)	4;0-8;11	0.92—for 6 yo (*SEM* = 1) 0.91—for 7 yo (*SEM* = 1)	*r* = 0.87-for 4-6 yo *r* = 0.88-for 7-8 yo	√
Morphological Completion (TOLD-P:4)	4;0-8;11	0.91—for 5 yo (SEM = 1)	*r* = 0.82-for 4-6 yo	√
Matrices (K-Bit-2)	4;0-90;0	0.78—for 5 yo (SEM = 5.8) 0.87—for 6 yo (SEM = 5.7)	*r* = 0.76-for 4-12 yo	√
Elision (CTOPP-2)	4;0-6;0	0.90 for 5 yo (*SEM* = 1) 0.92 for 6 yo (*SEM* = 1)	*r* = 0.93-for 4-6 yo	√
Word Identification (Sight-Word Reading) (WRMT-3)	2;0-80;0	0.99—for 5 yo (*SEM* = 1.50) 0.98—for 6 yo (*SEM* = 2.12)	*r* = 0.95-for 4-8 yo	√
Word Attack (Decoding) (WRMT-3)	2;0-80;0	0.94—for 5 & 6 yo (*SEM* = 3.67)	*r* = 0.89-for 4-8 yo	√
Spelling (word-level) (PSI)	5;0-10;0	Not standardized	Not standardized	n/a
Vocabulary (receptive) (PPVT-4)	6;0-90;0	0.97—for 5;0-5;5 (*SEM* = 3.3) 0.95—5:6-5;11 (*SEM* = 3.9)	*r* = 0.92-for 5-6 yo *r* = 0.93-for 7-10 yo	√

Note. yo = years-old(s); conclusions - √ = results in consistent performance across administrations with low test error; n/a = not applicable.

[a] This table summarizes the reliability of the measures in the language and literacy battery and the conclusions that can be reached based on these reliability outcomes.

Table 6

Language/Literacy Battery: Administration Characteristics

Measure	Administration	Practice items	Stimulus	Scoring	Basal	Ceiling
Sentence Imitation (TOLD-P:4)	Oral	2 total—only need one, if correct	None	1 = Correct 0 = Incorrect	None	5[a]
Morphological Completion (TOLD-P:4)	Oral	2 total—only need one, if correct	None	1 = Correct 0 = Incorrect	None	5[a]
Matrices (K-Bit-2)	Visual (e.g., from pictures)	3 total—for each item type	Stimulus book (pictures)	1 = Correct 0 = Incorrect	First 3 items correct	4[a]
Elision (CTOPP-2)	Oral	None—stimulus items used as practice	None	1 = Correct 0 = Incorrect	None	3[a]
Word Identification (Word-level Reading: Sight Words) (WRMT-3)	Visual (e.g., from words)	None	Stimulus book (words)	1 = Correct 0 = Incorrect	None	4[a]
Word Attack (Word-Level Reading: Decoding) (WRMT-3)	Visual (e.g., from pseudowords)	2 total—need both	Stimulus book (pseudowords)	1 = Correct 0 = Incorrect	None	4[a]
Spelling (word-level) (PSI)	Oral	None	None	1 = Correct 0 = Incorrect	None	None
Vocabulary (receptive) (PPVT-4)	Visual (e.g., from pictures)	None[b]	Stimulus book (pictures)	Total errors - minus ceiling item	Up to 1 error***	8+ [c]

[a] Consecutive incorrect.

[b]After a brief verbal introduction to the task - starting set used as practice items.

[c] Errors in a set

calculated across five age intervals. The average coefficient alpha for the Sentence Imitation subtest was reported as .92 for age six with a slight decrease to .91 for age seven (*SEM* = 1; for both). Additionally, the test-retest reliability coefficient for the Sentence Imitation subtest was $r = .87$ for the 4- to 6-year-old age group, yet only slightly increased to $r = .88$ for the 7- to 8-year old group. These findings suggested that the Sentence Imitation subtest results in mostly consistent performance across multiple administrations with a low level of test error.

This static subtest was administered orally without any additional materials required. Two practice items were included in this subtest, but the second practice item can be skipped if the children responded correctly to the first. An examiner then began scoring the children's responses from the first stimulus item administered. The children's responses were scored dichotomously using either one for a correct answer or zero for an incorrect answer. No basal rule was required, but a ceiling rule required a discontinuation of the administration following five consecutive incorrect answers.

General Cognitive Ability

The Matrices subtest (K-BIT-2; Kaufman & Kaufman, 2004) served as a measure of nonverbal, general cognitive ability. This nonverbal subtest assessed how well the children can complete patterns of increasing complexity from meaningful to abstract. The Matrices subtest was normed from ages 4- to 90-years old, which was the age range used to calculate the internal consistency measure for this instrument. The average coefficient alpha for this subtest (i.e., a nonverbal subtest) was reported as .87 (*SEM* = 5.7) for 6-year olds that increased to .89 for 7-year-olds (*SEM* = 5.6), which follows a general trend of increasing reliability with an increase in age. Additionally, the test-retest reliability coefficient for the nonverbal subtests was $r = .76$ for the 4- to 12-year-old group when adjusted for standard deviation range restriction. These findings suggested that the nonverbal subtest results in mostly consistent performance across multiple administrations with a relatively low level of test error.

This static subtest was administered orally with the use of a stimuli book. One practice item for each type of pattern was included in this subtest that oriented the children to the task (e.g., three total; but first two are the designated starting points for the 4-7 age group and the 8-and-older age group, respectively). An examiner began scoring the children's responses from the first stimulus item administered. The children's responses were scored dichotomously using either one for a correct answer or zero for an incorrect answer. The basal was obtained from the first three stimulus items passed and a ceiling rule required discontinuation of the administration following four consecutive incorrect answers.

Phonological Awareness

To measure phonological/phonemic awareness, the Elision subtest (CTOPP-2; Wagner et al., 2013) was administered in the language and literacy battery. This core subtest assessed the children's segmentation skill or how well the children can say a word after removing various parts including full-word segments in compound words and individual sounds (i.e., phonemes) in single words in later items. The core subtests of this instrument were normed for ages 4- to 6-years-old. The internal consistency measure for this instrument was calculated across 15 age intervals. The average coefficient alpha for the Elision subtest was reported as .92 for both 6- and 7-year-olds (*SEM*s = 1). Additionally, the test-retest reliability coefficient for the Elision subtest was $r = .93$ for the 4- to 6-year-old group, but decreased to $r = .77$ for the 7- to 11-year-old age group when adjusted for age range effects. These findings suggested that the Elision subtest results in consistent performance across multiple administrations with a low level of test error.

This static subtest was administered orally without any additional materials required. No practice items were included in this subtest. An examiner began scoring the children's responses from the first item administered. The children's responses were scored dichotomously using either one for a correct answer or zero for an incorrect answer. No basal rule was required, but a ceiling rule required discontinuation of the administration following three consecutive incorrect answers.

Word-Level Reading

To assess the children's word-level reading, two subtests from the Woodcock Reading Mastery Test, 3rd edition (WRMT-3; Woodcock, 2011)—Word Identification and Word Attack—were administered in the language and literacy battery. These subtests focused on the word-level reading of sight-words and pseudowords or decoding, respectively.

Word identification. This subtest assessed the children's ability to read sight-words in isolation, providing a progression from simple to complex words. This instrument was normed from 2-years-old to 80-years-old. The internal consistency measure for this instrument was calculated across seven age intervals. The split-half reliability coefficient alpha for the Word Identification subtest was reported as .99 (SEM = 1.50) for 6-year-olds with only a slight increase/decrease to .98 (SEM = 2.12) for 7-year-olds. These findings suggested that the Word Identification subtest results in consistent performance across multiple administrations with a low level of test error. This static instrument was administered orally with a stimulus book of words, either individually presented or grouped with one to three words on a line and up to three lines on a page. A word corresponded with a stimulus item. After a word stimulus was presented, the examiner asked the children, “What is this word?” No practice items were included on this subtest due to the natural task expectations. The children saw the letters and words on the page and were prompted to read a targeted word. The children’s responses were dichotomously scored as either a correct or incorrect answer with the overall number of errors totaled and subtracted from the highest ceiling item attained to calculate the children’s raw score. The examiner began with the administration of the first item on this instrument as a way to preempt the basal rule and thus potentially capture the best possible accurate performance for this research study. A ceiling rule required discontinuation of the administration following four consecutive incorrect responses and end with the last item on the bottom of the stimulus page.

Word attack. This subtest parallels the Word Identification subtest outlined above, but assessed the children’s ability to read pseudowords (i.e., decoding) in isolation, providing a progression from simpler to complex. This instrument was normed from 2-years-old to 80-years-old. The internal consistency measure for this instrument was calculated across seven age intervals. The split-half reliability coefficient alpha for the Word Attack subtest was reported as .94 (SEM = 3.67) for 7- and 7-year-olds. These findings suggested that the Word Attack subtest results in a consistent performance across multiple administrations with a low level of test error.

This static instrument was administered orally with a stimulus book of words, either individually presented or grouped with one to three words on a line and up to three lines on a page. A word corresponded with a stimulus item. After a word stimulus was presented, the examiner asked the children, “What is this word?” This subtest contained two practice items to orient the children to task expectations (i.e., read the stimuli items on the page). The children’s responses were dichotomously scored as either a correct or incorrect answer with the overall number of correct answers totaled to calculate the children’s raw score. For this subtest, it was recommended that an examiner begin with administering the first item on the instrument with all children, which preempted a basal rule. A ceiling rule required discontinuation of the administration following four consecutive incorrect responses and end with the last item on the bottom of the stimulus page. As alluded to above, the children’s raw score was calculated by subtracting the total number of errors from the highest ceiling item attained.

Word-Level Spelling

The Spelling Dictation Task utilized the Primary Spelling Inventory from the Words Their Way program (i.e., PSI; Bear et al., 2008) to assess the children’s spelling skill. The PSI was designed for children in kindergarten through third grade. The PSI was comprised of 26 stimulus items that progressively increase in difficulty from three letter words with short vowels (e.g., *fan*, *pet*) to single syllabic words with consonant clusters (e.g., *sled*) and long vowel patterns (e.g., *shine*) to multi-syllabic words with inflectional morphological patterns (e.g., *riding*). To administer this task, the examiner orally presented a target word in the following sequence: target word, a sentence using the target word, and the target word a final time. The children were expected to respond to the stimulus by writing the target word on their student response sheet. No basal was required. All administrations began with the first stimulus item as prescribed in the measure directions. However, no ceiling was applied as the latter stimuli include several of the inflectional morphemes targeted in the morphological measures (i.e., *-ed*, *-ez*, and –*ing*). The entire inventory was administered to each participant in order to provide a parallel, potentially stronger basis for comparison of performance. Responses were scored dichotomously as either correct or incorrect, as a general estimate of spelling skill.

Vocabulary

To assess the children's level of vocabulary understanding, the Peabody Picture Vocabulary Test (PPVT-4; Dunn & Dunn, 2007) was administered in the language and literacy battery. This instrument assessed the children's receptive vocabulary (i.e., how well the children understand vocabulary based on the children's age), which provides a progression from more general to specific words. This instrument was normed from two years, six months old to ninety years old. The internal consistency measure for this instrument was calculated across the ages from 2 years, 6 months old to 81 years old. The average coefficient alpha for Form A was reported as .96 (*SEM* = 2.8) for the 6;0—6;5 age range and increased slightly to .97 (*SEM* = 3.6) for the 6;6—6;11 age range, but remained at a similar level for seven-year-olds (i.e., .97; *SEM*—3.8). Additionally, the test-retest reliability coefficient for Form A was r = .92 for the five- to six-year-old age group and also increased slightly to r = .93 for the seven- to ten-year-old age group. These findings suggested that Form A results in consistent performance across multiple administrations with a low level of test error.

This static instrument was administered orally with a stimulus book of pictures. A set of four pictures corresponded with each stimulus item. After this set of pictures was presented, the examiner asked the children to "point to the target." Children's responses were dichotomously scored as either a correct or incorrect answer with the overall number of errors totaled and subtracted from the highest ceiling item attained (i.e., the last item in the ceiling stimuli set). It was recommended that an examiner begin administering the instrument with the items corresponding to the children's age level. A basal rule required that the children achieve no more than one error in a stimulus set of 12 items grouped by age, which can preempt the need for practice items by using words that are already familiar to the children. A ceiling rule required discontinuation of the administration following eight or more errors within one 12-item stimuli set.

Morphological Awareness

To assess the participants' morphological awareness, two total tasks were administered in the language and literacy battery: The Morphological Completion subtest (e.g., a standardized, static measure; TOLD-P4; Hammill & Newcomer, 2008) and the Dynamic Measure of Morphological Awareness (DMMA; experimental). The DMMA was made up of two experimental subtasks: The Receptive-Discrimination Task (DMMA-RD) and the Expressive-Production Task (DMMA-EP).

Morphological awareness—standardized: Morphological completion. Morphological completion was a core subtest of the Test of Language Development—Primary, 4th Edition (TOLD-P:4; Hammill & Newcomer, 2008) that assessed the children's ability to identify, understand, and use common morphological word components with a specific focus on the children's understanding of affixes (e.g., prefixes or suffixes). This core subtest was normed for ages 4 to 8 years, 1 months old. The internal consistency measure for this instrument was calculated across five age intervals. The average coefficient alpha for the Morphological Completion subtest was reported as .92 for 6-year-olds with a slight decrease to .90 for 7-year-olds (*SEMs* = 1). Additionally, the test-retest reliability coefficient for the Morphological Completion subtest was $r = .82$ for the 4- to 6-year-old age group, yet only slightly increased to $r = .85$ for the 7- to 8-year-old group. These findings suggested that the Morphological Completion subtest results in consistent performance across multiple administrations with a low level of test error.

This static subtest was administered orally without any additional materials required. Two practice items were included in this subtest, but if the children respond correctly to the first practice item, the examiner was instructed to proceed to the first stimulus item. The presentation of stimuli items utilized a cloze technique with the examiner reading a sentence missing a word at the end that the children were supposed to complete with the accurate morphologically complex component. Each stimulus item can be repeated once, if needed. The children's response was scored dichotomously using either a 1 for a correct answer or a 0 for an incorrect answer. No basal rule was required, but a ceiling rule required discontinuation of the administration following five consecutive incorrect answers.

Morphological awareness - experimental: The dynamic measure of morphological awareness (DMMA). The Dynamic Measure of Morphological Awareness was comprised of two experimental subtasks: The Receptive Discrimination Task (DMMA-RD) and the Expressive Production Task (DMMA-EP). The Receptive Discrimination Task (DMMA-RD) focused on a judgment of silliness based on the relationship between a targeted word pair that was either accurate (i.e., *makes sense*) or foil (i.e., *is silly*). The Expressive Production Task (DMMA-EP) focused on the production of a corresponding morphologically complex form to complete a stimulus that contained both the base root-word and a cloze sentence. These tasks are discussed in more detail in the following sections.

Receptive discrimination task (DMMA-RD): Procedures. The purpose of the DMMA-RD was to determine whether children who were developing literacy skills could use their knowledge of familiar base words and suffixes to distinguish whether a base word and morphologically complex word were related. Moreover, it was administered independently from the DMMA-EP because it required receptive-discrimination rather than expressive-production or generation, which was the primary target of the other subtask in the dynamic measure as outlined below. The DMMA-RD task was modeled after the revised judgment task of Carlisle (1995), the DATMA for sixth graders developed by Larsen and Nippold (2007) and the DAPMA for third graders by Wolter and Pike (2015). More specifically, the DMMA-RD task was adapted to utilize age-appropriate stimuli based on the selection criteria described in the following paragraphs. The number of stimuli was expanded to provide a more robust pool of potential stimulus items while also maintaining balance in morphological complexity across morpheme type and kind. Additionally, a response scale was designed and utilized to scaffold and structure the children's responses. The ultimate goal of these adaptations was to design a measure that was sensitive to assess emerging skill that may not be consistently mastered but was nonetheless emerging in development.

To facilitate the comparison between the two dynamic tasks of morphological awareness, the number of stimuli items on the DMMA-RD paralleled those used in the expressive-production task (outlined in a succeeding section). Stimuli were selected based on the following criteria: morphological complexity/type, word frequency, and transparency. First, as a way to control for basal and ceiling effects, stimuli of differing morphological complexity and type were included in this task. The task was composed of two practice items with one inflected form followed by a derived form and 60 total stimulus items, including 32 inflected forms (e.g., *-s* plural—all allophones, *-ed* regular past tense—all allophones, and *–ing* present progressive) equally divided between Form A (i.e., 15) and Form B (i.e., 17) with corresponding foil pairs (i.e., 4 on Form A and 6 on Form B; inclusively) and 28 total derived forms (e.g., *-er*, *-ly*, and *–y*) equally divided between Form A (i.e., 15) and Form B (i.e., 13) with corresponding foil pairs (i.e., 7 on Form A and 6 on Form B, inclusively). Second, all of the words within the word pairs utilized in the stimulus items were considered to be "high frequency" with a word frequency rating of 45 or higher according to Zeno et al. (1995). The inclusion of high frequency words increased the probability that the children will be familiar with a word and, therefore, attend more to the morphological nature (e.g., being able to flexibly use a targeted word by attaching additional word parts to it in order to modify the word's overall meaning) of the task rather than the initial semantic nature of the task (i.e., having to recognize a word and whether this word was within their mental lexicon or not). The targeted words used on the DMMA-RD had an average overall rating of 56.8 (*SD* = 5.3) with values ranging from 43.6 to 73.3. Finally, all of the stimulus targets whether inflected or derived contained a transparent form, with a clear, apparent relationship between the base root word and the morphologically complex word form (no phonological change; e.g., *egg* and *eggs*) because all the stimuli were presented orally.

Given the young age of children targeted for this task, a Likert scale, developed and piloted with kindergarten children (Quemart, Wolter, Deacon, & Chen, 2017) was used for this task in which five smiley face icons were superimposed on corresponding backgrounds to depict a range of the concepts of *making sense* and that of *silliness*. Each smiley face had specific characteristics to facilitate the distinction among response possibilities. On the left, a full smiley face indicated that the children think that the stimulus sentence "*makes sense*". A partial smile (i.e., on one side only) indicated that the children thought that the stimulus provided "*sort of makes sense*," but they were not completely confident in this answer. A straight face smiley indicated that the children "[*did*] *not know*" the answer or were not comfortable responding either way. A partial winking smiley face with a tongue sticking out indicated that the children think that the stimulus was "*sort of silly*," but they were not completely confident in this answer.

Finally, the small and large-eyed smiley face with a tongue sticking out indicated that the children thought that the provided stimulus was "*silly*."

In addition, colors were also used to help facilitate a range of choices. The traffic light colors of green, yellow, and red were used as these are commonly taught to young children to be associated with go, caution and stop, respectively. The color of dark green characterized by full saturation without transparency was paired with the "*makes sense*" choice and transitions to lighter green characterized by 50% transparency of full saturation for the "*sort of makes sense*" choice. Yellow with full saturation was paired with the "*don't know*" choice. Finally, a lighter red characterized by 50% transparency of full saturation for the "*sort of silly*" choice and transitions to a dark red characterized by full saturation without transparency for the "*is silly*" choice (see Figure 1). This response scale was developed by Quemart et al. (2017) expands the original dichotomous silliness answers of Carlisle (1995) of "*makes sense*" or "*is silly*" to provide for the possibility of a more equivalent scale to that used on the expressive task (e.g., allowing for prompting in

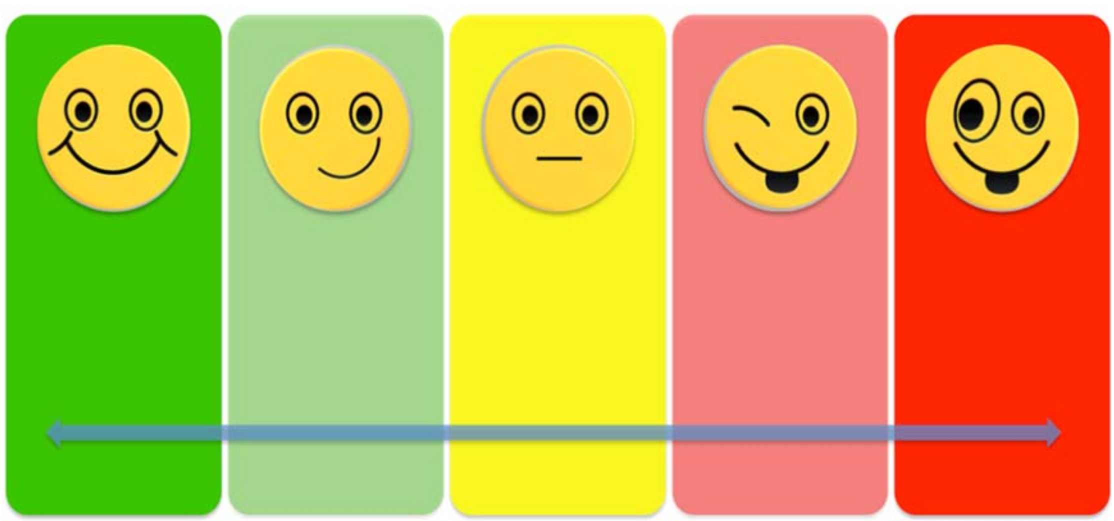

Figure 1. Likert smiley faces response scale for DMMA-RD.

multiple attempts). The DMMA-RD instructional protocol, practice items and stimuli are listed in Appendices A and B.

The DMMA-RD task involved the receptive-discrimination of the accuracy of the relationship between a base word and a morphologically complex word. For example, the examiner asked, “does this make sense or is it silly?” followed by a specific stimulus item like, “If there is more than one egg, there are eggs.” The children are then asked to discriminate whether the target words in the stimulus “*make sense*” (e.g., were accurately related; “If there is more than one egg, there are eggs.”) or “is silly” (e.g., were not accurately related; “If there is an arm, it is army.”). Administration of the DMMA-RD began with these two practice items to orient the child to the task and modeled an accurate stimulus containing an inflectional morpheme and a foil stimulus containing a derivational morpheme. Both practice items were required to orient the child to the complete task. If the child gave an incorrect response to an individual practice item, the examiner proceeded with the graduated prompting until either an accurate response was provided or all the prompts were administered. The general, non-contingent feedback was provided during the practice phase to encourage the child’s responses while not adding any additional instruction focused on the targeted task. This was consistent with the dynamic assessment evidence, which suggests that this type of feedback demonstrated a stronger relationship with performance (i.e., larger correlation; $r = .56$ vs. $r = .39$) rather than contingent, specific feedback (Caffrey, Fuchs, & Fuchs, 2008). The DMMA-RD instructional protocol, practice items and stimuli are listed in Appendices A and B.

Receptive discrimination task (DMMA-RD): Prompts. The DMMA-RD task began with a standard instructional protocol that introduced the task and acclimated the children to the task expectations (e.g., stimuli and targeted response formats; see Figure 1; Appendices A and B). This instructional protocol included two practice items, an inflected form that targeted plurality with the /z/-allophone; *egg* and *eggs* and a derived form that targeted –y; *arm* and *army*. Unique to the DMMA-RD, the practice items needed to include both an accurate relationship between the base root word and its morphologically complex word form or "*makes sense*" (i.e., "If there is more than one egg, there are eggs.") as well as an inaccurate relationship between the base root word and its morphologically complex word form or "*is silly*" (i.e., "If there is more than one arm, it is army.") so that the children were completely oriented to the task. Therefore, the inflected form was presented first and represented an accurate relationship. Subsequently, the derived form was presented second and represented an inaccurate relationship. Following this introduction, the DMMA-RD administration continued with the first stimulus item. The stimulus items were pseudo-randomized so that the stimulus items that targeted the same morphological ending were not consecutively sequenced. If the children did not respond correctly to the initial, static discrimination prompt for each individual stimulus, the dynamic portion of the instrument was implemented until either a correct response was given or all the dynamic prompts were administered.

The task prompts were designed to increase in explicit support to encourage correct responses (Bridges, 2009; Larsen & Nippold, 2007; Ram, Marinellie, Benigno, & McCarthy, 2013; Rubin et al., 1991; Wolter & Pike, 2015). The early administered prompts required more independence to accurately respond to compared to the latter prompts because the former only included verbal repetition and emphasis that was facilitative in spoken contexts, while the later prompts included visual support that also has been shown to facilitate word knowledge (Levin & Pressley, 1983; Paivio, 1983). The second prompt repeated the first prompt yet added verbal emphasis on the word pair (i.e., both the base root word and morphologically complex word individually) as the foundation for the targeted morphologically complex form; see Figure 2, which paralleled the first and second prompts utilized in the DSPA (Bridges, 2009; Bridges & Catts, 2011). The third prompt of the DMMA-RD differed from Bridges and colleagues by providing a visual cue for both the base root word and the morphologically complex form targeted. Providing a visual representation of the stimulus facilitates the basic recognition

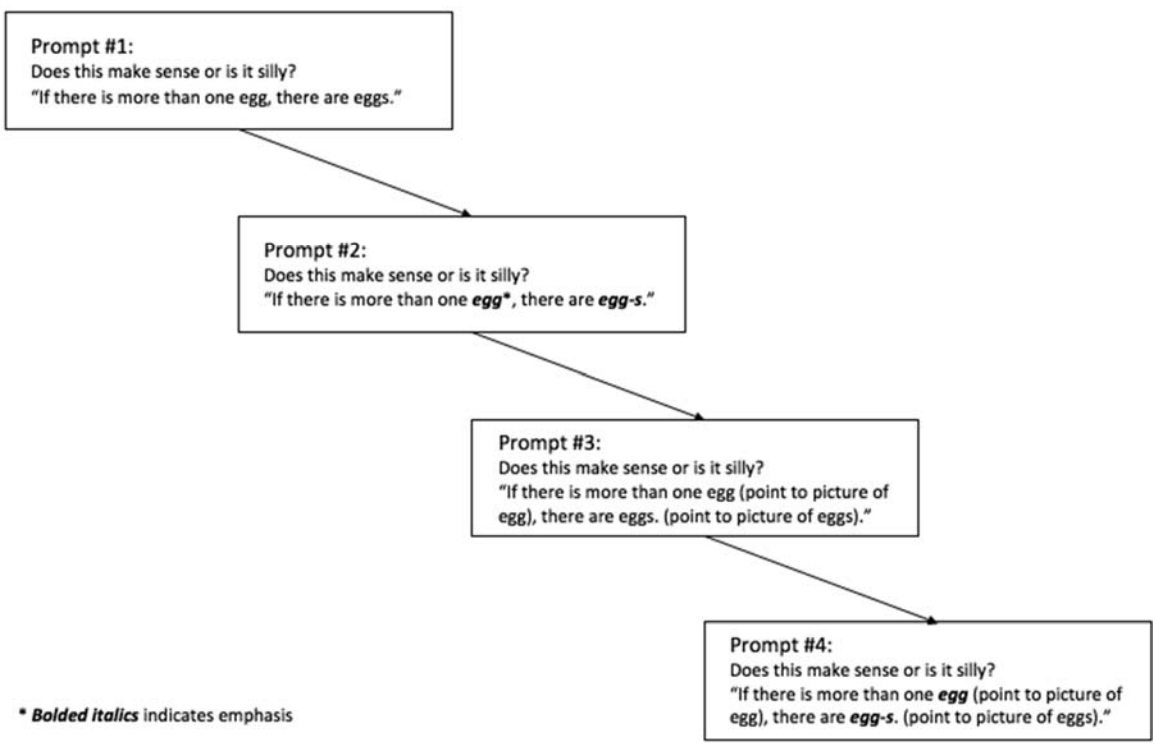

Figure 2. DMMA-RD prompting scheme.

of the word so that it can then be metalinguistically manipulated and learned (Levin & Pressley, 1983; Paivio, 1983). The fourth prompt combined the two prior prompts (e.g., the second and third prompts; emphasis and visual cues, respectively) as a final attempt to support the children's morphological skill to produce an accurate response. Unlike the DSPA prompting schema, the final answer was not included to maintain the graduated prompting and focus on the children's implicit morphological awareness (see Appendix A) rather than providing feedback that could facilitate additional learning and would have been more similar to a test-teach-test approach.

Therefore, the children's response on the first prompt that targeted static discrimination without additional prompting beyond the stimulus determined whether the subsequent prompts were to be administered. If the children answered incorrectly to the initial static discrimination prompt (e.g., "Does this '*make sense*' or '*is it silly*'? If there is more than one egg, there are eggs.") the second prompt was given. For the second prompt, the examiner repeated the original prompt with a "clue" of emphasis on the targeted word pair like, "If there is an *egg*, there are *egg-s*" (with italics indicating emphasis). If the children answered incorrectly to the second prompt, the third prompt was given. For the third prompt, the examiner repeated the original prompt with only the "clue" of two paired line drawings that corresponded to the base and its morphologically complex form, like, "If there is more than one egg—(Pointing to the picture of the egg), there are eggs—(Pointing to the picture of the eggs)." If the children answered incorrectly to the third prompt, the fourth, final prompt was given. For the fourth prompt, the examiner repeated the original prompt with "clues" of both the emphasis and the corresponding pictures, like, “If there is more than one *egg* - (Pointing to the picture of the egg), there are *egg-s* - (Pointing to the picture of the eggs).”

In an attempt to control for learning effects from repeated administration of the same stimulus item with prompting consecutively, the prompting was administered in waves. In the first wave, all stimulus items were administered with the first prompt. Next, in the second wave, the examiner re-administered the stimulus items that were incorrectly answered in the first wave with the second prompt. In the third wave, the examiner re-administered the stimulus items that were incorrectly answered in the second wave with the third prompt. Finally, in the fourth wave, the examiner re-administered the stimulus items that were incorrectly answered in the third wave with the final prompt, which provided one last chance to respond accurately. In sum, the waves followed the prompting levels with subsequent prompts only being administered until either an accurate response was provided or all the prompts were administered.

Receptive discrimination task (DMMA-RD): Scoring. Scores were awarded based on the level of prompting needed. The children received: a score of four if they responded correctly to the first prompt of the initial, static discrimination without additional support; a score of three for a correct response to the second prompt of emphasis only; a score of two for a correct response to the third prompt of visual only; a score of one for a correct response to the fourth prompt of both emphasis and visual; and a score of zero if no correct response was given (see Figure 3). A total of 240 points was possible on this task, which was divided equally between the forms—A and B, 120 points on each.

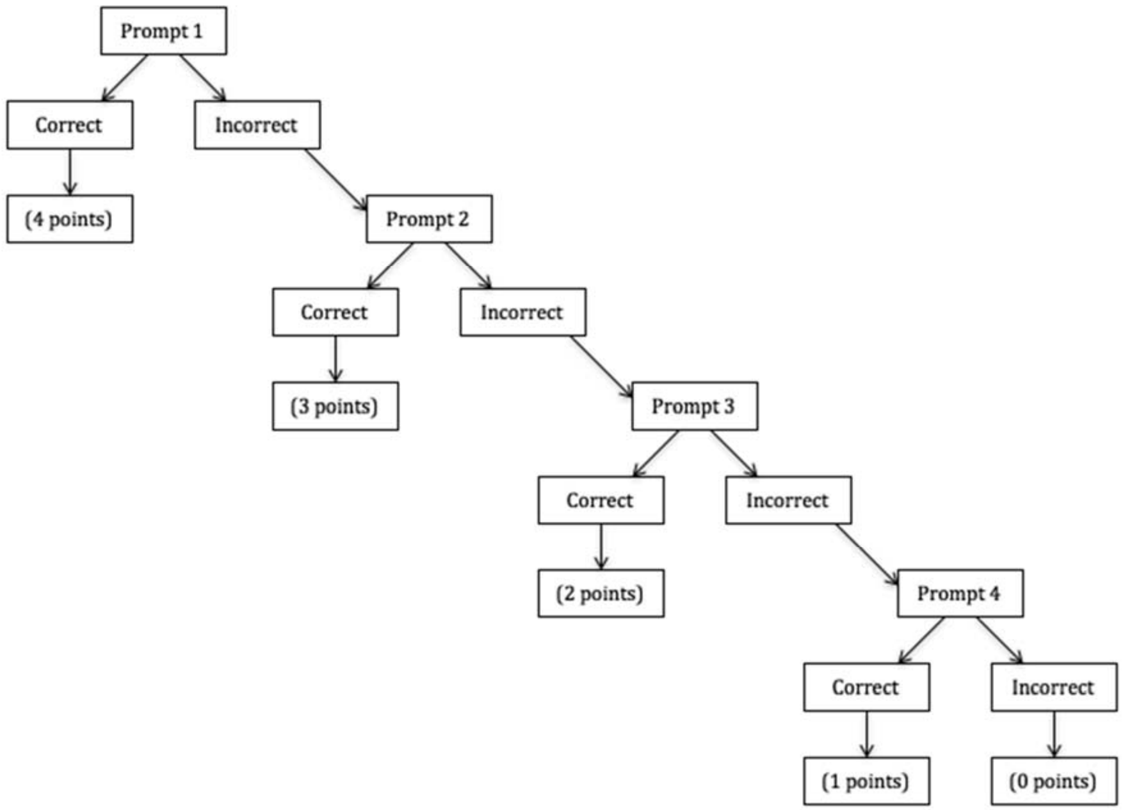

Figure 3. Flow chart illustrating the prompting and scoring of the Dynamic Measure of Morphological Awareness (DMMA).

Expressive production task (DMMA-EP): Procedures. This task was modeled after Carlisle's (1995) morphological production task and was originally included in the prompting schema in prior iterations of this dynamic measure (Wolter & Pike, 2015; Wolter et al., 2009). To facilitate comparison between the two dynamic measures of morphological awareness, the number of stimuli items paralleled those used in the receptive-discrimination task. Stimuli were selected based on the following criteria: morphological complexity/type, word frequency, and transparency. First, as a way to control for basal and ceiling effects, stimuli of differing morphological complexity and type were included in this task. The task was composed of two practice items (e.g., an inflected form followed by a derived form) and 60 total stimuli items, including 32 inflected forms (i.e., -*s* plural—all allophones, -*ed* regular past tense—all allophones, and –*ing* present progressive) equally divided between Form A (i.e., 15) and Form B (i.e., 17) and 28 total derived forms (e.g., -*er*, -*ly*, and –*y*) equally divided between Form A (i.e., 15) and Form B (i.e., 13). No foil pairs were needed in this task. Second, all of the words within the word pairs utilized in the stimulus items were considered to be "high frequency" with a word frequency rating of 45 or higher according to Zeno et al (1995). The inclusion of high frequency word pairs increased the probability that the children will be familiar with a word and, therefore, attend more to the morphological nature of the task (e.g., being able to flexibly use a targeted word by attaching additional word parts to it in order to modify the word's overall meaning) rather than the initial semantic nature of the task (e.g., having to recognize a word and whether this word is within their mental lexicon or not). The targeted words used in the DMMA-EP had an average overall rating of 57.9 (*SD* = 4.7) with values ranging from 41.4 to 72.6. Finally, all of the stimuli targets contained a transparent (e.g., inflected or derived) form, with a clear, apparent relationship between the base root-word and the derived word (no phonological change; e.g., *star* and *stars*) because all the stimuli were presented orally. The DMMA-EP instructional protocol, practice items and stimuli are listed in Appendices C and D.

The instrument of the confirmed test items was administered individually as a component of the language and literacy battery measures. The children were introduced to the task using a standard instructional protocol, which is included in Appendix C.

Following an introduction of two practice items that oriented the children, the task began with a traditional, static generation stimulus modeled after the tasks published by Carlisle (1995) and Wolter et al. (2009; 2015). Each stimulus targeted the generation of a morphologically complex word form that used the base root word provided in order to logically complete a cloze sentence. For example, the examiner read a base word stimulus, *star*, followed by a cloze sentence context, "The night sky is full of _____." The children are then asked to complete the sentence through providing the missing word (i.e., in this example, *stars*).

Expressive production task (DMMA-EP): Prompts. Paralleled the procedures of the receptive-discrimination task outlined above, the DMMA-EP task began with a standard instructional protocol that introduced the task and acclimated the children to the task expectations (see Appendix C). This protocol included two practice items, an inflected form (e.g., targeting plurality with the /z/-allophone; *star* and *stars*) and a derived form (e.g., targeting –y; *soap* and *soapy*). Following this introduction, the administration continued with the first stimulus item. The stimulus items were pseudo-randomized so that stimulus items targeting the same morphological ending were not consecutively sequenced. If the children did not respond correctly to an initial stimulus prompt, the dynamic portion of the instrument was implemented until either a correct response was given or all the dynamic prompts were administered. The task prompts were designed to increase in explicit support to encourage correct responses (Bridges, 2009; Larsen & Nippold, 2007; Ram et al., 2013; Rubin et al., 1991; Wolter & Pike, 2015). The early administered prompts required more independence to correctly respond compared to the latter prompts because the former only included verbal repetition and emphasis that was facilitative in spoken contexts while the later prompts included visual support that also has been shown to facilitate word knowledge (Levin & Pressley, 1983; Paivio, 1983). The second prompt repeated the first prompt yet added verbal emphasis on the word pair of both the base root word and morphologically complex word individually as the foundation for the accurate production of the targeted morphologically complex form (see Figure 4), which paralleled the first and second prompt utilized by the DSPA (Bridges, 2009; Bridges & Catts, 2011). The third prompt of the DMMA-EP differed from Bridges and colleagues (2010) by providing a visual cue for both the base root word and the morphologically complex form targeted. Providing a visual representation of the stimulus facilitates the basic recognition of the word so that it can then be linguistically

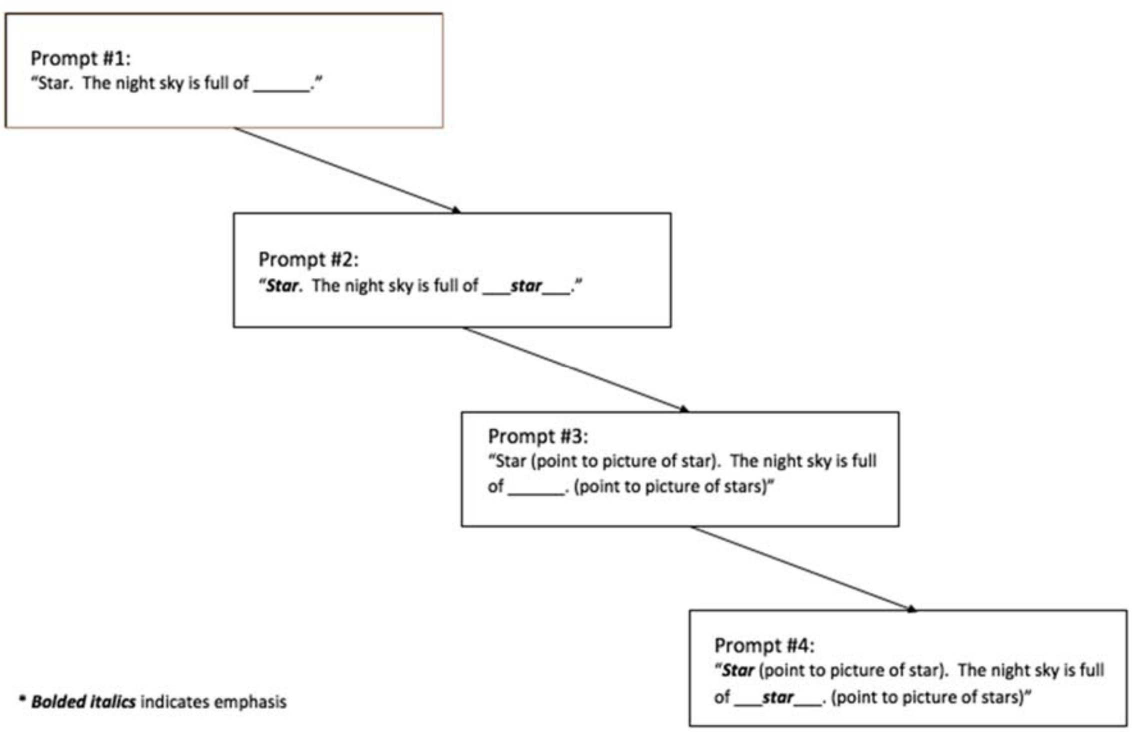

Figure 4. DMMA-EP prompting scheme.

manipulated and learned (e.g., used metalinguistically; Levin & Pressley, 1983; Paivio, 1983). The fourth prompt combined the two prior prompts (e.g., the second and third prompts; emphasis and visual cues, respectively) as a final attempt to support the children's morphological skill to produce an accurate response. Unlike the DSPA prompting schema, the final answer was not included to maintain the graduated prompting and focus on the children's implicit morphological awareness (see Appendix C) rather than providing feedback that could facilitate additional learning (e.g., more similar to a test-teach-test approach).

The children's response on the first prompt (e.g., static generation; without additional prompting beyond the stimulus) determined whether the subsequent prompts were to be administered. If the children answered incorrectly to an initial generation prompt (e.g., "Star. The night sky is full of ___."; expected response—*stars*), the second prompt was given. For the second prompt, the examiner repeated the original prompt with added emphasis as a "clue" to the original base like, "*Star*. The night sky is full of…___*star*-___" (italics indicates emphasis). If the children answered incorrectly to the second prompt, the third prompt was given. For the third prompt, the examiner repeated the original prompt with only the "clue" of two paired line drawings that corresponded to the base and its morphologically complex form, like, "Star. (Pointing to the picture of a star). The night sky is full of ___. (Pointing to the picture of the stars/night sky)." If the children answered incorrectly to the third prompt, the fourth, final prompt was given. For the fourth prompt, the examiner repeated the original prompt with the "clues" of both emphasis and corresponding pictures to the original base like, "*Star*. (Pointing to the picture of a star). The night sky is full of…___*star*- ___. (Pointing to the picture of the stars/night sky).” The pilot data of the original instrument suggested that these prompts did elicit a response from young children (e.g., 5-year-olds) who demonstrated a mean score of 40.50 out of 49 possible points with a standard deviation of 5.9 (Wolter & Pike, 2015).

Additionally, in an attempt to control for learning effects from repeated administration of the same stimulus item with prompting consecutively, the prompting was also administered in waves. In the first wave, all stimulus items were administered with the first prompt. Next, in the second wave, the examiner re-administered the stimulus items that were incorrectly answered in the first wave with the second prompt. In the third wave, the examiner re-administered the stimulus items that were incorrectly answered in the second wave with the third prompt. Finally, in the fourth wave, the examiner re-administered the stimulus items that were incorrectly answered in the third wave with the final prompt, which provided one last chance to respond accurately. In sum, the waves followed the prompting levels with subsequent prompts only being administered until either an accurate response was provided or all the prompts were administered.

Expressive production task (DMMA-EP): Scoring. Paralleling the DMMA-RD, scores were assigned based on the level of prompting needed. The children received: a score of four if they responded correctly to the first prompt of the initial, static generation without additional support; a score of three for a correct response to the second prompt of emphasis only; a score of two for a correct response to the third prompt of visual only; a score of one for a correct response to the fourth prompt of both emphasis and visual; and a score of zero if no correct response is given (see Figure 3). A total of 240 points was possible on this task that were divided equally between the forms—A and B with 120 points on each. A blinded reviewer rescored 30% of the protocols. The outcome of this review is discussed in the results section.

CHAPTER IV

RESULTS

Descriptive Statistics

The descriptive statistics (i.e., mean, standard deviation, range, skewness, and kurtosis) for the entire study sample are listed in Table 7 (except for the morphological awareness measures that are discussed below). The Sentence Imitation and Elision measures used scaled scores to compare performance to same aged peers with a mean of 10 and a standard deviation of 3 with an average range from 7 to 13. The Matrices, Vocabulary, Word Identification, and Word Attack measures used standard scores to compare performance to same aged peers with a mean of 100 and a standard deviation of 15 with an average range from 85 to 115.

Table 7

Language/Literacy Battery: Descriptive Statistics—Total Sample (N = 78)

Measure	Mean	*SD*	Min	Max	Skewness	Kurtosis
Age in months	79.32	3.87	73	88	0.257	-1.020
Sentence Imitation	10.46	2.98	3	16	-0.206	-0.515
Elision	9.73	2.98	1	17	-0.278	1.081
Vocabulary	110.50	12.79	64	138	-1.192	2.202
Word Identification	101.51	15.96	68	143	0.358	-0.025
Word Attack	105.68	13.56	78	142	0.192	-0.139
Spelling[a]	7.17	3.96	0	24	1.501	4.218
Matrices	95.69	12.69	75	126	0.484	-0.744

Note. Skewness—*SE* = 0.272, Kurtosis—*SE* = 0.538.

[a]Raw score.

The descriptive statistics were further analyzed to explore possible performance or gender differences in the children's language and literacy. The performance groupings were based on the Sentence Imitation measure. The literature base supports the use of the Sentence Imitation measure as an effective tool to classify general language ability and language impairment (Archibald & Joanisse, 2009; Conti-Ramsden, Botting, & Faragher, 2001; Oetting, McDonald, Seidel, & Hegarty, 2016; Seeff-Gabriel, Chiat, & Dodd, 2010). Morphological awareness as a metalinguistic component of language ability theoretically relates to the general language ability assessed by the Sentence Imitation measure. On the Sentence Imitation measure, the conventional cut-off of a scaled score of 7 or below that corresponds to one standard deviation and below the mean of 10 was used to identify children with weak, impaired language skill. For the current study, children who achieved a scaled score of 8 or above were classified as typically performing while those who achieved a scaled score of 7 or below were classified as at-risk for language and literacy difficulty. It is important to note that the results for the at-risk group should be interpreted with caution because of the small number of children (n = 13) from the current sample that were categorized into this subgrouping. The size of the at-risk group was much smaller than the generally acceptable minimum of 30. Additionally, females tend to develop language earlier than males. Not considering these subgroupings could impact the generalization of the findings to other samples and/or the general population. Therefore, these subgroupings were considered in subsequent analyses. The descriptive statistics for the performance groups are listed in Table 8 and for the gender groups in Table 9.

Table 8

Language/Literacy Battery: Descriptive Characteristics—By Performance Group with t*-Test Results*

	Typically performing group (n = 65)					At-risk group (n = 13)					Comparisons		
Measure	*M*	*SD*	Range	Skewness	Kurtosis	*M*	*SD*	Range	Skewness	Kurtosis	*t*	*p*	*d*
Age in months	79	4	73-88	0.29	-0.98	79	4	73-86	0.12	-1.20	0.16	.87	0.03
Sentence imitation	11	2	8-16	0.17	-0.87	6	1	3-7	-1.12	0.44	8.19	< .001	2.89
Elision	10	2	5-17	0.59	0.30	7	3	1-11	-0.37	-1.41	4.44	< .001	1.17
Vocabulary	113	10	86-138	-0.55	0.64	98	18	64-120	-0.68	-0.54	4.37	< .001	1.07
Word identification	101	16	68-143	0.28	-0.05	92	13	70-119	0.43	0.59	2.27	.03	0.58
Word attack	106	14	78-142	0.17	-0.09	100	12	78-109	0.09	-0.62	1.70	.11	0.50
Spelling[a]	8	4	1-24	1.47	3.72	5	2	0-9	-0.61	0.62	2.53	.02	0.63
Matrices	97	13	75-126	0.38	-0.93	91	9	78-109	0.53	-0.56	2.07	.04	0.69

Note. Skewness *SE*—Typically Performing Group = 0.297, At-Risk = 0.616.

Kurtosis - *SE*—Typically Performing Group = 0.586, At-Risk = 1.191.

[a]Used raw score.

Table 9

Language/Literacy Battery: Descriptive Characteristics—By Gender with t*-Test Results*

	Male (n = 40)					Female (n = 38)					Comparisons		
Measure	*M*	*SD*	Range	Skewness	Kurtosis	*M*	*SD*	Range	Skewness	Kurtosis	*t*	*p*	*d*
Age in months	80	4	74-86	0.21	-1.20	79	4	73-88	0.32	-0.79	0.71	.48	.16
Sentence imitation	11	3	4-16	-0.37	-0.59	10	3	3-16	-0.51	-0.31	0.57	.57	.13
Elision	10	3	2-17	-0.08	0.23	9	3	1-15	-0.97	2.85	0.86	.39	.20
Vocabulary	113	12	71-138	-1.46	3.24	108	13	64-127	-1.09	2.26	1.73	.09	.39
Word identification	100	16	70-143	0.78	0.58	100	16	68-133	-0.04	-0.43	0.85	.40	.19
Word attack	106	14	83-142	0.52	-0.16	104	13	78-128	-0.33	-0.37	1.10	.28	.25
Spelling[a]	7	5	0-24	1.73	3.73	7	3	1-13	-0.02	-0.27	0.19	.85	.04
Matrices	95	13	78-126	0.91	-0.05	97	13	75-121	0.04	-1.23	-0.80	.43	-.18

Note. Skewness—*SE*—Male = 0.374, Female = 0.383.
Kurtosis—*SE*—Male = 0.733, Female = 0.750
[a]Used raw score.

Mean Performance

In Table 7, the mean performance scores for the entire sample were within the average range of performance (i.e., scaled score 7 - 13, standard score 85 - 115) on the all tests. The entire sample achieved the highest standard scores on the Vocabulary measure and the lowest standard scores on the Matrices measure. The typically performing group achieved the highest standard scores on the Vocabulary measure and the lowest standard scores on the Matrices measure. Also, this typically performing group performed within the average range on the Elision measure (see Table 8). Moreover, their performance on the two word-level reading tests (Word Attack and Word Identification) were quite similar to each other. The at-risk group also performed within the normative standard score expectations (i.e., in the average range from 85-115) on the Vocabulary, Word Identification, Word Attack, and Matrices measures. The at-risk group's word-level reading performance was comparable with only eight points separating their mean sight-word reading performance and their mean decoding performance. The mean performance of the at-risk group on the Sentence Imitation and Elision was just below the normative average range of performance. Both gender groups achieved the highest score on the Vocabulary measure and the lowest score on the Spelling measure.

Standard Deviation and Variability

The performance on the scaled score measures for the entire sample was similar. Their performance on the Word Identification measure exhibited the most variability of the standard score measures. The standard deviations of the typically performing group's scaled score performance were comparable. Like the entire sample, the Word

Identification measure demonstrated the most variation of the standard score measures for this group. The at-risk group exhibited a more variable performance on the Elision measure versus the Sentence Imitation measure since both use scaled scores to report performance. Similarly, the at-risk group's performance on the Vocabulary measure demonstrated the most variability of the standard scores measures. The two gender groups demonstrated the equivalent variability on the Elision measure and the Sentence Imitation measures since both are based on scaled scores. Similarly, both gender groups exhibited the most variability on the Word Identification measure compared to the other standard score measures.

Skewness

The skewness statistic provides information on the symmetry of the distribution; this is an important consideration for measurement validity as well as an assumption in statistical calculation (e.g., Pearson's product-moment correlation). A skewness statistic of 0 indicates that the data are symmetric. Skewness statistics from -0.5 to 0.5 are considered to be "approximately symmetric," which is considered to be an acceptable level of skew to be considered "reasonably normal." If a skewness statistic falls between -1 and -0.5 or 0.5 and 1, skewness statistic indicates a clustering of performance towards lower or higher values, respectively, and is considered moderately skewed. Skewness ≥ 1 or ≤ -1 is considered to be highly skewed.

Histograms of the performance distributions with a skew greater than ± 0.5 by performance group are included in Figure 5. Both positive and negative skew were found in the sample performance. Several outliers were also noted in these distributions that

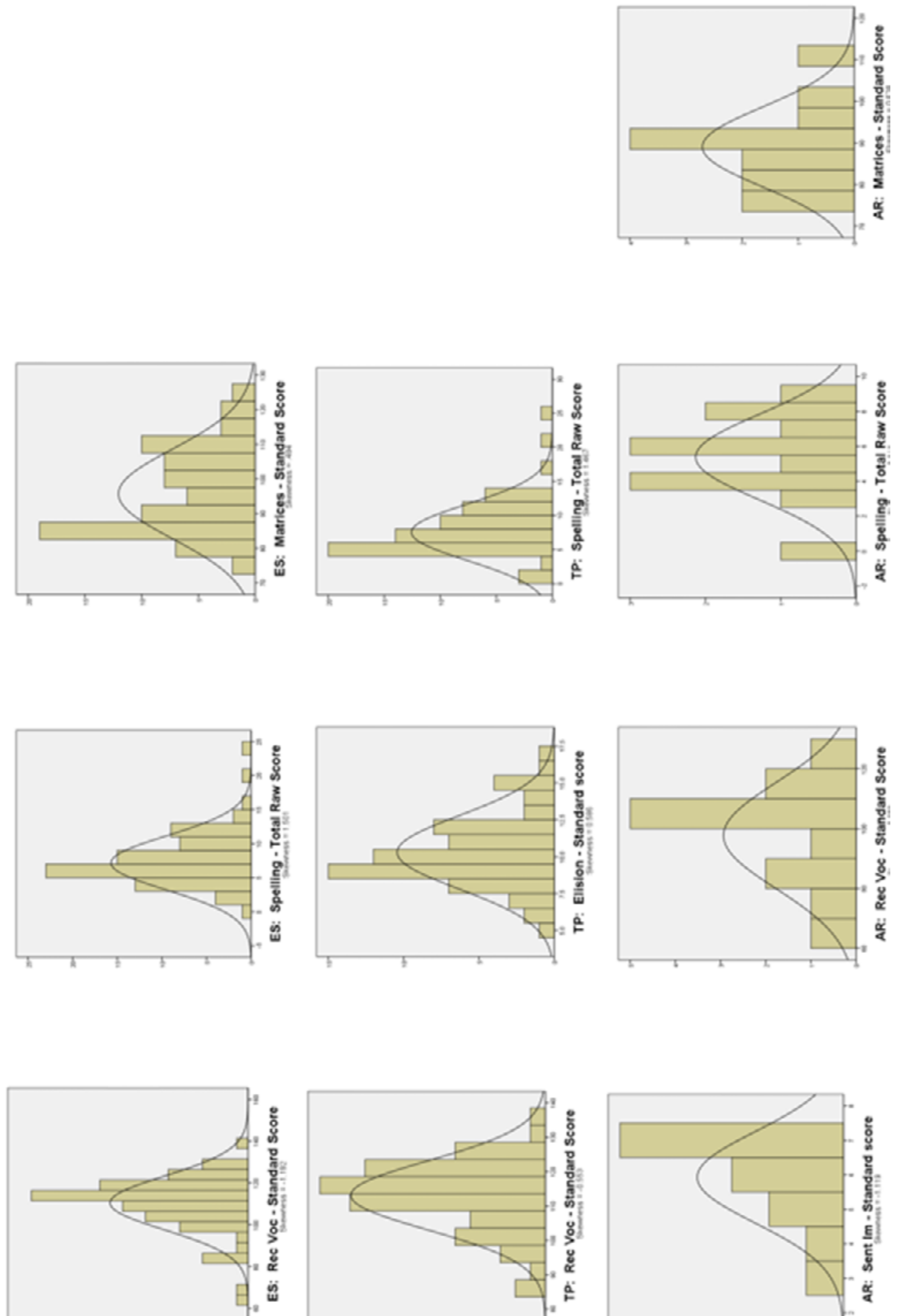

Figure 5. Histogram matrices of language and literacy measures with skewness ≥ ± 0.5 by performance group; ES = entire sample; TP = typically performing group; AR = at-risk group.

would appear to decrease the negative skew and positive skew if these outliers were removed. Based on the above guidelines, the performance on the Vocabulary and Spelling measures were highly skewed for the entire sample. The typically performing group exhibited a moderate skew on the Elision and Vocabulary measures. Furthermore, the typically performing group's performance on the Spelling measure was highly skewed. The performance of the at-risk group exhibited a moderate skew on the Matrices, Vocabulary and Spelling measures. The highly skewed distribution of the Sentence Imitation measure was an artifact of the performance subgroupings and was expected. Grouping by performance can distort the distribution and skewness by focusing on a smaller segment of the wider distribution and concentrating similar outcomes together also skews the results uniquely from when it is included in the entire sample. In sum, the performance of the current sample was not symmetrically distributed on several tests. It is important to note the sample size of the subgroupings especially the at-risk group that was extremely small compared to the general guidelines. Therefore, these results should be interpreted with caution.

Different patterns of skewness were found between the gender groupings. The male group's performance on the Sentence Imitation, and Elision measures approximated a symmetrical distribution while their performance on the Matrices, Word Identification, and Word Attack measures were moderately skewed and their performance on the Vocabulary and Spelling measures was highly skewed. The female group's performance on the, Matrices, Word Identification, Word Attack, and Spelling measures approximated a symmetrical distribution while their performance on the Sentence Imitation and Elision measure was moderately skewed and their performance on the Vocabulary measure was highly skewed.

In sum, the descriptive statistics indicated that the children's performance was relatively normally distributed with mean performance levels within expected ranges for the individual measures in the language and literacy battery including the performance of the at-risk group on all of these measures except for the Sentence Imitation and Elision tasks. However, the at-risk group's performance should be interpreted with caution due to the small size of this subgrouping that was well below the standards for what is expected for running the statistical analyses.

Group Comparisons Across Location, Performance Group, and Gender

Several factors within the study sample could be potential sources of variation that could impact the analyses of the sample performance as a whole including variation due to location (school), performance level, and gender. Thus, a 3 x 2 x 2 ANOVA with school location (Edith Bowen Lab School - EBLS, Wilson, and Adams), performance group (typically performing and at-risk), and gender (male and female) as between-subjects' factors was conducted to investigate the effects of these three independent variables on the individual language/literacy battery measures (see Tables 10 and 11). Homogeneity of variance was confirmed for all of the measures in the language/literacy battery. The general η^2 effect size guidelines to categorize effect sizes are as follows: .01 as small effects, .06 as medium effects, and .14 as large effects (Cohen, 1988; Huck, 2012; Miles & Shevlin, 2001).

Table 10

3 x 2 x 2 Two-Way ANOVA by Language/Literacy Battery Measure: Main Effects

	Levene's test		Location main effect			Performance main effect			Gender main effect		
Measure	*F*	*p*	*F*	*p*	η^2	*F*	*P*	η^2	*F*	*p*	η^2
Sentence imitation	1.464	.172	0.245	.783	.007	47.767	< .001*	.416	0.009	.925	.000
Elision	1.444	.181	2.166	.123	.061	14.177	< .001*	.175	0.467	.497	.007
Vocabulary	1.736	.091	3.284	.044*	.089	19.615	< .001*	.226	0.123	.727	.002
Word identification	1.016	.440	1.402	.253	.040	2.166	.146	.031	0.082	.775	.001
Word attack	0.893	.544	3.249	.045*	.088	2.953	.090	.042	0.482	.490	.007
Spelling [a]	1.450	.178	2.593	.082	.072	4.606	.035*	.064	0.948	.334	.014
Matrices	1.872	.065	1.200	.307	.035	4.619	.035*	.064	0.025	.874	< .001

Notes. Location *df* = (2,67).
Performance *df* = (1,67).
Gender *df* (1,67)

* Statistically significant at $p < .05$.
[a] Used raw score.

Table 11

3 x 2 x 2 Two-Way ANOVA by Language/Literacy Battery Measures: Interaction Effects

	Location x performance			Location x gender			Performance x gender			Location x performance x gender		
Measures	*F*	*P*	η^2	*F*	*p*	η^2	*F*	*P*	η^2	*F*	*p*	η^2
Sentence imitation	0.364	.697	.011	1.323	.273	.038	0.028	.868	< .001	0.365	.548	.005
Elision	2.537	.087	.070	0.492	.613	.014	2.947	.091	.042	0.074	.787	.001
Vocabulary	2.293	.109	.064	2.655	.078	.073	0.413	.523	.006	5.685	.020*	.078
Word identification	0.748	.477	.022	0.276	.760	.008	0.861	.357	.013	0.004	.949	< .001
Word attack	0.828	.442	.024	0.344	.710	.010	2.235	.140	.032	0.012	.913	< .001
Spelling [a]	0.299	.742	.009	0.520	.597	.015	1.254	.267	.018	0.867	.355	.013
Matrices	0.251	.779	.007	0.026	.975	.001	1.158	.286	.017	0.223	.638	.003

*Statistically significant at $p < .05$.

[a] Used raw score.

By Location

A moderate, statistically significant main effect of location was found for the Vocabulary (p = .044; η^2 = .089; medium effect size) and Word Attack (p = .045; η^2 = .088; medium effect size) measures (see Table 10). Therefore, follow-up pairwise comparisons with Bonferroni adjustment were calculated to further explore these main effects. A pairwise comparison for Vocabulary performance based on-location was only statistically significant between Wilson and EBLS (p = .051, d = .564; medium effect size) with the children at EBLS demonstrating stronger vocabulary skill than the children at Wilson. Similarly, a pairwise comparison based on location for the Word Attack performance was only statistically significant between Adams and EBLS (p = .007, d = .863; large effect size) with the children at Adams demonstrating stronger decoding skill than the children at EBLS.

By Performance Group

Main effects for performance group were found for all of the dependent variables with the exception of the Word Identification and Word Attack measures (see Table 10). The corresponding effect sizes of the performance group effects were small to large in size—in the range from .03 to .42—based on the general effect size guidelines for η^2. Follow-up pairwise comparisons were not necessary since there were only two performance groups. As expected, the typically performing group demonstrated stronger mean performance than the at-risk group. These findings should also be interpreted with caution because of the limited size of the subgroupings especially for the at-risk group that was below the expected standards for running the statistical analyses.

By Gender

In addition to the performance groupings above, it was also important to explore any possible difference between the gender groups because females tend to develop language skill earlier than males. The descriptive statistics for the gender groups are listed in Table 10. No main effects for gender were found for the dependent variables in the language and literacy battery. Only one significant interaction was found (see Table 10); on the Vocabulary measure, the three-way interaction among Location, Performance Group, and Gender was statistically significant with a medium effect size (p = .020; η^2 = .078). Since no significant difference in gender group performance was found, there was no need to analyze the performance of the current sample separately based on gender groups.

In sum, these results indicated that the children enrolled at EBLS demonstrated stronger vocabulary skills than those at Wilson and the children enrolled at Adams demonstrated stronger decoding skill than those at EBLS. Additionally, these results indicate that the children in the typically performing group outperformed their peers who were at-risk for literacy difficulty on language measures as expected, but surprisingly not on the measures that directly measured word-level reading.

Construct Representation

Internal Structure

An investigation of the relationship(s) between the performance on dissimilar and similar test items can provide evidence on the validity claim that a measure assesses the construct that it purports to assess based on the internal structure of the measure. The interpretations of the size of the correlations in the following sections were based on the general standards of Cohen (1988, 1992) who suggested that correlations less than .29 corresponded with weak relationship among the variables, correlations between .30 to .49 corresponded with moderate relationships among the variables, and correlations .50 or higher corresponded with strong relationships among variables.

Heterogeneity and homogeneity evidence. Performance variability can first be investigated in the relevant descriptive statistics (range and standard deviation) for the DMMA (see the subsequent section for more detail). Certain patterns in the descriptive statistics support the validity of DMMA interpretations including a wide range of performance that would indicate sensitivity to multiple levels of performance and a symmetrical distribution of performance outcomes with few outliers, which parallel a normal distribution. Descriptive statistics also provide an important foundation for inferences regarding the performance of the general population and future performance of the children in this sample based on their current performance.

The variation and correlations between the performance on the individual stimulus types on the DMMA-RD can be explored for both homogeneity and heterogeneity. Homogeneity can be explored by comparing performance on similar stimulus types that were included on both DMMA subtasks. Theoretically, similar stimulus items were expected to demonstrate strong relationships with larger, significant correlations that indicate homogeneity as compared to dissimilar stimulus items. These dissimilar items were expected to demonstrate weak relationships with smaller, insignificant correlations that indicate the heterogeneity.

Heterogeneity was explored by comparing performance on the dissimilar stimulus types of the accurate word pairs with the foil/inaccurate word pairs (i.e., *arm* and *army*) that needed to be included on the DMMA-RD task to provide some contrast for discrimination of the accurate word pairs (i.e., *egg* and *eggs*). Inaccurate/foil word pairs mimicked the addition of a morphemic ending without the added meaning versus the accurate word pairs that include the addition of a morpheme have fundamentally different relationships to meaning. The meaning of an inaccurate/foil word pair is an encapsulated whole within a word (i.e., one bound morpheme). The meaning of an accurate word pair involves two separate, meaningful units (i.e., morphemes) that both contribute to the overall meaning of a word (i.e., morphologically complex). Therefore, low correlations were expected between these two stimulus types because of their varied connection to meaning. The comparison of performance on the accurate versus the inaccurate/foil word pairs was theoretically expected to demonstrate heterogeneity in small and insignificant correlations. A weak relationship was expected between these stimulus types due to an underlying dissimilarity. The DMMA-EP task could only be used to explore homogeneity since all the stimuli represented accurate word pairs. The comparison of performance among the accurate word pairs was expected to demonstrate homogeneity because of the underlying similarity. Moreover, large and significant correlations among these items would indicate a strong relationship between similar stimulus types. The heterogeneity and homogeneity of the performance for the DMMA-RD and DMMA-EP was examined further using the correlations between stimulus types.

Correlations. Based on the construct, a heterogeneous relationship was expected in the performance between the stimulus types that targeted the accurate word pairs and the foil/inaccurate word pairs (i.e., dissimilar word pairs). The correlations between each stimulus type by DMMA task are listed in Table 12. Medium correlations (.22 - .32) were found between the accurate and foil/inaccurate stimulus types on the DMMA-RD. Likewise, small to medium correlations (.18 - .33) also were found between the accurate stimuli on the DMMA-EP and the foil/inaccurate stimuli on the DMMA-RD. These divergent findings indicated heterogeneity in the DMMA-RD performance only on these dissimilar items. Conversely, large correlations were found between both similar stimuli (i.e., accurate-to-accurate and foil-to-foil) that indicated homogeneity in the performance on both the DMMA-RD (.76 - .84) and DMMA-EP (.87). These convergent findings corresponded to homogeneity in the performance on both of the similar stimulus types. However, only very small to medium correlations (-.01 - .32) were found between the accurate stimulus items across the two tasks (DMMA-RD vs. DMMA-EP), which may reflect the difference between the tasks (task variance) themselves rather than morphological awareness (construct variance).

On the whole, the correlations among stimulus types on the DMMA-RD and DMMA-EP exhibited the expected patterns of heterogeneous and homogeneous performance. However, the correlations between performance on similar stimulus types within the two tasks (DMMA-EP and DMMA-RD) were substantially larger than the similar stimulus types between the two tasks that may be connected to the difference between the tasks.

Table 12

Correlation for Stimuli Type by DMMA Task and Morpheme Category

	DMMA-RD				DMMA-EP		MC
Stimulus type and morpheme category	Accurate inflections	Accurate derivations	Foil Inflections	Foil Derivations	Accurate Inflections	Accurate Derivations	Accurate[a]
DMMA-RD							
Accurate inflections	1						
Accurate derivations	.757**	1					
Foil inflections	.316**	.222	1				
Foil derivations	.324**	.229*	.844**	1			
DMMA-EP							
Accurate inflections	.319**	.135	.289*	.331*	1		
Accurate derivations	.158	-.010	.179	.239*	.870**	1	
MC							
Accurate	.408**	.276*	.264*	.305**	.790**	.687**	1

[a] Mix of inflection and derivation depending on individual ceiling item.

* Significant to the $p < .05$ level.

** Significant to the $p < .01$ level.

Reliability

The reliability of performance on a measure provides an additional source of evidence related to the validity of interpreting performance as an indicator of an underlying construct. Internal consistency, test-retest reliability, and interrater agreement can support the validity of interpretations of DMMA performance by providing evidence that performance was consistent across individual stimulus items, multiple administrations, and different raters, respectively. Consistency across test items demonstrates a measure's reliability in the measurement of targeted elements of the main construct (e.g., judgment or expression of the individual morphemes as an indicator of morphological awareness) whether within a single administration or across administrations. The consistency of individual rater's judgments of performance was also essential to the validity of interpretations of DMMA performance because different raters may demonstrate some variation in how they record/score individual performance (e.g., one rater may award full points for a response when another awarded only partial points). Consistent, reliable results were critical to validity of interpretations of DMMA performance because they rule out common alternative interpretations of results. It is important that a measure provides a relatively equivalent, consistent picture of performance between individual stimulus items, across multiple administrations, and across performance judgments.

Internal consistency. The internal consistency of a measure refers to whether the interpretations of performance on a measure yield similar results across the individual items. There are several ways to explore the internal consistency of test items: split-half and Cronbach's alpha.

Split-half. Split-half reliability focused on the internal consistency of performance interpretations from within an administration of a measure. Split-half reliability is calculated by dividing a measure into two equivalent parts (e.g., first half compared to second half or odd-numbered items compared to the even-numbered) and computing the correlation between an individual's scores on the two halves. The split-half reliabilities for each form of the DMMA using these divisions of equal halves and odd v. evens are included in Table 13. The split-half reliabilities for the DMMA ranged from 0.81 to 0.96. These findings indicated that the individual forms of the two DMMA tasks demonstrated strong internal consistency yielding similar, consistent performance when various combinations of the individual stimuli were examined.

Cronbach's alpha. The coefficient alpha or Cronbach's alpha (CA) expanded the calculations of split-half reliability to include an average of all the possible split-half combinations. As a measure of reliability, a high CA would indicate that the individual items of the DMMA result in reliable performance. It is important that additional stimuli help to explain the performance interpretations rather than repeating the information

Table 13

Split-Half Reliability for the DMMA

	Form A		Form B	
Measure	Equal halves[a]	Odds vs. evens[b]	Equal halves[a]	Odds vs. evens[b]
DMMA—RD	.809	.865	.895	.925
DMMA—EP	.938	.956	.935	.932
DMMA—Total		.922		.941

[a] Equal halves method—A (1-15 vs. 16-30) and B (31-45 vs. 46-60).

already provided by the prior stimulus items (i.e., limited redundancy within the test content). For example, high CA values may indicate a large amount of redundancy rather than true performance reliability. Similarity between the mean and standard deviations can support the comparability of the interpretations that the individual stimulus items measure morphological awareness. According to Nunnally (1978), the coefficient alpha can be quantified on a scale from 0.00 for no consistency to 1.00 for perfect consistency. Interpretations of reliability generally begin with at least .60 to .70 being acceptable for exploratory studies; however, reliability results between .80 to .90 are generally acceptable and applicable to all research whether exploratory or applied research (Nunnally, 1978). The CAs for each DMMA forms were within the acceptable category for all research ranging from .84 to .96 (see Table 14). These results confirmed the individual split-half reliabilities described in the prior section and provided additional evidence of the strong internal consistency of the DMMA measure.

Test-Retest reliability. Test-retest reliability focused on the consistency of performance across multiple administrations of a measure. Although similar to the above reliability analyses, test-retest reliability can be impacted by time and experience, which could directly contribute to a variation in performance. The separate administrations of

Table 14

Coefficient Alpha (CA) for the DMMA

Task	Form A	Form B	Total
DMMA-RD	.844	.887	.924
DMMA-EP	.936	.908	.955
DMMA-Total	.909	.913	.950

the forms of the DMMA tasks allowed for the calculation of test-retest reliability. The Form-A (i.e., stimulus items 1 to 30) and Form B (i.e., stimulus items 31 to 60) for both the DMMA-RD and DMMA-EP were administered in separate sessions within the scheduling parameters described in the Methods section. The test-retest reliability was calculated for the two DMMA tasks (i.e., DMMA-RD or DMMA-EP) and the DMMA-Total using an equal halves division (i.e., directly comparing Form A to B; see Table 15). Few general guidelines exist for the acceptable levels of test-retest reliability especially since outcomes can be directly impacted by time elapsed between administrations (Crocker & Algina, 2008). Since test-retest reliability is based on a correlation between multiple administrations of the same measure, the general correlation guidelines can be applied to these outcomes. The results showed large test-retest correlations for each subtask and the DMMA-Total.

Interrater agreement on responses. Interrater agreement on responses provides evidence of reliability based on the level of consensus on rating examinees' performance by multiple examiners. It can be quantified by both percent agreement and Cohen's kappa (Cohen, 1960). Interrater agreement was calculated for the DMMA-EP only since the responses were given orally and scoring could have been impacted by a subjective judgment of accuracy.

Percent agreement. A blinded reviewer (a graduate student in a communication sciences and disorders program who was not present when the items were originally scored by the primary scorer) rescored 30% of each location with the original scoring guidelines (i.e., 24 protocols of each form—A and B) of the DMMA-EP protocols from

Table 15

Test-Retest Reliability for the DMMA

Method	*DMMA-RD*	*DMMA-EP*	*DMMA-Total*
Equal Halves method—A v. B	.879	.870	.810

audio recordings that were randomly selected and balanced across form, location, and teacher/classroom to ensure a representative sample of the entire data set. The interrater agreement was calculated by comparing the scoring results of the primary author and the blinded scorer. The interrater agreement was 76% on the entire crosschecked sample with 74% agreement on the Form A and 78% agreement on Form B. The disagreements among the scorers appear to be due mainly to the discrepancy between the primary author having the benefit of scoring the children's "live" responses as they were given as compared to the blinded scorer having to rely on audio recording(s) of imperfect quality.

Cohen's kappa. Interrater agreement can also be evaluated by calculating Cohen's (1960) kappa. Cohen's kappa enhances the estimates of interrater agreement by adjusting for the chance level of agreement. This results in a more rigorous estimate of interrater agreement reliability. Cohen's kappa for the DMMA-Total was .749. The interrater agreement findings indicated an acceptable level of agreement between raters and suggested that the results were only minimally affected by differences among raters.

Comparisons of Morphological Awareness Performance

Since morphological awareness is an emergent skill in young children, floor effects have been found when current assessments have been used to assess this skill in this population. Floor and/or ceiling effects present two difficulties to the valid interpretations of DMMA performance as truly measuring morphological awareness. If morphological awareness is a continuously distributed skill, floor and/or ceiling effects would indicate that the DMMA performance did not correspond to this understanding of morphological awareness. Furthermore, floor and/or ceiling effects can prevent the detection of performance differences among children who achieve the same morphological awareness score. Comparison of the performance features, like distribution of performance, frequency of prompt use, correlation, and classification prediction, of the DMMA to an established measure of the same construct like the Morphological Completion measure, was another evidence source for the validity of interpreting DMMA performance as reflecting morphological awareness in young children with respect to the response processes. These performance features are discussed in the succeeding sections.

Descriptive Statistics for the Morphological Awareness Measures

The descriptive statistics (mean, standard deviation, range, skewness, and kurtosis) for the morphological awareness measures are presented in Table 16 for the entire group, Table 17 by performance group and Table 18 by gender. Additionally, the histograms of the score distributions with a skewness statistic of ±0.5 are shown in Figure 6.

Mean performance. The entire sample and typically performing group both outperformed the at-risk group on the DMMA-Total as expected. The entire sample

Table 16

Morphological Awareness Measures: Descriptive Statistics—Total Sample (N = 78)

Measure	Mean	*SD*	Min	Max	Skewness	Kurtosis
MC	22.17	6.01	0	30	-1.599	3.251
DMMA-Total	350.81	62.42	106	451	-1.558	3.100
DMMA-RD	169.65	37.63	60	229	-0.541	-0.011
DMMA-EP	181.49	39.86	16	222	-2.617	8.034

Note. All measures used raw scores; MC = Morphological Completion; Skewness SE = 0.272; Kurtosis SE = 0.538.

exhibited a higher mean performance on the DMMA-EP than the DMMA-RD. The typically performing group paralleled this performance trend. Conversely, the at-risk group exhibited a higher mean performance on the DMMA-RD than the DMMA-EP but these findings should be interpreted with caution because of the small size of this subgrouping.

A direct comparison of the mean performance between the Morphological Completion and DMMA cannot be completed due to the large difference in points possible on the measures. However, examining the percentage of points achieved by the children on these measures can provide the basis for a comparison of performance. The entire sample appeared to exhibit comparable skill as suggested by the 74% of points awarded on both measures. However, the entire sample achieved 76% of the points possible on the DMMA-EP that represented a more analogous task to the Morphological Completion measure while they achieved 71% of the points possible on the DMMA-RD. It is important to note that the typically performing group outperformed the at-risk group

Table 17

Language/Literacy Battery: Descriptive Characteristics—By Performance Group with t-*Test Results*

	Typically performing (*n* = 65)					At-risk (*n* = 13)					Comparisons		
Measure	*M*	*SD*	Range	Skewness	Kurtosis	*M*	*SD*	Range	Skewness	Kurtosis	*t*	*p*	*d*
MC	24	4	10-30	-0.78	0.83	15	9	0-29	-0.32	-0.72	5.72	< .001	1.29
DMMA													
Total	366	42	244-451	-0.75	0.55	276	91	106-407	-0.33	-0.69	5.60	< .001	1.27
RD	173	35	60-229	-0.62	0.40	152	47	80-229	0.10	-0.69	1.54	.144	0.51
EP	193	20	106-222	-1.41	3.86	126	63	16-190	-0.98	-0.55	3.79	.002	1.43

Note. All measures use raw scores; Skewness *SE*—Typically performing = 0.297, At-Risk = 0.616; Kurtosis - *SE*—Typically performing = 0.586, At-Risk = 1.191; MC = Morphological Completion - 38 total points possible; DMMA-Total = 480 total points possible; DMMA-RD and DMMA-EP = 240 total points possible for each.

Table 18

Language/Literacy Battery: Descriptive Characteristics—By Gender with t-*Test Results*

	Male (n = 40)					Female (n = 38)					Comparisons		
Measure	*M*	*SD*	Range	Skewness	Kurtosis	*M*	*SD*	Range	Skewness	Kurtosis	*t*	*p*	*d*
MC	23	5	0-30	-2.03	7.24	21	7	2-30	-1.31	1.51	1.26	.213	0.29
DMMA													
Total	360	53	170-451	-1.30	3.07	341	71	106-426	-1.53	2.40	1.40	.166	0.32
RD	178	35	99-229	-0.43	-0.68	160	39	60-224	-0.57	0.20	2.15	.035	0.49
EP	183	36	16-222	-2.93	11.72	180	44	17-219	-2.43	6.49	0.28	.782	0.06

Note. All measures use raw scores; Skewness—SE—Male = 0.374, Female = 0.383; Kurtosis—SE—Male = 0.733, Female = 0.750; MC = Morphological Completion - 38 total points possible; DMMA-Total = 480 total points possible; DMMA-RD and DMMA-EP = 240 total points possible for each; RS = Raw Score.

as expected on all of these morphological awareness measures based on the mean performance.

Standard deviation and variability. The entire sample and at-risk group exhibited more variability on the DMMA-EP versus the DMMA-RD as demonstrated in the standard deviation(s) and range of score(s) while the typically performing group demonstrated more variability on the DMMA-RD versus the DMMA-EP. The difference between the standard deviations was the smallest for the entire sample followed by the typically performing group. The at-risk group exhibited the largest difference in standard deviations between the DMMA subtasks, which supports the highest level of variability overall compared to the typically performing group and the entire sample. However, due to the small sample size of the at-risk group, these results should be interpreted with caution.

The Morphological Completion performance of the entire sample covered the range of the points possible on the measure. The typically performing group's performance reflected less variability as exhibited in a smaller range and standard deviation than the at-risk group. The at-risk group's performance, conversely, demonstrated more variability than the typically performing group exhibiting the same range as the entire sample and with a higher standard deviation. These patterns of performance for both subgroupings were also seen in their performance on the DMMA and subtasks.

Skewness. The top row of Figure 6 shows the morphological awareness performance distributions for the entire sample. The performance on all the entire sample

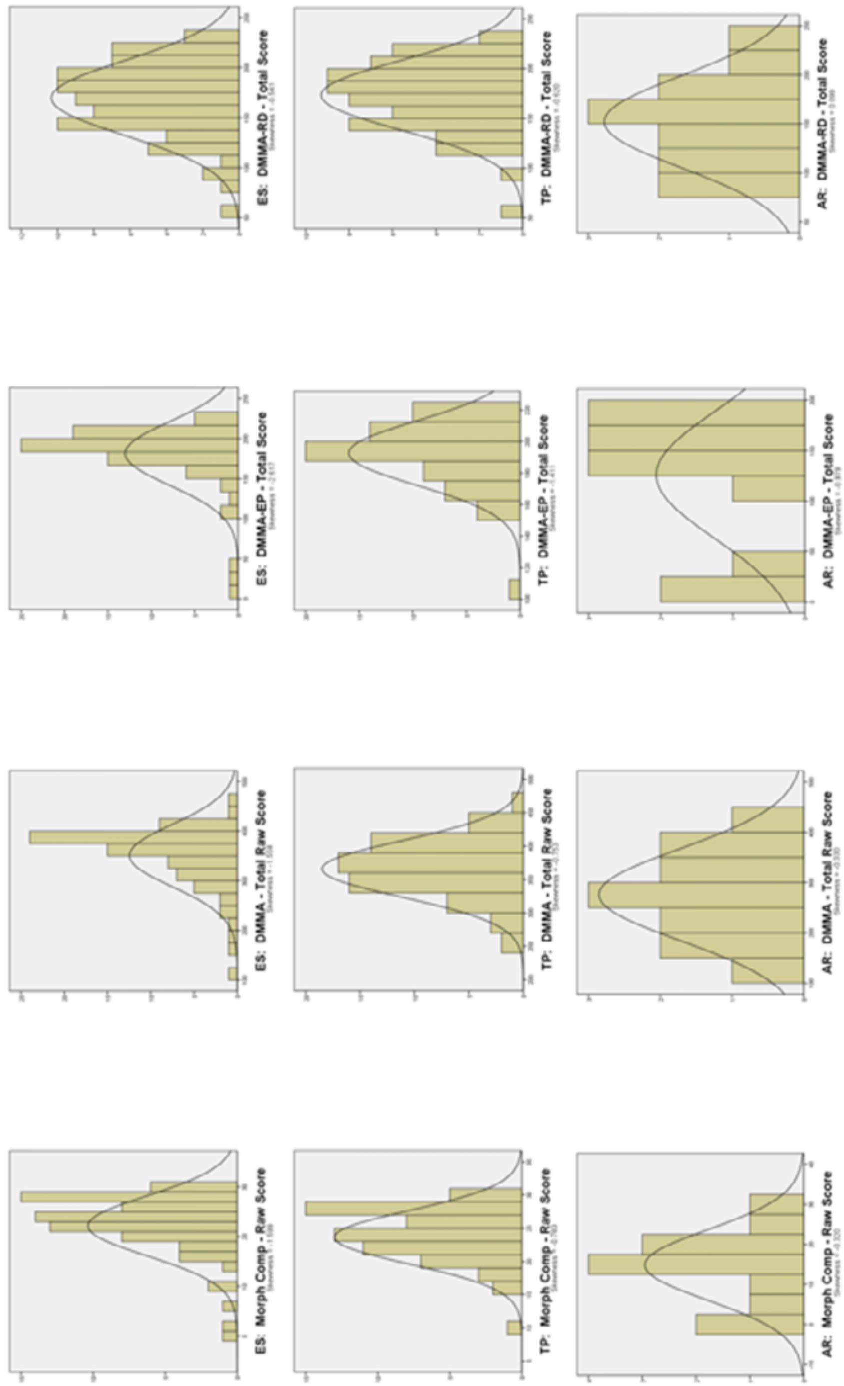

Figure 6. Histogram matrices of the morphological awareness measures. The rows correspond to the performance groups: top = entire sample, middle = the typically performing group, and bottom = at-risk group. The columns correspond to the measures: first column on left = Morphological Completion measure, second column = DMMA-Total, middle column = DMMA-EP, and the last column = DMMA-RD. ES = Entire Sample; TP = Typically performing; AR = At-Risk; Morph Comp = Morphological Completion

exhibited highly, negatively skewed performance on all the morphological awareness measures, except for a moderate skew on the DMMA-RD. A negative skew corresponded to the clustering of performance towards higher values with an extended tail reaching towards the lower values. Performance on the DMMA-EP reflected the most negative skew (Skewness = -2.62) followed by the Morphological Completion measure that was also highly, negatively skewed (Skewness = -1.60). The DMMA-RD was only moderately skewed for the entire sample (Skewness = -0.54). Some outliers were noted in the performance of the entire sample. The Morphological Completion and DMMA-Total measures appeared to exhibit similar skewness (i.e., within -0.04) for the entire sample. In conclusion, the DMMA performance distributions of the entire sample demonstrated a trend toward the presence of a possible ceiling effect for the DMMA-Total, but no perfect scores were achieved (i.e., highest score achieved = 451; total possible score = 480). The inclusion of the DMMA-RD in the DMMA-Total appeared to have provided some balance to the DMMA-Total (i.e., less skew and more symmetry) compared to the DMMA-EP alone.

The results for the typically performing group parallel those of the entire sample since this group contributed the majority of the performance to the entire sample (i.e., 65 children out of 78). The typically performing group exhibited less skew than the entire sample demonstrating a high, negative skew only on the DMMA-EP and moderate skew on the other measures of morphological awareness (refer to the middle row of Figure 6). The skewness of the typically performing group's performance on the Morphological Completion and DMMA-Total measures was also comparable (i.e., within -0.03) like the entire sample. The spread of scores noted on the Morphological Completion, DMMA-Total, and DMMA-RD for the typically performing group demonstrated a moderate negative skew. A few possible outliers were apparent on all the morphological awareness measures except the DMMA-Total. In brief, the results of the typically performing group mirror those of the entire sample with some negative skew noted in the DMMA and subtasks, but less than that of the entire sample.

The distributions of the at-risk performance group presented a different pattern of results. Although these results should be interpreted with caution because of the small performance group sample size of 13 (compared to the generally acceptable minimum group size of 30) and the limited proportion of the entire sample that this group represents. The distributions of the at-risk group exhibited a wider spread, which likely related to their less developed, emergent morphological skill compared to their more developed typically performing peers. The DMMA-RD distribution was roughly symmetrical as demonstrated by the minimal positive skewness statistic (Skewness = 0.10). Like the typically performing group and the entire sample, the at-risk group's performance on the Morphological Completion and DMMA-Total measures exhibited similar skewness (i.e., within -0.01). Interestingly, the DMMA-EP demonstrated a bimodal distribution with high variability (i.e., larger SD compared to DMMA-RD - the most comparable task). The skewness statistic of -.98 also indicates that this distribution lacks symmetry. A trend toward potential floor and ceiling effects appeared to be present for this measure. However, the distribution for the DMMA-Total for the at-risk group appeared to be the most symmetrical of all the comparable distributions among the performance groups and entire sample (Skewness = -.33).

In conclusion, the Morphological Completion and DMMA-Total resulted in similar distributions across the performance of the current study sample. On the whole, the DMMA-Total appeared to better approximate a symmetrical distribution of performance for the current sample compared to the individual subtasks, especially for the at-risk group. Moreover, the morphological awareness performance distributions of the entire sample demonstrated a trend toward the presence of a possible ceiling effect for the DMMA-EP and Morphological Completion measures; however, only a few perfect scores were achieved on the Morphological Completion measure. Finally, the typically performing group outperformed the at-risk group as expected, but the at-risk group's performance distribution was the most symmetrical. Nonetheless, the results for the at-risk group should be interpreted with caution because of the small size of this subgrouping.

Group Comparisons

The children's performance on the morphological awareness measures were also compared using a 3 x 2 x 2 ANOVA with location (EBLS, Wilson, and Adams), performance group (typically performing and at-risk), and gender (male and female) as between-subjects' factors. This analysis was conducted to investigate the effects of these three independent variables on each morphological awareness measure. Homogeneity of variance was confirmed for the Morphological Completion and DMMA-RD measures, but not for the DMMA-Total and DMMA-EP (i.e., dependent variables).

By location. A main effect of location was found for all of the morphological awareness measures (MC - $p < .001$, $\eta^2 = .244$; DMMA-Total—$p = .033$, $\eta^2 = .097$; DMMA-EP—$p < .001$, $\eta^2 = .218$) except for the DMMA-RD indicating a significant difference in the children's performance on these measures based on location (see Table 19); therefore, follow-up pairwise comparisons were also calculated to further explore these main effects. A pairwise comparison with a Bonferroni adjustment for the Morphological Completion performance based on location was only significant between Wilson and EBLS ($p < .001$, $d = .674$) with the children at EBLS demonstrating stronger morphological completion skill as compared to the children at Wilson. Similarly, a pairwise comparison with a Bonferroni adjustment for the DMMA-Total performance based on location was only significant between EBLS and Wilson ($p = .018$, $d = .587$) with the children at EBLS also demonstrating stronger skill on the DMMA-Total as compared to the children at Wilson. Finally, a pairwise comparison with a Bonferroni adjustment for the DMMA-EP performance based on location was significant between EBLS and Wilson ($p < .001$, $d = .717$) and Adams and Wilson ($p = .041$, $d = .443$) with the children at EBLS and Adams demonstrating stronger skill on the DMMA-EP as compared to the children at Wilson.

By performance group. Main effects for the performance groups were also found for the morphological awareness measures paralleling the previously reported results reported above for the language and literacy battery (see Table 19) with effect sizes ranging from .06 to .39 (i.e., medium to large effects) based on the guidelines of Cohen (1988) and Miles and Shevlin (2001). The subsequent interaction effects revealed significant interactions between location and performance, and among location,

Table 19

3 x 2 x 2 Two-Way ANOVA by Morphological Awareness Measure: Main Effects

	Levene's Test		Location main effect			Performance main effect			Gender main effect		
Measure	*F*	*p*	*F*	*p*	η^2	*F*	*p*	η^2	*F*	*p*	η^2
MC	1.445	.180	10.809	< .001*	.244	24.697	< .001*	.269	0.821	.368	.012
DMMA-Total	2.342	.020*	3.601	.033*	.097	29.713	< .001*	.307	0.559	.457	.008
DMMA-RD	1.007	.447	1.522	.226	.043	4.478	.038*	.063	5.584	.021*	.077
DMMA-EP	7.211	< .001*	9.338	< .001*	.218	42.402	< .001*	.388	2.690	.106	.039

Note. All measures used raw scores; MC = Morphological Completion.

Location *df* = (2,67).

Performance *df* = (1,67).

Gender *df* (1,67).

*Statistically significant at $p < .05$.

performance group, and gender for all the morphological awareness variables, except the DMMA-RD (see Table 20). In conclusion, the morphological awareness measures appeared to be sensitive to performance differences in young children who are developing this emergent skill.

By gender. A significant main effect for gender was found for the DMMA-RD with a medium effect size. The male group appeared to demonstrate stronger performance on the DMMA-RD as shown in the higher mean score and lower standard deviation (i.e., less variability).

In sum, these results indicated that the children enrolled at EBLS and Adams demonstrated stronger morphological awareness skill than those at Wilson. Additionally, these results indicated that the children in the typically performing group outperformed their peers who were at-risk for literacy difficulty on language measures as expected.

Frequency of Prompt Use

The frequency of prompt use provided another way to examine the response processes of the children on the DMMA. The frequency of prompt use reflected the level of independence of accurate responses. For example, an accurate response to Prompt #1 indicated the most independent accuracy because no additional prompting was provided to elicit this response; conversely, an accurate response to Prompt #4 indicated the least independent accuracy because maximal prompting was provided to elicit this response. Ideally, independent, accurate performance can be interpreted as mastery of this skill. It was expected that accurate responses would require a range of prompts across the children's performance depending on their morphological awareness skill development

Table 20

3 x 2 x 2 Two-Way ANOVA by Morphological Awareness Measure: Interaction Effects

	Location x performance			Location x gender			Performance x gender			Location x performance x gender		
Measure	*F*	*p*	η^2	*F*	*p*	η^2	*F*	*P*	η^2	*F*	*p*	η^2
MC	10.793	< .001*	.244	0.600	.552	.018	2.780	.100	.040	4.168	.045*	.059
DMMA-Total	3.283	.044*	.089	1.359	.264	.039	0.010	.920	< .001	7.104	.010*	.096
DMMA-RD	1.390	.256	.040	0.318	.729	.009	2.253	.138	.033	2.489	.119	.036
DMMA-EP	6.422	.003*	.161	2.346	.104	.065	2.857	.096	.041	7.086	.010*	.096

Note. All measures used raw scores. MC = Morphological Completion.

*Statistically significant at $p < .05$.

because of the emergent nature of this skill. Prompt use can provide supportive evidence for the valid interpretation of DMMA performance across children with varying levels of morphological awareness. Several patterns of performance were demonstrated in the frequency of prompt use among the prompts, performance group and DMMA task (see Table 21).

Prompt #1, the initial stimulus prompt without any additional prompting, successfully elicited accurate responses with the most frequency for both performance groups. As expected, Prompt #1 was more effective for the typically performing group than the at-risk group. Prompt #2, the stimulus prompt with emphasis on the targeted word pair, successfully elicited accurate responses about 15% of the time for both performance groups and across the tasks. Both Prompt #3, the stimulus prompt with visuals of the targeted word pair, and Prompt #4, the stimulus prompt with both emphases on and visuals of the targeted word pair, successfully elicited accurate responses less than 10% of the time across both the tasks and performance groups.

Table 21

DMMA: Frequency of Prompt Use—By Performance Group

	Typically performing group (*n* = 65)		At-risk group (*n* = 13)	
Prompt	DMMA-RD (%)	DMMA-EP (%)	DMMA-RD (%)	DMMA-EP (%)
Prompt #1	56	65	48	39
Prompt #2	14	18	13	13
Prompt #3	8	2	9	5
Prompt #4	4	2	4	4
Inaccurate	17	12	26	39

Inaccurate responses remained for the at-risk group in about a quarter of the stimulus items following the administrations of all the prompts—this is approximately twice the percentage of inaccurate responses in the typically performing group. These results highlighted the differences in-prompt effectiveness based on group and task performance and have implications for the use of dynamic assessment with young children who demonstrate performance differences. Prompt #1 resulted in success most frequently although slightly more often for the typically performing group as compared to the at-risk group. The frequency of accurate responses decreased through the successive prompting levels from Prompt #2, #3, and #4 with very similar patterns across both performance groups. Differences between performance groups were a result of the increased accuracy with Prompt #1 exhibited by the typically performing and number of items that were inaccurate after all prompts were exhausted for the at-risk group.

Correlations

Correlations were calculated to investigate the relationships

between the measures of morphological awareness in young children (see Table 22). The correlations between the Morphological Completion measure and the DMMA ranged from medium to large (i.e., range from $r = .42$ to $r = .77$) and were significant at the $p < .01$ level. More specifically, a large correlation was found between the Morphological Completion measure and DMMA-Total ($r = .744$), which suggests that these two measures may be tapping into the same construct. The large correlations of $r = .79$ and $r = .82$ between the DMMA-Total and the subtasks (DMMA-RD and DMMA-EP, respectively) were expected since the subtasks were combined to create the DMMA-

Table 22

Correlations for the Morphological Awareness Measures: All Participants (N = 78)

Measure	1	2	3	4
1. MC	1			
2. DMMA-Total	.744**	1		
3. DMMA-RD	.421**	.787**	1	
4. DMMA-EP	.767**	.819**	.294**	1

Note. All measures used raw scores. MC = Morphological Completion.
** significant at the $p < .01$

Total score. Although the medium correlation ($r = .29$) between the DMMA subtasks was unexpected, this could demonstrate that each of the tasks is measuring two separate aspects of morphological awareness, judgment and expression. The overall performance of the entire sample indicated that performance on the Morphological Completion measure strongly and significantly related to performance on both the DMMA- Total ($r = .74$; $p < .01$) and DMMA-EP ($r = .77$; $p < .05$).

Classification: Sensitivity and specificity. Beyond the correlational similarity of performance scores on the Morphological Completion and DMMA measures, another way to investigate the relationship between these measures is to examine the degree of agreement between each measure's classifications of student performance. Similarity of performance can be examined on how accurately classifications based on one test can be predicted based on their scores on another morphological test. In this analysis, the performance classification, based on the Morphological Completion measure as the established standard, was used to determine how well the DMMA scores approximated this classification. These further analyses of the results were included in the current study to assess the utility of the Morphological Completion measure and DMMA as classification tools to inform early literacy screening in young children. Performance predictions can be classified into four groups: (a) true positives = weak morphological skill accurately identified; (b) false positives = strong morphological skill identified as weak; (c) true negatives = strong morphological skill accurately identified; and (d) false negatives = weak morphological skill identified as strong. Predictive sensitivity refers to the number of children *accurately* identified with weak morphological skill (i.e., true positives) compared to the *total* number identified as weak in morphological skill. Conversely, predictive specificity refers to the number of children *accurately* identified with strong morphological skill (i.e., true negatives) compared to the *total* number identified as strong in morphological skill. A guideline of .70 or greater is generally considered an acceptable level of prediction for these statistics. A useful measure balances its predictive sensitivity with predictive specificity.

Cut-off scores were needed as the basis for classification into performance groups (i.e., typically performing and impaired) to inform diagnosis. On the Morphological morphological awareness measures reflected a negative skew for the entire sample. The Completion measure, the conventional cut-off of a score of 7 or below (i.e., one standard deviation below the mean of 10) was used to identify children with weak morphological skill. Two cut-off scores for the DMMA and subtasks were calculated using the SPSS software (Version 21; IBM Corp., 2012) in order to provide a basis for comparison of the accurate classifications of the children's performance. A cut-off score of 329 or 341 (AUC = .764) provided the most balance among sensitivity (i.e., 85.71% for both) and specificity (i.e., 78.87% and 76.06%, respectively) for the DMMA when compared to the classifications of the Morphological Completion measure. Table 23 compares the classification predictions based on the children's performance on the DMMA-Total and the Morphological Completion measure. For example, the Morphological Completion measure and DMMA-Total with a cut-off score of 341, both classified six children as at-risk while 17 children were classified as typically performing by the Morphological Completion measure but as at-risk by the DMMA-Total at this same cut-off. As detailed in the beginning of this section, these comparisons provided additional evidence for the relationship between the Morphological Completion and DMMA-Total measures beyond the correlational evidence. In sum, with Morphological Completion measure used as a standard, the DMMA-Total showed acceptable levels of classification accuracy for identifying performance differences.

Relationship with Other Language and Literacy Skills

Comparing the performance on the DMMA with the performance on other language and literacy skills can also provide evidence for the valid interpretation of the

Table 23

Classification prediction matrices for the Morphological Completion and the DMMA

	Morphological completion				Morphological completion		
DMMA 341 CO	Weak MA	Strong MA	Total	DMMA 329 CO	Weak MA	Strong MA	Total
Weak MA	6	17	23	Weak MA	6	15	21
Strong MA	1	54	55	Strong MA	1	56	57
Total	7	71	78	Total	7	71	78

CO = Cut-off Score; MA = Morphological awareness.

former through convergence or divergence. A strong relationship between two variables demonstrates convergence while a weak relationship demonstrates divergence. Validity is supported when the obtained correlations correspond with the expected patterns of correlations based on theoretical relationships among the constructs. The skills targeted in the language and literacy battery varied in their expected levels of convergence and divergence with the DMMA. The language measures (i.e., Elision—phonological awareness, Vocabulary, Sentence Imitation - general language ability) and the literacy measures (Word Identification, Word Attack, and Spelling) were expected to demonstrate the most convergence while the Matrices measure (i.e., general cognitive ability) was expected to demonstrate the most divergence with the morphological awareness measures (DMMA and MC). Like the above comparison within the construct of morphological awareness, both correlation and classification comparisons were conducted among the language and literacy skills included in the assessment battery of the current study.

Correlations

The correlations among the language and literacy skills, including the DMMA, ranged from very small to large. For the entire sample, the DMMA-Total moderately correlated (i.e., range from r = .12 to r = .74; p < .01) with all the other language and literacy skills, except for the Matrices - general cognitive ability measure (see Table 24). The DMMA-Total exhibited the strongest correlation with the Morphological Completion measure (r = .74) followed by other language measures (Vocabulary, Sentence Imitation - general language ability, and Elision - phonological awareness) and the literacy measures (i.e., Spelling, Word Attack, and Word Identification) while it

Table 24

Correlations for the Language/Literacy Battery Measures: Total Sample (N = 78)

Measure	1	2	3	4	5	6	7	8	9	10	11
1. Sentence imitation	1										
2. Elision	.597**	1									
3. Vocabulary	.556**	.379**	1								
4. Word identification	.305**	.472**	.194	1							
5. Word attack	.409**	.606**	.335**	.671**	1						
6. Spelling	.346**	.534**	.347**	.516**	.714**	1					
7. Matrices	.328**	.276*	.302**	.216	.311**	.225*	1				
8. Morphological completion	.637**	.438**	.659**	.214	.298**	.245*	.152	1			
9. DMMA—Total	.573**	.470**	.650**	.311**	.336**	.374**	.125	.744**	1		
10. DMMA-RD	.251*	.236*	.298**	.238*	.230*	.273*	-.019	.421**	.787**	1	
11. DMMA-EP	.648**	.492**	.742**	.255*	.298**	.319**	.218	.767**	.819**	.294**	1

Note. Measures 8 to 11 used raw scores.

* Significant at the $p < .05$ level.

** Significant at the $p < .01$.

exhibited the weakest correlation with the measure of general cognitive ability (i.e., Matrices). This pattern of results paralleled the expectations that the DMMA-Total should correlate more strongly with skills that, theoretically, should be highly related than with lesser-related skills.

The Morphological Completion measure demonstrated small to medium correlations with the other language measures (Sentence Imitation, Elision and Vocabulary; ranging from $r = .44$ to $r = .66$; $p < .01$) and with both Word Attack significantly ($r = .30$; $p < .01$) and Spelling ($r = .25$; $p < .05$). The Morphological Completion measure and the DMMA-Total exhibited roughly equivalent, strong correlations with Vocabulary. The DMMA-Total exhibited slightly larger correlations with the Elision, Word Identification, Word Attack, and Spelling measures than Morphological Completion. However, the Morphological Completion measure exhibited a larger correlation with Sentence Imitation than the DMMA-Total. Interestingly, Morphological Completion was not significantly correlated with Word Identification ($r = .21$; $p = .059$) like the DMMA-Total ($r = .31$; $p < 001$). In conclusion, the correlational results between the DMMA-Total and other measures followed the expected pattern of performance by demonstrating the strongest connection to the other morphological awareness measure, followed by the language measures, then literacy measures, and finally, the general cognitive measure.

Classification: Sensitivity and specificity. The degree to which the performance classifications between the DMMA-Total and the Sentence Imitation measure related can also inform the validity of the interpretations of overall DMMA performance. Like the above comparison with the Morphological Completion measure, it was important to explore whether the overall DMMA performance can predict classification outcomes of a distinct yet related language construct (i.e., general language ability) for the typically performing and at-risk performance groups. As discussed above, several cut-off scores for the DMMA-Total and subtasks were calculated using SPSS software (Version 21; IBM Corp., 2012) in order to provide a basis for comparison of the accurate classifications of the children's performance. A cut-off score of 341 (AUC = .764) provided the most balance among sensitivity (i.e., .77) and specificity (i.e., .80) for the DMMA-Total predicting performance group membership. Table 25 compares the classification predictions based on the children's performance on the DMMA-Total and the Sentence Imitation measure.

For example, Sentence Imitation and DMMA-Total with a cut-off score of 341 both classified ten children as at-risk while 13 children were classified as typically performing by Sentence Imitation but as at-risk by the DMMA-total at this same cut-off. This is an example of a classification error with this cut-off score resulting in over-identification of children into the at-risk group. There were relatively few errors of under-

Table 25

Classification prediction matrices for the Sentence Imitation and the DMMA-Total

DMMA	Sentence imitation			DMMA	Sentence imitation		
341 CO	AR	TP	Total	329 CO	AR	TP	Total
At-risk (AR)	10	13	23	At-risk (AR)	9	12	21
Typical (TP)	3	52	55	Typical (TP)	4	53	57
Total	13	65	78	Total	13	65	78

CO = Cut-off Score; Performance Groups - AR = At-Risk Group; TP = Typically performing group.

identification. These comparisons provide additional evidence for the relationship between the Sentence Imitation and DMMA-Total measures beyond the correlational evidence. In summary, the comparison of the sensitivity and specificity between the DMMA-Total and Sentence Imitation measure revealed acceptable levels of classification accuracy for identifying performance differences and provided support for the validity of the interpretations of overall DMMA performance.

Morphological Awareness Contributions to Literacy Skills

The second research question of the current study explores the contributions of morphological awareness as measured by the DMMA to literacy skills.

Correlations of Phonological Awareness and Morphological Awareness with Other Literacy Skills

The correlations among the language and literacy skills assessed in the current study are included in Table 24. For the entire sample, phonological awareness (i.e., the Elision measure) moderately correlated with the early language (i.e., Sentence Imitation $r = .60$; Vocabulary $r = .38$), and general cognition measures (i.e., Matrices $r = .28$). Furthermore, phonological awareness also moderately correlated with the literacy measures: Word Identification ($r = 0.47$), Word Attack ($r = 0.61$) and Spelling ($r = 0.53$). Finally, phonological awareness moderately correlated with the Morphological Awareness measures: Morphological Completion ($r = .44$) and DMMA ($r = .47$). All of these correlations were significant at either the $p < .05$ or $p < .01$ levels. All the correlations between the phonological awareness measure and language and literacy battery were greater than the correlations between morphological awareness and the same language and literacy measure(s).

Contributions to the Variance in Literacy Skills

Since phonological awareness and morphological awareness significantly correlated with each other and with word-level reading (i.e., a composite score of the Word Identification and Word Attack measures) and spelling, step-wise linear regressions were performed to examine the contributions of the independent variables, phonological awareness and morphological awareness, to the literacy skills (see Table 26). The first step-wise regression revealed that phonological awareness accounted for 34% of the unique significant variance in word-level reading ($R^2 = .34$; $p < .001$) after controlling for performance group with the whole equation accounting for 58% of the variance in word-level reading. The performance groupings did not significantly contribute to this equation ($p = .478$). When morphological awareness was added to this equation the overall variance increased by 2%; however, this change was not statistically significant (R^2 change = .017; $p = .170$). The performance groupings also remained non-significant ($p = .223$).

In a second step-wise regression, morphological awareness was entered first and incremental improvement was investigated by entering phonological awareness. This analysis revealed that morphological awareness accounted for 12% of the variance in word-level reading ($R^2 = .12$; $p = .002$) after controlling for performance groupings with the whole equation accounting for 35% of the variance in word-level reading. The

Table 26

Step-Wise Linear Regression Results

	Entire sample (N = 78)				Typically performing (n = 65)				At-risk (n = 13)			
Model	F	p	R^2	ΔR^2	F	p	R^2	ΔR^2	F	p	R^2	ΔR^2
Reading												
Model 1												
1st - PA	19.04	< .001	.34		24.93	< .001	.28		14.95	.003	.58	
2nd - MA	1.92	.170	.35	.017	3.43	.069	.32	.038	0.02	.883	.58	.001
Model 2												
1st - MA	10.63	.002	.12		6.61	.013	.10		1.74	.214	.14	
2nd - PA	13.21	< .001	.35	.231	20.64	< .001	.32	.226	10.41	.009	.58	.440
Spelling												
Model 1												
1st - PA	15.10	< .001	.29		23.72	< .001	.27		5.38	.041	.32	
2nd - MA	3.27	.075	.32	.030	6.49	.013	.34	.069	0.12	.735	.34	.008
Model 2												
1st - MA	12.36	.001	.14		10.18	.002	.14		1.42	.258	.11	
2nd - PA	9.61	< .001	.32	.177	19.16	< .001	.34	.203	3.35	.097	.34	.222

Notes. PA = phonological awareness; MA = morphological awareness.

performance groupings did not contribute significantly to this variance (p = .223). When phonological awareness was added to this equation the overall variance increased to 60% of accounted variance with phonological awareness adding 23% unique, significant variance (R^2 change = .231; p < .001) to this regression equation. Morphological awareness and the performance groupings did not remain statistically significant contributors to this equation.

Step-wise linear regressions were performed to examine the contributions of each independent variable (phonological awareness and morphological awareness) to spelling skills (see Table 26). The first step-wise regression revealed that phonological awareness accounted for 29% of the unique significant variance in word-level spelling (R^2 = .287; p < .001) after controlling for performance groupings with the whole equation accounting for 54% of the variance in word-level spelling. The performance groupings did not significantly contribute to this equation (p = .649). When morphological awareness was added to this equation the overall variance rose 3%, but morphological awareness did not appear to add any additional significant variance, although it approached significance (R^2 change = .030; p = .075). Phonological awareness remained significant while the performance groupings remained non-significant (p = .244).

A second step-wise regression was conducted for spelling. Morphological awareness was entered first and then incremental improvement was investigated by entering phonological awareness. This analysis revealed that morphological awareness accounted for 14% of the variance (R^2 = .140; p = .001) after controlling for performance groupings with the whole equation accounting for 37% variance in spelling. When phonological awareness was added to this equation the overall variance increased to 56% with phonological awareness adding 18% unique, significant variance (R^2 change = .177; $p < .001$) to this regression equation. The performance groupings did not contribute significantly to this variance ($p = .244$).

CHAPTER V

DISCUSSION

The purpose of the current study was to examine the validity of a dynamic measure of morphological awareness (DMMA) as an evaluative tool to assess morphological awareness (MA) in young children. Traditionally, the quality of a measure is comprised of two main constructs—validity and reliability. According to more recent standards, the definition of validity has been clarified and promoted to paramount importance and subsumes reliability (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014). The validity of the interpretations of performance on a measure is rooted in the degree to which performance on the measure can be interpreted as representing the construct that it purports to assess. Both claims that performance on a measure can be interpreted as an indication of the children's current level of skill and as an indication of future skill development in the targeted area require validation.

The current validity investigation began with an examination of the sample to ensure that it was appropriate for the subsequent validity testing. This validity investigation examined several aspects of validity information including: (a) evidence based on internal structure and reliability and (b) evidence based on relations to other variables (including another measure of MA and measures of other language and literacy constructs with various theoretical relations to MA). Each of these sources of validity evidence will be discussed more fully in the following sections.

Sample Characteristics

The performance distribution of the sample approximated the performance of the general population. For the standardized measures, the majority of the performance scores for both performance groups fell within the average range based on the normative sample of each individual measure. However, the at-risk group's mean for the Sentence Imitation and Elision measures fell slightly below the lower bound of the average range for these measures—this is not surprising because the at-risk group was selected based on low performance on the Sentence Imitation test. Although the phonological awareness Elision performance for this group was one standard deviation below the mean, this performance was more variable than the typically performing group as shown by the higher standard deviation. This finding suggested a wider range of phonological awareness skill for the children who struggle with developing language and literacy skills as compared to the children who are typically performing. In general, the typically performing group exhibited stronger performance on the language and literacy measures than the at-risk group as expected. However, the performance of the at-risk group should be interpreted with caution due to the small size of this subgrouping.

Moreover, the skewness of the sample's language and literacy performance exhibited a trend towards floor or ceiling effects, which indicated varying degrees of challenge for the children, either too easy or too difficult. For the typically performing group, the negative skewness statistic of their Word Identification performance was more than the acceptable standard, which indicates a clustering of performance towards higher values compared to the normal distribution. This finding was not surprising because an apparent increased skew can be caused by the added focus on the specific performance of a sample subgroup rather than the performance distribution of the whole sample; however, this apparent increase was only seen in the Sentence Imitation performance of the typically performing group versus all other measures of the language and literacy battery. Conversely, the positive skewness statistic of the Spelling performance for the typically performing group was more than the acceptable standard, which indicates the presence of a floor effect with these children's performance clustered towards lower values for the Spelling measure. This trend towards a floor effect could have been related to the emergent nature of this skill in young children. Young children are known to be in the initial stages of developing their literacy skill from simply interpreting the written word during reading to using the written word for expression through writing (Walker & Hauerwas, 2006).

Variability due to differences in location (EBLS, Wilson, or Adams), performance (typically performing v. at-risk), and/or gender (male v. female) could have impacted the interpretations of this study's results. Because of this, further analysis (i.e., main effects and interaction) was conducted to examine any differences in the sample performance based on these factors. In the current study, several significant main effects and interactions were found. The two performance groups showed statistically significant differences with the typically performing group scoring higher than the at-risk group on all measures except those that were focused on word-level reading (i.e., sight-word reading and decoding). This confirmed that our selection procedures were successful in distinguishing a group of students who are at-risk for difficulty developing language and literacy skills.

Overall, the results on the language and literacy measures appeared to approximate the performance expectations of the general population. The sample appeared to be appropriate for evaluating the validity of the DMMA. The distinct performance of the typically-developing and at-risk groups provided the opportunity to evaluate the characteristics of DMMA performance of these distinct groups as well as the broader sample.

Question #1: Validity of the Dynamic Measure of Morphological Awareness

As stated in the introduction to this chapter, the current investigation examined several aspects of validity for the DMMA including: (a) evidence based on internal structure and reliability and (b) evidence based on relations to other variables (including another measure of MA and measures of other language and literacy constructs with various theoretical relations to MA). Each of these sources of validity evidence will be discussed more fully in the following sections.

Evidence Based On Internal Structure

First, a source of evidence for the interpretation of DMMA performance as measuring morphological awareness stemmed from the analyses of the internal structure and reliability of the measure. The internal structure was demonstrated by the relationships among the children's performance on dissimilar (i.e., heterogeneity) and similar (i.e., homogeneity) stimulus items. Based on this underlying construct, we would expect high correlations among scores on similar stimulus types (i.e., targeting the same morphemes or morphemic approximations; either accurate-to-accurate or foil-to-foil items) because of their complementary meaning structure. Also, based on the underlying construct we would expect weak or no correlation among performance on dissimilar stimuli even though they were focused on the similar morphemic targets (i.e., true or approximated; accurate-to-foil items). Foil word pairs should demonstrate some correlation with the accurate word pairs since they approximated the targeted morpheme, but large, strong correlations were not expected because the foil pair only mimicked an accurate morphemic ending.

Large correlations between similar stimulus items (i.e., ranging from .76 to .87; accurate-to-accurate stimulus types or foil-to-foil stimulus types on the individual subtasks) were found as expected based on the construct. The correlations between similar stimulus types were much smaller than what was expected based on the theoretical construct between the DMMA subtasks than within each individual subtask (i.e., ranging from $r = -.01$ to .32; more comparable to the relationships among the dissimilar stimulus types). The correlations between the sample's performances on dissimilar stimulus items (i.e., accurate-to-foil stimulus types) targeting the same morpheme indicated a range of correlation from .18 to .32 (i.e., medium sized relationships). Moreover, a reversed pattern of performance for the at-risk group provided evidence for the different morphological awareness tasks having a varied impact performance. For example, some tasks resulted in stronger, increased performance (DMMA-RD for this group) versus weaker, decreased (DMMA-EP), but this variation in performance may not be a true, significant difference but a difference related to the way a construct was assessed.

Therefore, these findings provided evidence of the homogeneity of performance on similar stimulus types (i.e., accurate-to-accurate and foil-to-foil items) and heterogeneity of performance on dissimilar stimulus types (i.e., accurate-to-foil). However, the correlations of the similar stimulus items between the two DMMA subtasks were surprisingly low but this weak relationship can be more attributable to the difference in the targeted tasks rather than the stimulus type. This appeared to be a more plausible reason for this weak relationship since the relationships among the similar stimulus items were much closer to expectations. Overall, this performance heterogeneity and/or homogeneity supported the validity of the interpretation of DMMA performance measuring different aspects of morphological awareness in young children.

Evidence Based On Reliability

Second, as a subcomponent of validity, reliability provides evidence for the consistent measurement of performance as assessed by the targeted measure. Children's performance exhibits some variation in general whether within a single administration or between multiple administrations of a measure, which impacts measurement consistency. There were several different ways to test internal consistency reliability including the split-half, the coefficient alpha, and test-retest procedures. In this study, the DMMA items were divided (a) between the first half of the test and the second half, and (b) between the odd-numbered items and the even-numbered items. In addition to these two split half methods, coefficient alphas were also calculated. Coefficient alpha (a.k.a., Cronbach's alpha) represents the average of all the possible split-half reliabilities. Although some variation is seen in the interpretations of reliability outcomes, it is generally agreed that the acceptable level of reliability is .70 for exploratory research. Higher levels of the coefficient alpha are expected for basic research (i.e., .80) and applied research (i.e., .90) as per Nunnally (1978). The test-retest procedure, like the split-half procedure, compares the results of two halves of a measure that were administered at two separate times. The use of two separate forms of the DMMA provided an avenue for this reliability comparison.

Both tasks of the DMMA demonstrated strong internal consistency across all reliability comparison procedures. All split-half correlations were above .80 with 5 of 8 greater than .90. Similarly, all coefficient alpha reliabilities were greater than .80 with 7 of the 9 above .90. Finally, all test-retest reliabilities were also greater than .80 with 5 out of 7 above .90. Notably, all estimates of internal consistency were above .90 for the full DMMA. Although the expressive production task appeared to be slightly more reliable than the receptive discrimination task based on the higher internal consistency results with more reliabilities .90 or greater compared to .80 or greater for the DMMA-RD, this difference in reliability may be attributed to the need for the inclusion of foil word pairs that mimicked an accurate morphological relationship between base-root word and morphologically complex form, but were unrelated (e.g., *arm* and *army*) on the DMMA-RD versus the DMMA-EP to provide a basis for discrimination. The inclusion of these foil word pairs was contrary to the underlying assumption of the reliability procedures that all of the measure items strongly correlate. The larger reliabilities for the DMMA-EP may be due to the homogeneous nature with all items utilizing accurate morphologically-related word pairs. Overall, the reliabilities were strong indicating that the DMMA and its subtasks demonstrated strong internal consistency. The DMMA stimulus items appeared to measure the same construct of morphological awareness as seen in the similar interpretations of DMMA performance from both forms, e.g., forms A and B in the current study.

Interrater reliability also examined the reliability of DMMA results based on the amount of agreement between different scorers of an individual DMMA performance. Interrater reliability was of particular importance when scoring the responses that required subjective judgments. Because of the potential subjectivity in the DMMA-EP responses, the interrater reliability of this DMMA subtask using percent agreement and Cohen's kappa statistic was calculated through a cross-check of 30% of the children's performance that was balanced across location and task form. Cohen's kappa statistic (Cohen, 1960) calculates an even more rigorous level of interrater agreement by removing the agreement due to chance from the general agreement statistic. All the interrater agreement reliability calculations were 70% or above, which demonstrates an acceptable agreement between the two raters and provides evidence for the interrater reliability for the DMMA. Overall, each area of reliability (i.e., internal consistency, test-retest reliability, and interrater agreement) supported the validity of the interpretations of DMMA performance through the demonstration of consistent measurement of skill in young children.

Evidence Based On Response Distributions

Third, evidence for the valid interpretation of DMMA performance can also be gathered from examining the distribution of the children's performance. A valid measure of morphological awareness would capture the full range of proficiency from emergence to mastery—that is, it would show a smooth curve of performance with neither floor nor ceiling effects. Since both the Morphological Completion measure and DMMA-EP target morphological awareness expressively and the DMMA-Total was still being developed, the comparison of raw scores on these measures provided the best picture of performance equivalence among these measures. The DMMA performance for the entire sample revealed a negative skew of the DMMA-Total and DMMA-EP scores while the distribution of DMMA-RD scores was more symmetric.

A trend toward potential ceiling effects was noted in the highly-skewed distributions of the DMMA-Total and Morphological Completion measures for the entire sample. A large part of this skewness appears to be related to the inclusion of the highly-skewed DMMA-EP rather than the moderately skewed DMMA-RD that appears to add some balance to the measure as a whole. The highly-skewed findings for the DMMA-EP and Morphological Completion were not surprising since these tasks both target the expressive production of morphological complex forms that are known to be mastered by age five-years in spoken language (Brown, 1973), but the evidence is growing regarding the part that morphological skills play in literacy. To contrast, the DMMA-RD resulted in a more clustering of performance with less skew and variability (i.e., lower standard deviation) than the expressive morphological awareness tasks for the entire sample.

The comparison of raw score performance can also be made using the percentage of accurate responses. The comparison of the raw score performance of the entire sample on the DMMA with Morphological Completion measure revealed higher performance on the DMMA with an average score of 73% of the points possible on the full DMMA, 71% on the DMMA-RD, and 76% on the DMMA-EP when compared to 58% on the Morphological Completion measure. The current performance distribution without any apparent floor effects appeared to adequately measure low morphological awareness like that of the at-risk performance group. The measurement of higher levels of morphological performance appeared to be slightly limited as seen in the trend towards a possible ceiling effect, but this issue can be addressed by the addition of stimulus items focused on derivations for more challenge.

The performance groups also showed varied patterns of distribution in the DMMA subtasks as expected, although the DMMA-Total resulted in similar distributions. The typically performing group's distribution showed some slight negative skew in their DMMA-RD performance, but a more pronounced negative skew on the DMMA-EP was consistent with a trend towards a possible ceiling effect. These findings suggested that more challenging stimulus items, like those focused on additional derivational morphemes, need to be added to further differentiate the children with strong morphological awareness skills in the typically performing group. The at-risk group's distribution for both subtasks appeared normal. However, their DMMA-EP performance appeared bimodal demonstrating a trend towards both potential floor and ceiling effects. In conclusion, the DMMA and subtasks appeared to capture a wide range of performance variability in the current sample, but also appeared to result in different patterns for different performance groups. These findings suggested that the interpretations of DMMA performance may be more sensitive to measuring the morphological performance of at-risk children who exhibit weak and/or still emergent skill rather than the stronger skill of typically performing children. It is important to note, however, that the performance of the at-risk group should be interpreted with caution due to the small size of this subgrouping, which was well below accepted standards. Additionally, the addition of more stimuli focused on derivations could increase the differentiation of skill for the typically performing group.

Evidence Based On Prompt Use

The frequency of prompt usage indicates the amount of support needed by a child to achieve an accurate response. In general, the graduated prompting led to a response accuracy rate of 75% for the full DMMA and subtasks. The first and second prompts that represent the lowest levels of cueing (i.e., either no additional support or emphasis on the word pair only, respectively) were the most frequently effective for eliciting an accurate response from the children; however, the third and fourth prompts were also required for some children to achieve accuracy (i.e., successfully prompted about 10% of the responses). A high level of achievement indicated more independent morphological awareness shown in an accurate response than a lower level of achievement. Further examination of prompting may lead to more effective prompts especially for the at-risk group to reduce any potential floor effects and detect morphological impairment. Some variation in the prompt usage was noted with more prompting being required by the at-risk group to achieve an accurate response on the DMMA-RD when compared with the DMMA-EP, which may be slightly inflated due to the trend towards ceiling effect found especially for the typically performing group who exhibited less need for prompting. The current pattern of prompt use appeared to support morphological performance as seen in increased accurate responses especially for the at-risk group. The graduated prompts also provided an opportunity for the assessment of a wider range of morphological awareness skill as assessed by the DMMA that in turn supported the validity of this measure. In conclusion, these results supported the validity of the interpretations of DMMA performance. That is, the DMMA appeared to measure morphological awareness, and based on the current response distributions, the DMMA may assess various levels of morphological awareness achievement in young children.

Evidence Based On Correlations

Evidence about the valid interpretation of DMMA performance can also be found in the relationships between the DMMA and other measures of language and literacy skills. These other measures assessed either the same construct as the DMMA, morphological awareness (i.e., morphological completion), or related constructs (i.e., phonological awareness, vocabulary, general language ability, word-level reading and spelling, or general cognitive ability). The specific relationships will be discussed in the following sections.

Evidence Based On Relations Between Morphological Awareness Measures

The DMMA and Morphological Completion measure were both designed to assess morphological skill. All the correlations between the DMMA and Morphological Completion measure ranged in size from small to large (i.e., 0.42 to 0.77) and were significant to the $p < .01$ level for the entire sample. More specifically, large significant correlations were found between the DMMA-Total and DMMA-EP with the Morphological Completion measure. However, a different pattern was found between the DMMA-RD and Morphological Completion measures. The DMMA-RD appeared to add minimally to the correlation between the DMMA and the Morphological Completion measure. The DMMA-EP exhibited a stronger correlation with the Morphological Completion measure than did the total DMMA, which may be due to the similarity of the tasks since both target expressive production. The DMMA-Total included the additional discrimination task to factor into the relationship with the Morphological Completion measure.

One explanation for these findings could be that the DMMA-RD was measuring a separate component of morphological awareness that the DMMA-EP and the Morphological Completion measure do not. This explanation was consistent with a test design comprised of two separate tasks to target demonstration of both the expression and comprehension of morphological awareness. Since the DMMA-RD was designed to assess judgment rather than expression, a stronger relationship may be found between the DMMA-RD and another judgment-based task like a word analogy task as measures of morphological awareness; however, this is a point for future study. Equally important for the support of this explanation that the DMMA-RD can be validly interpreted as measuring a separate aspect of morphological awareness was the stimulus selection criterion and test design for both of the DMMA tasks implemented to control for other confounding variables (i.e., frequency, transparency, imageability, syntax length, and similarity).

Overall, these correlations provided evidence for the validity of the interpretations of DMMA performance since the Morphological Completion measure was an established, standardized measure of morphological awareness. Moreover, the correlations between the performance of the entire sample indicated that the DMMA and the Morphological Completion measure appeared to be assessing the same construct (i.e., morphological awareness). This finding also showed promising support for the use of dynamic assessment as an alternative to the traditional, static assessment approach.

Evidence Based On Relations to Other Skills

Validity evidence for the interpretations DMMA performance can also be found in its relationship to other literacy skills. Morphological awareness as demonstrated in the DMMA performance exhibited moderate, significant (r = .31-.74; p < .01) correlations with all measures, except with Matrices, i.e., the general cognitive ability measure. Because of a metalinguistic connection, language-based literacy skills, like phonological awareness and vocabulary, should exhibit stronger relationships to morphological awareness than other skills like general cognitive ability. Moreover, literacy-based skills, like word-level reading and spelling, should demonstrate weaker relationships with morphological awareness than the language-based skills because of the metalinguistic foundation of literacy skills yet a stronger relationship than general cognitive ability. This expected pattern of results was found in the current study.

Although the DMMA assessed the morphological awareness in the entire sample, the differences between the two performance groups showed slightly varied outcomes. The DMMA demonstrated a stronger relationship to the at-risk group's performance when compared to the typically performing group as indicated by the patterns of correlation between the DMMA and the other language and literacy measures for each performance group. The DMMA and the phonological awareness measure were significantly related for the typically performing group, but not for the at-risk group who may still be acquiring and developing both areas of skill. Similarly, word-level reading and spelling also moderately and significantly correlated with the DMMA for the typically performing group yet not significantly for the at-risk group, which could have been due to the latter group's language and literacy development and the restricted sample size as compared to the former group. The DMMA exhibited a stronger correlation to vocabulary for the at-risk group versus the typically performing group. This might be indicative of early developing morphological skill connected to awareness of meaning in whole units (i.e., words) before the discrimination of more specific, sub-units (i.e., morphemes). Therefore, further investigation is needed to examine the performance of various population subgroupings to inform the application of the DMMA as a screening measure of morphological awareness in young children.

The Morphological Completion measure exhibited weaker correlations with the other measures, including three statistically non-significant results between Morphological Completion measure and Matrices, Word Identification or Spelling measures. The Morphological Completion measure was related with the language-based literacy skills at levels comparable to those of the DMMA (i.e., r = .43-.66 versus r = .58-.65, respectively). Likewise, the expected pattern of results was exhibited by the literacy-based skills, these relationships were weaker than those of the DMMA (i.e., r = .21-.30 versus r = .31-.37). Finally, as expected the weakest relations were found between the Morphological Completion measure and the Matrices measure and was similar that of the DMMA (i.e., r = .15 versus r = .13). In conclusion, the pattern and strength of the correlations between the DMMA and the other language and literacy skills was consistent with the expected relationships based on the underlying constructs, which supports the idea that morphological awareness as assessed by the DMMA exhibited a pattern of relationships indicative of a valid measure of this construct.

Evidence Based On Classification Prediction

Last, the classification sensitivity and specificity were important to assessment and intervention tools. In order to calculate the sensitivity and specificity, cut-off scores were needed to separate the children into two separate groups. Since the Morphological Completion measure was standardized, a scaled score of 7 or below indicates impairment. However, cut-off scores need to be established for the DMMA and the subtasks. Several cut-off scores were used to calculate the classification statistics. The DMMA cut-off scores of 329 and 341 demonstrated the best classification prediction with high sensitivity to performance difference equal to 85.71% (i.e., typical performance vs. at-risk) that was also balanced with an acceptable level of specificity (i.e., ≥ 70%; range from 70.42% to 78.87%) when analyzed with the Morphological Completion classification predictions. These cut-off scores for the DMMA appeared to demonstrate high sensitivity and specificity versus the Morphological Completion measure, which indicated an accurate prediction of performance classification based on the interpretations of DMMA performance.

Additionally, as alluded to in prior sections, strong evidence exists for the classification of performance based on the Test of Language Development's Sentence Imitation task. The classification prediction sensitivity and specificity of DMMA performance was an important aspect of establishing the validity of the interpretations of this measure. The DMMA cut-off scores demonstrated strong sensitivity to performance difference equal to 76.92% (i.e., typical performance vs. at-risk) that was also balanced with an acceptable level of specificity (i.e., ≥ 70%; range from 70.77% to 80.00%). In sum, the sensitivity and specificity of the DMMA also supported the valid interpretations of the DMMA as a measure of morphological awareness.

In the current study, the moderately strong correlations between Sentence Imitation and DMMA performance for the at-risk group indicated a promising potential use of the DMMA as a screening measure of morphological awareness. As an established measure with strong sensitivity and specificity for classifying language performance (i.e., typically performing and at-risk for impairment or impaired), the Sentence Imitation subtest provided an important standard to measure the classification of the DMMA against since it can be potentially applied to these same categories. The strength and significance of these correlations suggested that given further study and examination, the DMMA could potentially be used to classify children's performance like the Sentence Imitation measure.

There are important caveats to this classification comparison between the DMMA and the Morphological Completion and Sentence Imitation measures (see below). First, as noted throughout, the performance of the at-risk group should be interpreted with caution due to the small size of this subgroup compared to the generally accepted standard. Second, although the intent was to compare the DMMA to a standardized measure, the MC subtest is not a gold-standard measure for this purpose. The MC measure was chosen because of its general use in the field of speech-language pathology, standardization, and inclusion of the targeted sample population. Future studies may expand this comparison to other morphologically-related measures like the Test of Early Grammar Impairment (TEGI) by Rice and Wexler (2001) and/or the Test for Examining Expressive Morphology (TEEM) by Shipley (1983). Finally, the classification outcomes used for the comparison of the Morphological Completion task and Sentence Imitation task were not individually reported in the test materials. The TOLD-P:4 Sentence Imitation task only reported the sensitivity (i.e., 0.74) and specificity (i.e., 0.88) statistics for the Spoken Language Index, a global composite of all the core subtests. No further information was provided regarding how the TOLD-P:4 established the validity of the measure. Thus, future investigation might include the exploration of a more direct comparison of the specific classification outcomes of these measures to each other and/or measures of similar constructs. In summary, the sensitivity and specificity of the DMMA compared to the Morphological Completion task supported the valid interpretations of the DMMA as a measure of morphological awareness.

Question #2: Contributions to Early Literacy Skills

Since phonological awareness has been proven to be an integral skill in literacy development and develops early in the literacy process, it was important to investigate and distinguish the role of morphological awareness (i.e., from interpretations of DMMA performance) in literacy skill beyond phonological awareness, especially in young children who are in the initial stages of developing their literacy skill. A finding of significant, unique contributions of morphological awareness to literacy skill would suggest a need to include an assessment of this skill within comprehensive early literacy assessment and provide a potential avenue for intervention that supports literacy development in young children. This research focus was intended to begin to establish the practical application of the interpretations of DMMA performance as an indicator of morphological awareness in young children. Interpretations could indicate that morphological awareness contributes to and/or facilitates literacy skill since it focuses on another metalinguistic aspect of literacy that goes beyond phonological awareness. This evidence was also important to establish in this age group whose morphological awareness skill is thought to be emerging versus fully developed. Finally, evidence based on a unique, significant contribution of morphological awareness to early literacy skill can facilitate the interpretations of DMMA performance as a predictor of future literacy skill development.

Since phonological awareness and morphological awareness moderately correlated in the total sample ($r = 0.47$; $p < .01$), an exploration of their potentially unique, significant roles in the early literacy skills of word-level reading and spelling was conducted. Phonological awareness and reading or spelling correlated moderately and significantly with a mean correlation of 0.53 ($p < .01$) and a range from 0.47 to 0.60 for the total sample. Similarly, morphological awareness and reading or spelling also moderately and significantly correlated with a mean correlation of 0.34 ($p < .01$) and a range from 0.31 to 0.37 for the total sample. The current results were consistent with the evidence-base that supports the critical role of phonological awareness in early literacy skills. Phonological awareness made larger contributions to reading and spelling when compared to morphological awareness. However, when morphological awareness was entered first into the equations, it contributed significant variance. The evidence is building to support this unique contribution, but it is often complicated by the amount already accounted for by other variables like phonological awareness, like the research findings of Apel et al. (2013), Deacon (2012), Kirby et al. (2012) who found respective variance to be 11% in Kindergarten, 1% in First Grade, and 4%-5% in Second and Third grade in sight-word reading, respectively. The current findings suggested a more complicated picture of the contributions of unique variance made by morphological awareness to literacy skills (i.e., word-level reading and spelling) over and beyond phonological awareness. Further investigation is needed to examine the unique, significant contributions that morphological awareness makes to early literacy based on the trends in the current results. These findings could be related to the emergent nature of morphological and phonological awareness and skill in young children who are in the initial stages of literacy development. The morphological performance may still be intermittent and variable as the young children progress towards mastery.

Limitations of the Current Study

It is important to acknowledge some limitations of the current study. First, although the sample size was sufficient when taken as a whole, the significant performance difference between the typically performing and at-risk groups required separate analyses for each group. Smaller sample sizes are less likely to reflect a true representation of the sampled population when compared to larger samples. Although the typically performing group was represented with an adequate sample size in the current study ($n = 65$), the at-risk group sample size was less than adequate (i.e., about half the generally acceptable minimum of 30). The results of the at-risk group performance group should be interpreted with caution because of the limited size of this group; however, the proportion of the children who typically performed to those who were at-risk for difficulty appears to mirror the general distribution of these performance groups in the wider population based on population statistics concerning students with special needs (National Center for Education Statistics, 2017). Although this limitation could be potentially mitigated by the use of qualification measures that can facilitate early grouping of the sample into performance groups and recruitment management of additional children, the status of the current sample exhibited some similarity with the general population so that the current findings and conclusions can be interpreted with cautious optimism. Furthermore, the study sample mirrored the demographics of the area from which it was drawn and lacked some of the racial/ethnic diversity typical of the general U.S. population, which may also limit the generalizability of these results to the general student population.

Future Directions

The findings of the current study point to several essential directions for future research. First, future studies should focus on the additional analysis of the content, form, and use of the individual stimuli, especially word types and sentence types, both statistically and by expert reviewers. Similarly, the relation between Form A and Form B should also be examined. The breadth of the current study precluded this additional in-depth analysis of the individual stimulus items. However, this in-depth analysis should be undertaken in order to examine any additional factors may directly impact overall performance. For example, the length and syntactic complexity of a certain stimulus item may result more frequently in an inaccurate response rather than just weak morphological skill. Moreover, these in-depth analyses can help to address alternative explanations for performance on the individual stimulus items that would then strengthen the support for the valid interpretations of DMMA performance as truly measuring morphological awareness.

Second, the DMMA performance could have been impacted by the type of tasks chosen to be included in the measure since the current evidence suggested that different morphological awareness tasks result in different patterns of performance. The DMMA tasks could have benefitted from comparison with another morphological awareness task such as a word analogy task with the DMMA-RD as both require judgment. The high performance for the typically performing group on the DMMA-EP signaled a potential trend towards the presence of a possible ceiling effect that may correspond to a higher level of mastery of morphological awareness. This pattern of performance may also indicate an emergent level of development required to establish mastery.

Additionally, only word-level reading outcome measures were included in the language and literacy battery. Word-level reading has limited connection to word meaning. An important characteristic of morphological awareness is its bridge between language form (i.e., words) and meaning. Future research could expand to consider providing comparison of morphological awareness in naturalistic, functional contexts (i.e., connected speech/language sample, text-level reading, or narratives). More naturalistic, functional contexts can provide additional benefit beyond the current study. These contexts might provide increased support for the use and understanding of morphological awareness through the added integration with meaning and linguistic structure like syntactical support within a sentence. However, morphological awareness may also be confounded by these other linguistic factors like syntactical structure and demand. This additional complexity may be worth the risk for a potentially more beneficial connection between morphological awareness and meaning. Morphological awareness is fostered through opportunities to use and increase children's understanding within naturalistic, functional contexts. Furthermore, these contexts facilitate both assessment and intervention focused on the acquisition, development, and generalization of morphological awareness skill. These contexts could be particularly important for the children in the current study who are in these initial stages of literacy as demonstrated by their emergent morphological awareness skill.

Third, future research should focus on confirming and extending the reliability and validity findings present in this study. Further exploration of the validity and classification analysis could inform educational programming (i.e., for curriculum planning and added support for student performance, respectively). The validity and reliability can be further explored by utilizing the items to the opposite DMMA task to provide a comparison of the individual stimulus item effectiveness (i.e., using the current stimulus items on the DMMA-RD for the DMMA-EP since the stimuli were developed for both tasks rather than for only one task specifically). Each individual stimulus item could then be assigned to the task that it elicited the strongest performance on or could provide a pool of stimulus items from which to create an alternative form, if too many items are strongest on one task versus the other. Using the stimuli on both tasks could provide a basis to compare item effectiveness that could potentially strengthen the full measure when items are sorted based on their strongest effect to measure performance. Future research could also explore and determine basal and ceiling items using item response theory, which could increase the ease of administration of the measure overall.

Fourth, the analysis of the effectiveness of the graduated prompting scheme was limited in the current study. Future research can build on the potential of the graduated prompting approach as a dynamic assessment technique by further analysis of the prompt effectiveness, especially by performance group, to generalize to other language and literacy measures. The current findings indicated that the prosodic prompt involving emphasis on the targeted word pair was the most effective level of prompt while also corresponding to the lowest level of support. The addition of a prosodic prompting level to other language and literacy measures could be a beneficial adaptation to utilize while assessing these skills in young children. More in-depth adaptations to current assessments could include visual picture prompts of targets but these would be more involved and may not be feasible for all measures. Additional methods of prompting like movement and tactile/physical prompting not included in the current graduated prompting scheme could also be explored for effectiveness in future research.

Finally, future research should focus on replication and extension of these findings to more diverse populations including special populations (e.g., language impaired, language/literacy impaired, reading disabled) especially since significant performance effects were found in the current sample as noted above.

Summary and Conclusion

The current study adds to the evidence base by exploring the validity of a new measure to assess morphological awareness dynamically and documenting morphological awareness in young children who were in the emergent stages of literacy acquisition. As outlined above, strong evidence was found to support the validity of interpretations of DMMA performance as measuring morphological awareness in young children based on sources from the test content, internal structure, reliability, and relations to other measures. The DMMA design included multiple elements of morphological awareness to adequately assess this construct in young children while also being sensitive to developmentally appropriate elements due to the emergent nature of this skill in this population. Dissimilar (i.e., accurate to foil word pairs) and similar (i.e., accurate to accurate) stimulus items were included as the basis to target both the receptive judgment and expressive production of morphological awareness to provide a more comprehensive look at the children's understanding and use of morphology. The reliability statistics showed strong, consistent measurement of DMMA performance within a single administration or across administrations. The DMMA demonstrated the ability to classify performance in ways similar to the Morphological Completion measure, an established, standardized measure of morphological awareness, which supports the claim that both measures targeted the same construct—morphological awareness—in young children. Furthermore, this finding was also supported by the large, significant correlation between the DMMA and the Morphological Completion measure. The moderate correlations between the DMMA and the other language and literacy measures also indicate that morphological awareness as assessed by the DMMA was related to early language and literacy acquisition and development in young children. Moreover, these patterns of correlation match the theoretically expected findings among the language and literacy constructs, which also supported the interpretations of DMMA performance as truly measuring morphological awareness. Finally, the findings indicated a more complicated picture of role that morphological awareness plays in early literacy based on the unique contributions it potentially makes to word-level reading and spelling. In conclusion, the DMMA appeared to demonstrate multiple sources of evidence that support the validity of the interpretations of DMMA performance as measuring morphological awareness at this exploratory stage of investigation.

The above evidence also indicated a potential benefit of dynamic assessment to assess the emergent skill of morphological awareness in young children. Significant performance effects indicated that the typically performing and at-risk groups demonstrated different patterns of performance across the language and literacy battery. However, the DMMA assessed morphological awareness in both performance groups, i.e., a wide range of skill. The DMMA provided additional opportunity for the at-risk group to demonstrate morphological awareness as compared to the Morphological Completion measure as evidenced in the frequency of prompt use. The DMMA exhibited larger, moderate correlations with the other language and literacy skills for the at-risk group's performance compared to Morphological Completion measure, but these findings should be interpreted with caution due to the small size of the at-risk group. In sum, dynamic assessment has potential as an alternative approach to assess language and literacy skills in young children to provide the opportunity for more performance and information beyond that of the current, traditional approach to assessment, especially for students who are at-risk for language and literacy difficulty.

REFERENCES

Adams, M. J. (1990). *Beginning to read: Learning and thinking about print*. Cambridge, MA: 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.

Anderson, R. C., & Nagy, W. E. (1989). *Word meanings (*Technical Report No. 485). Cambridge, MA: Center for the Study of Reading.

Anglin, J. (1980). Acquiring linguistic skills: A study of sentence construction in preschool children. In D. R. Olson (Ed.), *The social foundations of language and thought: Essays in honor of Jerome S. Bruner* (pp. 111-155). New York, NY: Norton.

Anglin, J. (1993). *Vocabulary development: A morphological analysis. Monographs of the Society for Research in Child Development, 58*. Retrieved from http://www.jstor.org/stable/1166112?seq=1#page_scan_tab_contents

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

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

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., & Masterson, J. J. (2001). Theory-guided spelling assessment and intervention: A case study. *Language, Speech, and Hearing Services in Schools*, *32*(3), 182-195.

Apel, K., Wilson-Fowler, E. B., Brimo, D., & Perrin, N. A. (2012). Metalinguistic contributions to reading and spelling in second and third grade students. *Reading and Writing*, *25*(6), 1283-1305.

Archibald, L. M., & Joanisse, M. F. (2009). On the sensitivity and specificity of nonword repetition and sentence recall to language and memory impairments in children. *Journal of Speech, Language, and Hearing Research*, *52*(4), 899-914.

Bear, D., Invernizzi, M., Templeton, S., & Johnston, F. (2008). *Words their way: Word study for phonics, spelling, and vocabulary instruction*. Columbus, OH: Merrill/ Macmillan.

Berko, J. (1958). The child's learning of English morphology. *Word*, 14, 150-177. doi: 10.1080/00437956.1958.11659661

Berninger, V. W., 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. https://doi.org/10.1007/ s10936-009-9130-6

Bowerman, M. (1982). Starting to talk worse: Clues to language acquisition from children's late speech errors. In S. Straus (Ed.), *U shaped behavioral growth* (pp. 101-145). New York, NY: Academic.

Bowey, J. A., & Francis, J. (1991). Phonological analysis as a function of age and exposure to reading instruction. *Applied Psycholinguistics*, *12*(01), 91-121.

Bridges, M. S. (2009). The use of a dynamic screening of phonological awareness to predict reading risk for kindergarten students (Doctoral dissertation). Available from ProQuest Dissertations & Theses database. (UMI 193989260). Retrieved from http://dist.lib.usu.edu/login?url=https://search-proquest-com.dist.lib.usu.edu/ docview/193989260?accountid=14761

Bridges, M., & Catts, H. (2010). *Dynamic screening of phonological awareness*. East Moline, IL: LinguiSystems.

Bridges, M. S., & Catts, H. W. (2011). The use of a dynamic screening of phonological awareness to predict risk for reading disabilities in kindergarten children. *Journal of Learning Disabilities*, *44*(4), 330-338. doi: 10.1177/0022219411407863.

Brown, A., Bryant, N., & Campione, J. (1983, April). *Preschool children's learning and transfer of matrices problems: Potential for improvement*. Presented at the Society for Research in Child Development meetings, Detroit, MI.

Brown, R. (1973). *A first language: The early stages.* Cambridge, MA: Harvard U. Press. Retrieved from http://doi.apa.org/psycinfo/1973-30971-000

Bryant, P., Nunes, T., & Bindman, M. (1998). Awareness of language in children who have reading difficulties: Historical comparisons in a longitudinal study. *Journal of Child Psychology and Psychiatry*, *39*(4), 501-510.

Budoff, M. (1974). *Learning potential and educability among the educable mentally retarded*. Cambridge, MA: Research Institute for Educational Problems, Cambridge Mental Health Association.

Budoff, M. (1987). The validity of learning potential assessment. In C. S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 53-81). New York, NY: Guilford.

Caffrey, E., Fuchs, D., & Fuchs, L. S. (2008). The predictive validity of dynamic assessment: A review. *The Journal of Special Education*, *41*(4), 254-270.

Campione, J. C. (1989). Assisted assessment: A taxonomy of approaches and an outline of strengths and weaknesses. *Journal of Learning Disabilities*, *22*(3), 151-165.

Campione, J. C., & Brown, A. L. (1987). Linking dynamic assessment with school achievement. In C. S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 82-115). New York, NY: Guilford.

Campione, J. C., Brown, A. L., & Ferrara, R. A. (1982). Mental retardation and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 392-490). New York, NY: Cambridge University Press.

Campione, J. C., Brown, A. L., Ferrara, R. A., Jones, R. S., & Steinberg, E. (1985). Breakdowns in flexible use of information: Intelligence-related differences in transfer following equivalent learning performance. *Intelligence*, *9*(4), 297-315.

Carlisle, J. F. (1988). Knowledge of derivational morphology and spelling ability in fourth, sixth, and eighth graders. *Applied Psycholinguistics*, *9*(03), 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: Erlbaum.

Carlisle, J. F. (1996). An exploratory study of morphological errors in children's written stories. *Reading and Writing*, *8*(1), 61-72.

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

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

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*(02), 177-195.

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

Carlisle, J. F., Stone, C. A., & Katz, L. A. (2001). The effects of phonological transparency on reading derived words. *Annals of Dyslexia*, *51*(1), 249-274.

Cazden, C. B. (1968). The acquisition of noun and verb inflections. *Child Development*, *39*(2), 433-448. https://doi.org/10.2307/1126956

Champion, A. (1997). Knowledge of suffixed words: A comparison of reading disabled and nondisabled readers. *Annals of Dyslexia*, *47*, 29-55.

Chesnick, M. A., Menyuk, P., D'Agostino, R., & Belanger, A. (1992, November). *Development of metalinguistic skills in children of varying language ability*. Paper presented at the Annual Convention of the American Speech-Language-Hearing Association, San Antonio, TX.

Cho, E., Compton, D. L., Gilbert, J. K., Steacy, L. M., Collins, A. A., & Lindström, E. R. (2017). Development of first-graders' word reading skills: For whom can dynamic assessment tell us more? *Journal of Learning Disabilities*, *50*(1), 95-112.

Clark, E. V. (1978). Awareness of language: Some evidence from what children say and do. In A. Sinclair, R. J., Jarvella, & W. J. M., Levelt (Eds.), *The child's conception of language* (pp. 17-43). Berlin-Heidelberg, Germany: Springer.

Clark, E. V. (1982). The young word maker: a case study of innovation in the child's lexicon. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: The state of the art* (p. 390). New York, NY: Cambridge University Press.

Clark, E. V., & Cohen, S. R. (1984). Productivity and memory for newly formed words. *Journal of Child Language*, *11*(03), 611-625.

Clark, E. V., & Hecht, B. F. (1982). Learning to coin agent and instrument nouns. *Cognition*, *12*(1), 1-24.

Clin, E., Wade-Woolley, L., & Heggie, L. (2009). Prosodic sensitivity and morphological awareness in children's reading. *Journal of Experimental Child Psychology*, *104*(2), 197-213.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37-46.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155.

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry*, *42*(6), 741-748.

Crocker, L., & Algina, J. (2008). *Introduction to classical & modern test theory*. Mason, OH: Cengage Learning.

Cunningham, A. E. (2006). Accounting for children's orthographic learning while reading text: Do children self-teach? *Journal of Experimental Child Psychology*, *95*(1), 56-77.

Cunningham, A. E., Perry, K. E., & Stanovich, K. E. (2001). Converging evidence for the concept of orthographic processing. *Reading and Writing*, *14*(5), 549-568.

Daniel, M. H. (1997). Intelligence testing: Status and trends. *American Psychologist*, *52*(10), 1038-1045.

Day, J. D., Engelhardt, J. L., Maxwell, S. E., & Bolig, E. E. (1997). Comparison of static and dynamic assessment procedures and their relation to independent performance. *Journal of Educational Psychology*, *89*(2), 358-368.

Day, J. D., & Zajakowski, A. (1991). Comparisons of learning ease and transfer propensity in poor and average readers. *Journal of Learning Disabilities*, *24*(7), 421-426.

de Groot, A. (1989). Representational aspects of word imageability and word frequency as assessed through word association. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *15*(5), 824-845.

De Villiers, J. G., & De Villiers, P. A. (1973). A cross-sectional study of the acquisition of grammatical morphemes in child speech. *Journal of Psycholinguistic Research*, *2*(3), 267-278.

Deacon, S. H. (2012). Sounds, letters and meanings: The independent influences of phonological, morphological and orthographic skills on early word reading accuracy. *Journal of Research in Reading*, *35*(4), 456-475.

Deacon, S. H., & Bryant, P. (2005). The strength of children's knowledge of the role of root morphemes in the spelling of derived words. *Journal of Child Language*, *32*(02), 375-389.

Deacon, S. H., & Kirby, J. R. (2004). Morphological awareness: Just "more phonological"? The roles of morphological and phonological awareness in reading development. *Applied Psycholinguistics*, *25*(2), 223-238.

Deacon, S. H., Whalen, R., & Kirby, J. R. (2011). Do children see the danger in dangerous? Grade 4, 6, and 8 children's reading of morphologically complex words. *Applied Psycholinguistics*, *32*(3), 467-481.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody picture vocabulary test: PPVT-4* (4th ed.). Bloomington, MN: Pearson.

Dye, C. D., Walenski, M., Prado, E. L., Mostofsky, S., & Ullman, M. T. (2013). Children's computation of complex linguistic forms: A study of frequency and imageability effects. *PloS One*, *8*(9), e74683. doi:10.1371/journal.pone.0074683.

Ehri, L. C. (2000). Learning to read and learning to spell: Two sides of a coin. *Topics in Language Disorders*, *20*(3), 19-36.

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly*, *36*(3), 250-287.

Ehri, L. C., & Wilce, L. S. (1985). Movement into reading: Is the first stage of printed word learning visual or phonetic? *Reading Research Quarterly*, *20,* 163-179.

Elbro, C., & Arnbak, E. (1996). The role of morpheme recognition and morphological awareness in dyslexia. *Annals of Dyslexia*, *46*(1), 209-240.

Elliott, J. (2003). Dynamic assessment in educational settings: Realizing potential. *Educational Review*, *55*(1), 15-32.

Embretson, S. E. (1987). Toward development of a psychometric approach. In C. S. Lidz (Ed.), *Dynamic testing* (pp. 141-172). New York, NY; Guilford.

Ferrara, R. A., Brown, A. L., & Campione, J. C. (1986). Children's learning and transfer of inductive reasoning rules: Studies of proximal development. *Child Development*, *57,* 1087-1099.

Feuerstein, R., Rand, Y., Hoffman, M., & Miller, R. (1980). *Instructional enrichment*. Baltimore, MD: University Park.

Fowler, A. E. (1988). Grammaticality judgments and reading skill in grade 2. *Annals of Dyslexia*, *38*(1), 73-94.

Freyd, P., & Baron, J. (1982). Individual differences in acquisition of derivational morphology. *Journal of Verbal Learning and Verbal Behavior*, *21*(3), 282-295.

Gleason, J. B. (2012). *The development of language* (8th ed.). Boston, MA: Allyn & Bacon.

Gonnerman, L. M., Seidenberg, M. S., & Andersen, E. S. (2007). Graded semantic and phonological similarity effects in priming: Evidence for a distributed connectionist approach to morphology. *Journal of Experimental Psychology: General*, *136*(2), 323-345.

Good, R. H., Kaminski, R. A., Smith, S., Laimon, D., & Dill, S. (2003). *Dynamic indicators of basic early literacy skills*. Longmont, CO: Sopris West Educational Services.

Goswami, U. (2000). Phonological representations, reading development and dyslexia: towards a cross-linguistic theoretical framework. *Dyslexia*, *6*(2), 133-151.

Grigorenko, E. L., & Sternberg, R. J. (1998). Dynamic testing. *Psychological Bulletin*, *124*(1), 75.

Gutierrez-Clellen, V. F., & Pena, E. (2001). Dynamic assessment of diverse children: A tutorial. *Language, Speech & Hearing Services in Schools*, *32*(4), 212-224.

Hammill, D., & Newcomer, P. (2008). *Test of language development-primary:4 (TOLD-P:4)* (4th ed.). Austin, TX: Pro-Ed.

Hasson, N., & Joffe, V. (2007). The case for dynamic assessment in speech and language therapy. *Child Language Teaching and Therapy*, *23*(1), 9-25.

Haywood, H. C., & Tzuriel, D. (1992). Epilogue: The status and future of interactive assessment. In H. C. Haywood & D. Tzuriel (Eds.), *Interactive assessment* (pp. 38-63). New York, NY: Springer-Verlag.

Hessels, M. G. (1997). Low IQ but high learning potential: Why Zeyneb and Moussa do not belong in. *Educational and Child Psychology*, *14*(4), 121-146.

Huck, S. W. (2012). *Reading statistics and research* (6th. ed.). Boston, MA: Pearson Education.

Jitendra, A. K., & Kameenui, E. J. (1993). Dynamic assessment as a compensatory assessment approach: A description and analysis. *Remedial and Special Education*, *14*(5), 6-18.

Justice, L. M., & Ezell, H. K. (2004). Print referencing: An emergent literacy enhancement strategy and its clinical applications. *Language Speech and Hearing Services in Schools*, *35*(2), 185-193.

Katz, R. B., Shankweiler, D., & Liberman, I. Y. (1981). Memory for item order and phonetic recoding in the beginning reader. *Journal of Experimental Child Psychology*, *32*(3), 474-484.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman brief intelligence test: K-BIT-2* (2nd ed.). Bloomington, MN: American Guidance Service.

Kemper, R. L. (1985). *Metalinguistic correlates of reading ability in second grade children* (Doctoral dissertation). Kent State University, Kent, OH.

Kester, E. S., Peña, E. D., & Gillam, R. B. (2001). Outcomes of dynamic assessment with culturally and linguistically diverse students: A comparison of three teaching methods within a test-teach-retest framework. *Journal of Cognitive Education and Psychology*, *2*(1), 37-54.

Kintsch, W., & Kozminsky, E. (1977). Summarizing stories after reading and listening. *Journal of Educational Psychology*, *69*(5), 491-499.

Kirby, J. R., Deacon, S. H., Bowers, P. N., Izenberg, L., Wade-Woolley, L., & Parrila, R. (2012). Children's morphological awareness and reading ability. *Reading and Writing*, *25*(2), 389-410.

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

Larsen, J. A., & Nippold, M. A. (2007). Morphological analysis in school-age children: Dynamic assessment of a word learning strategy. *Language, Speech, and Hearing Services in Schools*, *38*(3), 201-212.

Laughon, P. (1990). The dynamic assessment of intelligence: A review of three approaches. *School Psychology Review, 19*, 459-470.

Leonard, L. B. (1998). *Children with specific language impairment*. Cambridge, MA: MIT press.

Levin, J. R., & Pressley, M. (1983). *Understanding mnemonic imagery effects: A dozen "obvious" outcomes*. (M. L. Fleming & D. W. Hutton, Eds.). Englewood Cliffs, NJ: Educational Technology.

Libben, G., Gibson, M., Yoon, Y. B., & Sandra, D. (2003). Compound fracture: The role of semantic transparency and morphological headedness. *Brain and Language*, *84*(1), 50-64.

Liberman, I. Y., Rubin, H., Duques, S., & Carlisle, J. (1985). Linguistic abilities and spelling proficiency in kindergartners and adult poor spellers. *Biobehavioral Measures of Dyslexia*, *1*, 163-176.

Lidz, C. S., & Pena, E. D. (1996). Dynamic assessment: The model, its relevance as a non-biased approach, and its application to Latino American preschool children. *Language, Speech, and Hearing Services in Schools*, *27*(4), 367-372.

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

Manis, F. R., Doi, L. M., & Bhadha, B. (2000). Naming speed, phonological awareness, and orthographic knowledge in second graders. *Journal of Learning Disabilities*, *33*(4), 325-333.

Mann, V. A., Shankweiler, D., & Smith, S. T. (1984). The association between comprehension of spoken sentences and early reading ability: The role of phonetic representation. *Journal of Child Language*, *11*(03), 627-643.

Masterson, J., Druks, J., & Gallienne, D. (2008). Object and action picture naming in three-and five-year-old children. *Journal of Child Language*, *35*(02), 373-402.

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

Miles, J., & Shevlin, M. (2001). *Applying regression and correlation: A guide for students and researchers*. London, UK: Sage.

Miller, J. F., & Chapman, R. S. (1981). The relation between age and mean length of utterance in morphemes. *Journal of Speech, Language, and Hearing Research*, *24*(2), 154-161.

Missiuna, C., & Samuels, M. T. (1989). Dynamic assessment of preschool children with special needs: Comparison of mediation and instruction. *Remedial and Special Education*, *10*(2), 53-60.

Moats, L. C., & Smith, C. (1992). Derivational morphology: Why it should be included in language assessment and instruction. *Language, Speech, and Hearing Services in Schools*, *23*(4), 312-319.

Nagy, W. E., & Anderson, R. C. (1984). How many words are there in printed school English? *Reading Research Quarterly*, *19,* 304-330.

Nagy, W., Anderson, R. C., Schommer, M., Scott, J. A., & Stallman, A. C. (1989). Morphological families in the internal lexicon. *Reading Research Quarterly*, *24*, 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. https://doi.org/http://dx.doi.org.dist.lib.usu.edu/10.1037/0022-0663.98.1.134

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*(4), 730-742.

Nagy, W. E., Carlisle, J. F., & Goodwin, A. P. (2014). Morphological knowledge and literacy acquisition. *Journal of Learning Disabilities*, *47*(1), 3-12. https://doi.org/10.1177/0022219413509967

National Center for Education Statistics. (2017). *The condition of education: Children and youth with disabilities.* Retrieved from https://nces.ed.gov/programs/coe/indicator_cgg.asp

National Reading Panel, National Institute of Child Health, & Human Development. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Rockville, MD: National Institute of Child Health and Human Development, National Institutes of Health.

Nelson, N. W. (1998). *Childhood language disorders in context: Infancy through adolescence* (2nd ed.). Boston, MA: Allyn & Bacon.

Nippold, M. A. (2007). *Later language development: School-age children, adolescents, and young adults*. Austin, TX: Pro-ed.

Nippold, M. A., Allen, M. M., & Kirsch, D. I. (2001). Proverb comprehension as a function of reading proficiency in preadolescents. *Language, Speech, and Hearing Services in Schools*, *32*(2), 90-100.

Nippold, M. A., Moran, C., & Schwarz, I. E. (2001a). Idiom understanding in preadolescents: Synergy in action. *American Journal of Speech-Language Pathology*, *10*(2), 169-179.

Nippold, M., Moran, C., & Schwarz, I. (2001b). Proverb comprehension in New Zealand and American youth: A cross-cultural comparison. *New Zealand Journal of Speech-Language Therapy*, *55*, 14-21.

Nippold, M. A., & Sun, L. (2008). Knowledge of morphologically complex words: A developmental study of older children and young adolescents. *Language, Speech, and Hearing Services in Schools*, *39*(3), 365-373.

Nunes, T., & Bryant, P. (2006). *Improving literacy by teaching morphemes*. New York, NY: Routledge.

Nunes, T., Bryant, P., & Barros, R. (2012). The development of word recognition and its significance for comprehension and fluency. *Journal of Educational Psychology*, *104*(4), 959-973.

Nunes, T., Bryant, P., & Bindman, M. (2006). The effects of learning to spell on children's awareness of morphology. *Reading and Writing*, *19*(7), 767-787.

Nunnally, J. C. (1978). *Psychometric theory*. New York, NY: McGraw-Hill.

Oetting, J. B., McDonald, J. L., Seidel, C. M., & Hegarty, M. (2016). Sentence recall by children with SLI across two nonmainstream dialects of English. *Journal of Speech, Language, and Hearing Research*, *59*(1), 183-194.

Olswang, L. B., & Bain, B. A. (1996). Assessment information for predicting upcoming change in language production. *Journal of Speech, Language, and Hearing Research*, *39*(2), 414-423.

Oregon Department of Education. (2005). *Oregon Statewide Assessment*. Salem, OR: Office of Assessment and Evaluation.

Owens, R. E., Jr. (1996). *Language development: An introduction* (4th ed.). Boston, MA: Allyn & Bacon.

Paivio, A. (1983). The empirical case for dual coding. In J. C. Yuille (Ed.), *Imagery, Memory and Cognition*: Essays in honor of Allan Paivio (pp. 307-332). Hillsdale, NJ: Lawrence Erlbaum.

Paivio, A. (1991). *Images in mind: The evolution of a theory.* New York, NY: Harvester Wheatsheaf.

Paul, R. (1995). *Language disorders from infancy through adolescence: Assessment and intervention.* St. Louis, MO: Mosby.

Pena, E. D. (2000). Measurement of modifiability in children from culturally and linguistically diverse backgrounds. *Communication Disorders Quarterly*, *21*(2), 87-97.

Peña, E., & Gillam, R. (2000). Dynamic assessment of children referred for speech and language evaluations. In C. S. Lidz & J. Elliott (Eds.), *Dynamic assessment: Prevailing models and applications* (Vol. 6). Oxford, UK: Elsevier Science.

Peña, E., Iglesias, A., & Lidz, C. S. (2001). Reducing test bias through dynamic assessment of children's word learning ability. *American Journal of Speech-Language Pathology*, *10*(2), 138-154.

Pena, E., Quinn, R., & Iglesias, A. (1992). The application of dynamic methods to language assessment: A non-biased approach. *Journal of Special Education, 26,* 269-280.

Pence-Turnbull, K. L., & Justice, L. M. (2011). *Language development from theory to practice*. Boston, MA: Pearson Higher Ed.

Petersen, D. B., Allen, M. M., & Spencer, T. D. (2016). Predicting reading difficulty in first grade using dynamic assessment of decoding in early kindergarten: A large-scale longitudinal study. *Journal of Learning Disabilities*, *49*(2), 200-215.

Petersen, D. B., & Gillam, R. B. (2015). Predicting reading ability for bilingual Latino children using dynamic assessment. *Journal of Learning Disabilities*, *48*(1), 3-21.

Peterson, R. L., Pennington, B. F., Shriberg, L. D., & Boada, R. (2009). What influences literacy outcome in children with speech sound disorder? *Journal of Speech, Language, and Hearing Research*, *52*(5), 1175-1188.

Prado, E. L., & Ullman, M. T. (2009). Can imageability help us draw the line between storage and composition? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 849-866.

Quemart, P., Wolter, J. A., Deacon, S. H., & Chen, B. (2017, July). *The development of morphological representations in school-age children: A cross-language study in a symposium entitled 'Individual Differences in Word Learning.'* Paper presented at the 24th Annual Conference of the Scientific Studies of Reading, Halifax, Nova Scotia.

Ram, G., Marinellie, S. A., Benigno, J., & McCarthy, J. (2013). Morphological analysis in context versus isolation: Use of a dynamic assessment task with school-age children. *Language, Speech, and Hearing Services in Schools*, *44*(1), 32-47.

Reinharth, B. M. (1989). *Cognitive modifiability in developmentally delayed children* (Unpublished doctoral dissertation). Yeshiva University, New York, NY.

Resing, W. (1993). Measuring inductive reasoning skills: The construction of a learning potential test. In J. H. M. Hamer, K Sijtsma, & A. J. J. M. Ruijssenaars (Eds.), *Learning Potential Assessment*, (pp. 219-242). Amsterdam, The Netherlands: Swets & Zeitlinger.

Resing, W. (1997). Learning potential assessment: The alternative for measuring intelligence? *Educational and Child Psychology*, *14*, 68-82.

Resing, W., & Van Wijk, A. (1996). Leerpotentieel: Onderzoekbij allochtone leerlingen uit het basisonderwijs.[Learning potential: Research with minority children in primary education]. *Tijdschrift Voor Orthopedagogiek*, *35*, 432-442.

Rice, M., & Wexler, K. (2001). *Rice Wexler Test of Early Grammatical Impairment*. San Antonio, TX: Hove.

Roman, A. A., Kirby, J. R., Parrila, R. K., Wade-Woolley, L., & Deacon, S. H. (2009). Toward a comprehensive view of the skills involved in word reading in Grades 4, 6, and 8. *Journal of Experimental Child Psychology*, *102*(1), 96-113.

Rothman, H., Semmel, M., & Gaylord-Ross, R. (1990). Dynamic assessment: A comprehensive review of literature. *Issues and Research in Special Education*, *1*, 355-386.

Rubin, H., Patterson, P. A., & Kantor, M. (1991). Morphological development and writing ability in children and adults. *Language, Speech, and Hearing Services in Schools*, *22*(4), 228-235.

Schwiebert, C., Green, L., & McCutchen, D. (2002). *The contribution of morphology to reading and spelling achievement.* Retrieved from http://eric.ed.gov/?id= ED465986

Seeff-Gabriel, B., Chiat, S., & Dodd, B. (2010). Sentence imitation as a tool in identifying expressive morphosyntactic difficulties in children with severe speech difficulties. *International Journal of Language & Communication Disorders*, *45*(6), 691-702.

Seymour, P. H., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*(2), 143-174.

Shankweiler, D., Liberman, I. Y., Mark, L. S., Fowler, C. A., & Fischer, F. W. (1979). The speech code and learning to read. *Journal of Experimental Psychology: Human Learning and Memory*, *5*(6), 531-545.

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition*, *55*(2), 151-218.

Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology*, *72*(2), 95-129.

Share, D. L. (2004). Orthographic learning at a glance: On the time course and developmental onset of self-teaching. *Journal of Experimental Child Psychology*, *87*(4), 267-298.

Share, D. (2011). On the role of phonology in reading acquisition: The self-teaching hypothesis. In S. Brady, D. Braze, & C. Fowler (Eds.), *explaining individual differences in reading: Theory and evidence* (pp. 45-68). New York, NY: Psychology Press.

Sharoni, V., & Greenfeld, T. (1999). Applications of dynamic assessment and mediated learning principles in a reading remedial workshop: Case study. In D. Tzuriel (Ed.), *Mediated learning experience: Theory applications and research.* (pp. 121-141). Haifa: Ach and the International Center for Enhancement of Learning Potential.

Shipley, K. G. (1983). *Test for examining expressive morphology (TEEM)*. Online open library: Communication Skill Builders.

Siegel, L. S. (2008). Morphological awareness skills of English language learners and children with dyslexia. Topics in Language Disorders, *28*(1), 15-27.

Singson, M., Mahony, D., & Mann, V. (2000). The relation between reading ability and morphological skills: Evidence from derivational suffixes. *Reading and Writing*, *12*(3), 219-252.

Speece, D. L., Cooper, D. H., & Kibler, J. M. (1990). Dynamic assessment, individual differences, and academic achievement. *Learning and Individual Differences*, *2*(1), 113-127.

Spencer, T. D., Petersen, D. B., & Adams, J. L. (2015). Tier 2 language intervention for diverse preschoolers: An early-stage randomized control group study following an analysis of response to intervention. *American Journal of Speech-Language Pathology*, *24*(4), 619-636.

Sternberg, R. J. (1996). *Successful intelligence: How practical and creative intelligence determine success in life*. New York, NY: Simon & Shuster.

Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1140-1154.

Swanson, H. L., & Lussier, C. M. (2001). A selective synthesis of the experimental literature on dynamic assessment. *Review of Educational Research*, *71*(2), 321-363.

Swanson, H. L., Trainin, G., Necoechea, D. M., & Hammill, D. D. (2003). Rapid naming, phonological awareness, and reading: A meta-analysis of the correlation evidence. *Review of Educational Research*, *73*(4), 407-440.

Tunmer, W. E., Herriman, M. L., & Nesdale, A. R. (1988). Metalinguistic abilities and beginning reading. *Reading Research Quarterly*, *69*(3), 134-158.

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

Tzuriel, D. (2000). Dynamic assessment of young children: Educational and intervention perspectives. *Educational Psychology Review*, *12*(4), 385-435.

Tzuriel, D., & Caspi, N. (1992). Cognitive modifiability and cognitive performance of deaf and hearing preschool children. *The Journal of Special Education*, *26*(3), 235-252.

Tzuriel, D., & Kaufman, R. (1999). Mediated learning and cognitive modifiability dynamic assessment of young Ethiopian immigrant children to Israel. *Journal of Cross-Cultural Psychology*, *30*(3), 359-380.

Tzuriel, D., & Klein, P. S. (1985). The assessment of analogical thinking modifiability among regular, special education, disadvantaged, and mentally retarded children. *Journal of Abnormal Child Psychology*, *13*(4), 539-552.

Tzuriel, D., & Klein, P. S. (1987). Assessing the young child: Children's analogical thinking modifiability. In C. S. Lidz (Ed.), *Dynamic assessment* (pp. 268-282). New York, NY: Guilford.

U.S. Census Bureau. (2015a). *Quick facts: Logan City, UT*. Retrieved from https://www.census.gov/quickfacts/fact/table/logancityutah,US/PST045216

U.S. Census Bureau. (2015b). *Educational attainment in the United States: 2015*. Retrieved from https://www.census.gov/content/dam/Census/library/publications/2016/demo/p20-578.pdf

Valtin, R. (1984). Awareness of features and functions of language. In *Language awareness and learning to read* (pp. 227-260). New York, NY: Springer.

Van Kleeck, A. (1982). The emergence of linguistic awareness: A cognitive framework. *Merrill-Palmer Quarterly 28*, 237-265.

Vogel, S. A. (1983). A qualitative analysis of morphological ability in learning disabled and achieving children. *Journal of Learning Disabilities*, *16*(7), 416-420.

Vygotsky, L. S. (1978). *Mind in society: The development of higher mental process*. (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Cambridge, MA: Harvard University Press.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2013). *Comprehensive Test of Phonological Processing: CTOPP-2* (2nd ed.). Austin, TX: ProEd.

Walker, J., & Hauerwas, L. B. (2006). Development of phonological, morphological, and orthographic knowledge in young spellers: The case of inflected verbs. *Reading and Writing*, *19*(8), 819-843.

Webster, P. E., & Plante, A. S. (1992). Effects of phonological impairment on word, syllable, and phoneme segmentation and reading. *Language, Speech, and Hearing Services in Schools*, *23*(2), 176-182.

Wiig, E. H., Semel, E. M., & Crouse, M. A. (1973). The use of English morphology by high-risk and learning disabled children. *Journal of Learning Disabilities*. Retrieved from http://psycnet.apa.org/psycinfo/1974-05491-001

Windsor, J. (2000). The role of phonological opacity in reading achievement. *Journal of Speech, Language, and Hearing Research*, *43*(1), 50-61.

Wolter, J. A. (2014). Imageability and transparency in morphological awareness: A study of how third-grade children made lemonade from lemon. *Topics in Language Disorders*, *34*(3), 228-239.

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., & Dzatko, 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.

Woodcock, R. W. (2011). *Woodcock reading mastery test: WRMT-3* (3rd ed.). Circle Pines, MN: American Guidance Service.

Zeno, S., Ivens, S., Millard, R., & Duvvuri, R. (1995). *The educator's word frequency guide*. Brewster, NY: Touchstone Applied Science Associates.

APPENDICES

Appendix A

DMMA-RD: Standard Instructional Protocol for Task Introduction

DMMA-RD: Standard Instructional Protocol for Task Introduction

Introduction

Examiner: We're going to play a word game.

Here are smiley pictures to go with your answers.
(Explanation of cards) - After I say a sentence…
If it makes sense—point to this one—the full smile on the dark green (Point to full smiley w/dark green).
If it is silly—point to this one—the silly smile with the tongue out on the dark red (Point to silly smiley w/dark red).
If you don't know—point to this one—the straight smile on the yellow (Point to straight smiley w/yellow).
If you think it's sort of makes sense, but you are not sure—point to this one—the side smile on the light green (Point to side smiley w/light green).
If you think it's sort of silly, but you are not sure—point to this one—the winking smile on the light red (Point to winking smiley w/light red).

Note to examiner: Both practice items need to be administered to acclimate the examinees to the two kinds of question

Remember after I say something, you point to which smiley goes with your answer…Okay, let's try one…Listen: *If there is more than one egg, there are eggs.* Does this make sense or is it silly?

Practice Item—Prompting Schema

Prompt #1

> Accurate—"Great work. Let's try another one…" Then present Practice Item B.
> Inaccurate—"Good job. How about if I say it this way? Then, administer Prompt #2.

Prompt #2—Administer Prompt #2 with the added prompting of "...Does that change your answer?"

> Accurate - "Great work. Let's try some more"
> Inaccurate - "Good job. How about if I say it this way? Then, administer Prompt #3.

Prompt #3 - Administer Prompt #3 with the added prompting of "...Does that change your answer?

Accurate - “Great work. Let’s try some more”
Inaccurate - “Good job. How about if I say it this way? Then, administer Prompt #4.

Prompt #4 - Administer Prompt #4 with the added prompting of “...Does that change your answer?

Accurate - “Great work. Let’s try some more”
Inaccurate - “Good job. Let’s try some more"

Appendix B

DMMA-RD: Practice Items and Stimuli

DMMA-RD: Practice Items & Stimuli

Practice Items

A. If there is more than one egg, there are eggs.
B. If there is an arm, it is army.

Stimuli - Form A

1. If there is more than one eye, there are eyes.
2. A person who likes to climb is a climber.
3. After you drive a car, it is card.
4. After the boat sets sail, it is sailing.
5. If there is more than one boat, there are boats.
6. After she got the bean count, the beans were counted.
7. When the machine is set on the wash cycle, it is washing.
8. If there is more than one let, there is lettuce.
9. If there is a smile, it is smiley.
10. If there is more than one glass, there are glasses.
11. A person who likes to dance is a dancer.
12. If it happens once a month, it happens monthly.
13. A person who takes care of a moth is a mother
14. After you save your money, it is saved.
15. If there is a grave, it is gravy.
16. If the lamp is off, it is offer.
17. If there is more than one fish, there are fishes.
18. After he puts a block in front of his toy car, it is blocked.
19. If it is your own, it is only.
20. If there is rain, it is rainy.
21. If he is your friend, you are friendly.
22. When you use a box, you are boxing.
23. If there is more than one bed, there are beds.
24. If there is care, it is carry.
25. If there is more than one cat, there are cats.
26. A person who asks, “May I?” is a mayor.
27. When you win, you are wing.
28. If the feather is light, it will fall down lightly.
29. Since he likes to eat corn, he is a corner.
30. If there is hair, it is hairy.

Stimuli - Form B -

31. If there is more than one cow, there are cows.
32. If she took a pill, she is a pillar.
33. After you print something, it is printed.
34. If there is a bull, it is a bully.

35. After you start to push the door open, it needs more pushing.
36. After you found a ten, it is tent.
37. If there is a bell, it is belly.
38. If it happens at night, it happens nightly.
39. When you are too thin, you are thing.
40. If there is more than one circle, there are circles.
41. If you are on time, you are timely.
42. Turning up the volume makes a loud noise louder.
43. After he asks them, “which costumes do they want?”; They will choose the witches.
44. When you eat dinner, you are eating.
45. If there is more than one ten, there is tennis.
46. If there is a part, it is party.
47. A person who makes a doll is a dollar.
48. If you have more than one pal, you have palace.
49. If they are safe, they cross the road safely.
50. After you watch a cartoon, it is watched.
51. If there is eight, it is eighty.
52. After you clean something, it is cleaned.
53. If there is more than one no, there are nose.
54. When you print your name, you are printing.
55. If you add sugar to a sweet dessert, it is sweeter.
56. If there is more than one box, there are boxes.
57. If they like to go east, they are Easter.
58. After he gives the door a push, it is pushed.
59. If there is more than one hole, there are holes.
60. If there is dirt, it is dirty.

Appendix C

DMMA-EP: Standard Instructional Protocol for Task Introduction

DMMA-EP: Standard Instructional Protocol for Task Introduction

Examiner: We're going to play a word game. First, let's use the word *star*. Listen to this sentence: *Star. The night sky was full of* _____. Can you put the word *star* into this sentence so that it makes sense?

Examiner:

Practice Item—Prompting Schema

Prompt #1

Accurate—"Great work. Let's try another one…" Then present Practice Item B.

Inaccurate—"Good job. How about if I say it this way? Then, administer Prompt #2.

Prompt #2—Administer Prompt #2 with the added prompting of "...Does that change your answer?"

Accurate - "Great work. Let's try some more"

Inaccurate - "Good job. How about if I say it this way? Then, administer Prompt #3.

Prompt #3 - Administer Prompt #3 with the added prompting of "...Does that change your answer?

Accurate - "Great work. Let's try some more"

Inaccurate - "Good job. How about if I say it this way? Then, administer Prompt #4.

Prompt #4 - Administer Prompt #4 with the added prompting of "...Does that change your answer?

Accurate - "Great work. Let's try some more"

Inaccurate - "Good job. Let's try some more"

Appendix D

DMMA-EP: Practice Items and Stimuli

DMMA-EP: Practice Items & Stimuli

Practice Items

A. Star. The night sky was full of _____.
B. Soap. The shampoo made her hair very _____.

Stimuli - Form A—

1. Cook. On the table, she put the meal that she had _____.
2. Strong. The boy is strong, but the man is _____.
3. Table. The classroom had three _____.
4. Quiet. He told him the secret _____.
5. Count. She used her fingers for _____.
6. Bird. In the sky, she saw two flying _____.
7. Week. During the summer, they go to the pool _____.
8. Close. The school and hospital were close, but the library was _____.
9. Cloud. When it began to rain, the sky became partly _____.
10. Wash. Before bedtime, the dishes needed to be _____.
11. Write. She wondered what he was _____.
12. Slow. Since it has a heavy shell, a snail moves very _____.
13. Sleep. After a hard day of work, he was very _____.
14. Line. This paper is _____.
15. Small. The sister is small, but her brother is _____.
16. Horse. The rancher owned five _____.
17. Fix. He took the broken toy to his father for _____.
18. Tie. Here is the knot that he _____.
19. Slow. A turtle is slow, but a snail is _____.
20. Coat. For the winter trip, the family remembered to bring their _____.
21. Salt. For her, the food was too _____.
22. Like. Since that is his favorite, his choice of it is very _____.
23. Hunt. The deer tracks were followed by the _____.
24. Wish. The genie gave him three _____.
25. Jump. Mom said, "Your bed is not for _____."
26. Fluff. After a bath, the cat was very _____.
27. Name. The teacher learned all the students' _____.
28. Lead. Today, she is the class' line _____.
29. Cream. The milkshake was very _____.
30. Rope. While waiting, he held the horse's _____.

Stimuli - Form B -

31. Dress. She owned many _____.
32. Grass. He wanted a large yard that was very ___.
33. Even. They wanted the candy to be handed out _____.
34. Boy. The team added some new ____.

35. Read. Now that he is in first grade, he is a good _____.
36. Cook. He wanted to know what she was _____.
37. Soft. His mother told him to pet the rabbit very _____.
38. Animal. We went to the zoo to see all the _____.
39. Fast. A car is fast, but a jet is _____.
40. Play. The fans cheered loudly during the game the team _____.
41. Clean. To get the house ready for the party, he is _____.
42. Reach. This is the cookie jar that he ______.
43. Home. The city had many _____.
44. Speed. He liked doing everything fast and very _____.
45. Keep. You can learn about the animals from the zoo ___.
46. Arm. The man had two strong _____.
47. Fair. Everyone was happy because the judge decided the case _____.
48. Long. The bottom string is long, but the top string is _____.
49. Ball. He was good at juggling three _____.
50. Part. Since she did not have all the pieces, the picture could only be put together ____.
51. String. Her dirty hair looked very _____.
52. Walk. She was tired after all the steps they had _____.
53. Size. The brothers were all different _____.
54. Work. She did not want to be interrupted while she was _____.
55. Sail. Here is the boat that he _____.
56. Inch. The long ruler measured many _____.
57. Sand. After a day at the beach, the car was very _____.
58. Play. This is the game they are _____.
59. Leg. The bear stood up on his back two _____.
60. Win. The prize was given to the _____.

CURRICULUM VITAE

FRANCES E. GIBSON, CCC-SLP

EDUCATION

Doctor of Philosophy, Specialization in Speech-Language Pathology & Language and Literacy, PhD (to be awarded—May 2017)

THE UTAH STATE UNIVERSITY—Logan, UT

Co-Chairs: Julie A. Wolter, Ph.D. CCC-SLP and Dr. Tim Slocum

Master of Arts in Speech-Language Pathology, MA (May 2002)

THE GEORGE WASHINGTON UNIVERSITY—Washington, DC

Bachelor of Science in Speech-Language Pathology and Audiology with a minor in Rehabilitation Services, BS (May 1999)

THE FLORIDA STATE UNIVERSITY—Tallahassee, FL

UNIVERSITY EXPERIENCE

2013 to Present

Doctoral Student, Speech-Language Pathology

As part of the College of Education and Human Services—Department of Special Education and Rehabilitation, this program focuses on both content-area and interdisciplinary coursework and training

Clinical Supervisor

Supervision focuses on developing the clinical skills of graduate students who provide therapy to preschoolers with communication disorders, including both language, phonological processing, and fluency disorders

Research Assistant

Assisted with research focused on the multi-linguistic contributions to early school-aged literacy development as part of Dr. Julie Wolter's School-Age Language and Literacy Lab (SALL)

Instructor

Fall 2014—COMD 2500: Language, Speech, and Hearing Development

Instruction focused on an introduction to the key aspects of speech, language, and hearing development

Teaching Assistant

Spring 2014—COMD 5200: Language Assessment and Intervention for Children Birth to 5

Fall 2013—COMD 6020: Language Assessment and Intervention for School-Aged Children &

CLINICAL & TEACHING EXPERIENCE

Certified Speech-Language Pathologist 2008 to 2013
THE HOWARD SCHOOL - Atlanta, GA
Caseload focused on a combination of ninth, tenth, eleventh and twelve graders with a combination of language and learning needs served through a collaborative service delivery model

Certified Speech-Language Pathologist 2006 to 2008
FORSYTH COUNTY PUBLIC SCHOOLS - Cumming, GA
Caseload focused on a combination of second, third, and fourth graders

Certified Speech-Language Pathologist 2005 to 2006
ATLANTA PEDIATRIC THERAPY - Atlanta, GA
Caseload focused on a combination of early intervention assessments for *Babies Can't Wait* and early elementary students, including preschoolers, kindergarteners, first, second, and third graders

Certified Speech-Language Pathologist 2003 to 2005
DOUGLAS COUNTY PUBLIC SCHOOLS - Douglasville, GA
Caseload focused on the community-based preschoolers

Clinical Fellow, Speech-Language Pathology 2002 to 2003
COBB COUNTY PUBLIC SCHOOLS - Marietta, GA
Caseload focused on a variety of students from kindergarten to fifth grade

RESEARCH

Doctoral: A Dynamic Assessment of Morphological Awareness in Young Children 2015 to Present
Focused on designing a dynamic assessment sensitive to emerging levels of morphological awareness in kindergarteners. Dr. Julie Wolter, Department of Communication Sciences and Disorders

PUBLICATIONS

Gibson, F.E. & Wolter, J.A. (2015). *Morphological Awareness Intervention to Improve Vocabulary Success.* Perspect Lang Learning Educ*., 22(4), 147-155.*

Wolter, J.A. & Gibson, F.E. (2015). *Morphological Awareness Assessment and Intervention to Improve Language and Literacy*. Seminars in Speech and Language, 36*(1), 31-41.*

NATIONAL & INTERNATIONAL PRESENTATIONS (Refereed)

Gibson, F. & Wolter, J. (2017). *A Dynamic Screening of Morphological Awareness in First-Grade Children: A Preliminary Validation*. A poster session presented at the 2017 American Speech Language and Hearing Association Annual Convention, Los Angeles, CA.

Gibson, F. & Wolter, J. (2017). *Dynamic Assessment of Morphological Awareness and Early Literacy Achievement*. A technical research session presented at the 24th Annual Meeting of the Society for the Scientific Study of Reading Conference, Halifax, Nova Scotia.

Gibson, F. & Wolter, J. (2016). A Dynamic Screening of Morphological Awareness in Kindergarten Children: A Preliminary Validation*.* A technical research session presented at the 2016 American Speech Language and Hearing Association Annual Convention, Philadelphia, PA.

Gibson, F. & Wolter, J. (2016). *Early Spelling Development and Children's Sensitivity to Morphological Patterns*. A technical research session presented at the 23rd Annual Meeting of the Society for the Scientific Study of Reading Conference, Porto, Portugal.

Gibson, F. & Wolter, J. (2015). *A Dynamic Measure of Morphological Awareness in Young Children: A Feasibility Study.* A technical research session presented at the 2015 American Speech Language and Hearing Association Annual Convention, Denver, CO.

Wolter, J., Gibson, F., & Edrington, C (2014). *Multiple Linguistic Awareness and Literacy in Students with Language Impairment.* A poster presented at the 2014 American Speech Language and Hearing Association Annual Convention, Orlando, FL.

Walsh, J., Gibson, F., & Giles, S. (2013) *Syntactic Attainments in Adolescent Language: Exploiting the Lexical/Syntactic Interface.* A poster presented at the 2013 American Speech Language and Hearing Association Annual Convention, Chicago, IL.

OTHER PRESENTATIONS (Non-Refereed)

Gibson, F. (2008, February) *A Speech RTI Process*. Presented at the Georgia Department of Education's Elluminate Session on Response-to-Intervention (RTI).

Gibson, F. (2004, August) *Speech Referrals and the Student Support Team Process*. Presented to Douglas County Schools

Gibson, F. (2004, March) *Electronic Portfolio Assessment: Job Overview*. A Preschool Handicapped Module sponsored by the Georgia Project for Assistive Technology

Gibson, F. (2004, January) *Multicultural Issues in Communication Disorders.* Presented at the Cobb County SLP Roundtables

Gibson, F. (2002, April) *Health Law 101: A Basic Introduction for Speech-Language Pathologists.* Presented to the Providence Hospital Speech Department

HONORS, AWARDS, & RECOGNITIONS

Research Assistantship: Evidence-based Language and Literacy: U.S. Department of Education, Office of Special Education and Rehabilitation, Preparation of Leadership Personnel
Frederick Q. Lawson Fellowship: Utah State University Department of Communication Disorders and Deaf Education
ASHA Award for Continuing Education (ACE), 2013

CERTIFICATIONS & LICENSES

ASHA Certified Speech-Language Pathologist, CCC-SLP
Licensed Speech-Language Pathologist, State of Utah
Licensed Speech-Language Pathologist, State of Georgia
Certified Educator, Speech & Language Pathology (P-12), State of Georgia
With a Preschool Handicapped Endorsement Certification Add-on

ADVANCED TRAINING

Getting Started as a Successful Proposal Writer and Academician Workshop, Office of Research and Graduate Studies, Utah State University, Logan, UT, April 2012

MEMBERSHIPS IN PROFESSIONAL ASSOCIATIONS

American Speech and Hearing Association (ASHA)
American Speech and Hearing Association (ASHA), Special Interest Group #1
Society for the Scientific Studies of Reading (SSSR)
Utah Speech-Language and Hearing Association (USHA)
Georgia Speech-Language and Hearing Association (GSHA)
Georgia Organization of School-based Speech-Language Pathologists (GOSSLP)
Omicron Delta Kappa
Phi Sigma Pi National Honor Fraternity
Golden Key International Honor Society

LEADERSHIP & CIVIC INVOLVEMENT

Coordinator of Language, Communication, & Speech Services at The Howard School in Atlanta, Georgia
Awarded as an Advanced Communicator Silver (ACS) by Toastmasters International
Awarded as a Competent Leader (CL) by Toastmasters International