January 2014

Vocabulary Assessment: A Comparison Between Clinical And Academic Measures

Lindsay Mae Anderson

VOCABULARY ASSESSMENT: A COMPARISON BETWEEN CLINICAL AND ACADEMIC MEASURES

by

Lindsay Mae Anderson
Bachelor of Arts, University of North Dakota, 2012

A Thesis

Submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

Master of Science

Grand Forks, ND
May
2014

This thesis, submitted by Lindsay Mae Anderson in partial fulfillment of the requirements for the Degree of Master of Science from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

_____
Dr. Sarah Robinson, Chairperson

_____
Dr. John Madden

_____
Dr. Manish Rami

This thesis is being submitted by the appointed advisory committee as having met all of the requirements of the School of Graduate Studies at the University of North Dakota and is hereby approved.

_____
Dr. Wayne Swisher
Dean of the School of Graduate Studies

_____
Date

PERMISSION

Title	Vocabulary Assessment: A comparison between Clinical and Academic Measures
Department	Communication Sciences and Disorders
Degree	Master of Science

In presenting this thesis in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my thesis work or, in her absence, by the Chairperson of the department or the dean of he School of Graduate Studies. It is understood that any copying or publication or other use of this thesis or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my thesis.

Lindsay Mae Anderson
4/28/14

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to my thesis advisor Dr. Sarah Robinson for all her guidance and support. I would also like to thank my thesis committee members Dr. John Madden and Dr. Manish Rami for taking time to provide feedback in the writing process. A special thanks goes out to the Grand Forks Public School District for playing an instrumental role in the recruiting process, as well as a thank you to all the participants and their families for taking part in this study. Thank you to my family for their awesome support throughout my graduate school studies. I am so blessed to have such a wonderful and supportive family. Finally, a huge thank you goes out to my classmates Ellie Lindsey and Emily Stephens for their endless words of encouragement throughout the journey!

ABSTRACT

Researchers have shown vocabulary knowledge to be highly linked to reading comprehension, and ultimately a strong predictor of academic success. Therefore, students' vocabulary skills should be carefully monitored and assessed as the demands of education increase.

The purpose of this study was to compare 28 fourth grade students' vocabulary scores on the Measures of Academic Progress (MAP) test with scores on the Peabody Picture Vocabulary Test-4 (PPVT-4) (Dunn & Dunn, 2007) and Expressive Vocabulary Test-2 (EVT-2) (Williams, 2007), two clinically used vocabulary measures with established validity. Strong correlations would potentially allow for the use of the MAP vocabulary test when making clinical decisions.

Pearson r correlations between EVT-2 and MAP vocabulary scores yielded a moderate correlation, and correlations between PPVT-4 and MAP vocabulary scores yielded a fair to moderate correlation. Data does not provide strong enough support to use the MAP vocabulary as a basis for clinical decisions, but it is recommended that scores be used as part of a comprehensive speech-language assessment to gain a better understanding of overall vocabulary abilities.

CHAPTER I

INTRODUCTION

Written and spoken words are used to express thoughts, share information, and exchange ideas with others. The words that an individual uses and understands are part of their vocabulary. Vocabulary is considered to be divided into two categories: expressive vocabulary and receptive vocabulary. Expressive vocabulary involves using words to communicate through speaking or writing, whereas receptive vocabulary involves comprehending words through hearing or reading (Jalongo & Sobolak, 2010). Both expressive and receptive vocabularies begin to grow in early childhood, and continue to grow across the lifespan through life experiences and interactions with others. Words become stored in an individual's lexicon, or "mental dictionary of word knowledge" (Vadasy & Nelson, 2012, p. 6).

As children progress through the elementary school years, academic demands begin to increase, and students need adequate vocabulary knowledge to achieve expected academic success (Corson, 1997). Around fourth grade, students become challenged to understand a more extensive vocabulary and to synthesize vocabulary with background information to make meaning of text (Chall & Jacobs, 2003). Without the ability to do this, comprehension of academic materials becomes difficult and students fall behind their peers (Buly & Valencia, 2002).

Therefore, students' vocabulary skills should be carefully monitored and assessed as the demands of education increase. The available vocabulary assessment tools can be divided into two primary categories: clinical tools and academic tools. Clinical tools are standardized assessments designed to measure a particular content area. These tests typically have well established reliability and validity; however, are generally only administered by special education personnel to diagnose disabilities and determine eligibility for special education programs.

Two commonly used clinical measures of vocabulary assessment are the Peabody Picture Vocabulary Test—Fourth Edition (PPVT-4) (Dunn & Dunn, 2007) and Expressive Vocabulary Test—Second Edition (EVT-2) (Williams, 2007), measuring receptive and expressive vocabulary knowledge, respectively. These measures have strong, established validity but are only administered to students referred for special education testing.

In contrast, academic tools are created or administered by general education faculty to all students, such as classroom tests and district or statewide assessments. Under the No Child Left Behind Act (NCBL) of 2001, schools are required to conduct annual assessments in core academic areas to ensure students are receiving adequate academic education (Cortiella, 2005). Through these tests, information about children's strengths and challenges in different areas are provided and achievement growth can be compared over time.

The Measure of Academic Progress (MAP) test is a computer based, dynamic assessment measure used by many school districts across the nation to assess

students' performance in core areas (Northwest Evaluation Association, 2012a). These tests are typically taken twice a year, with progress being measured over time. Vocabulary is one sub-section of the reading portion of this test.

Results from the vocabulary portion of the MAP test could potentially allow for quick identification of students with vocabulary deficits. If the vocabulary sub-test of the MAP yields strong correlations when compared with clinically used measures of vocabulary (i.e. PPVT-4 and EVT-2), it may be possible to use MAP test scores when making clinical decisions.

The question remains whether speech-language pathologists should use MAP vocabulary scores when determining vocabulary abilities in fourth grade students. The purpose of this study is to compare vocabulary scores on the MAP test with scores on the PPVT-4 and EVT-2, two clinically used vocabulary measures that have already established validity and reliability. Strong correlations would indicate that MAP testing is a valid measure of vocabulary skills in fourth grade students. This may allow for a more comprehensive identification of children who may need further speech-language evaluations. It may also reduce the need for administration of additional vocabulary assessments.

CHAPTER II

LITERATURE REVIEW

Vocabulary is referred to as the "knowledge of words and word meanings" (Honig, Diamond, & Gutlohn 2008, p. 407). Beginning in early childhood, children are continually expanding their word knowledge through experiences and interactions with peers, educators, and other individuals in their lives. These words may be understood receptively in ways such as hearing or reading, or expressed in ways such as speaking or writing. Therefore, vocabulary is typically divided into two different categories: receptive vocabulary and expressive vocabulary.

Both expressive and receptive vocabularies are known to most rapidly grow during childhood (Read, 2000). Although children receptively understand words prior to the age of one, children don't typically begin to produce their first words until around 12 months. These first words typically contain earlier acquired sounds and a simple syllabic pattern, such as 'mama' or 'papa.' Children do not learn these words on their own, however. A child may coincidentally be babbling 'mama' without any intended meaning. Parents and caretakers are quick to reinforce these sounds and meaning becomes attached as children infer the connection (O'Grady, 2005). Parents and caretakers continue to play an influential role in their child's early vocabulary development by providing exposure to new words.

Vocabulary Development

It has been hypothesized that these early stages of word learning are a product of what is known as fast mapping. Dollaghan (1987) describes fast mapping as "the initial step in lexical acquisition, in which a listener rapidly constructs a representation for an unfamiliar word on the basis of a single exposure to it" (Dollaghan, 1987, p. 218). In an earlier study, Dollaghan (1985) examined this hypothesis by exposing 35 preschool children to a nonsense word and a nonsense object referent. After just one exposure to the nonsense word and its referent, 81% of children were able to select the correct object referent when hearing the matching nonsense word a second time (Dollaghan, 1985). These results would suggest that children are 'fast mapping' the words they hear.

It would be anticipated that for these children who are 'fast mapping,' the more exposure to words the greater the vocabulary size would become. Several studies have shown that greater exposure to vocabulary does lead to a larger vocabulary size. A study completed by Huttenlocher, Haight, Bryk, Seltzer, and Lyons (1991) researched the relationship between vocabulary exposure through parents' speech and it's effect on vocabulary size for children ages 14 to 26 months, which has been shown to be a time of rapid vocabulary growth. Although individual differences in capacity to learn are assumed to play a role in overall language development, results found that frequency of exposure to words does have an impact on vocabulary size during this time period (Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991).

Following a child's production of first words, vocabulary size increases slowly until approximately 50 words have been learned. This would be expected around the age of 18 months. Once this level is achieved, a vocabulary spurt occurs and vocabulary size rapidly increases and by 30 months a child will likely have an expressive vocabulary of over 500 words (Li, Xiaowei, & MacWhinney, 2007).

After the occurrence of the 'vocabulary spurt,' vocabulary development continues; however, it occurs at a slower rate. Receptively, a child will typically understand more than 300 words between nineteen and twenty four months, more than 10,000 words between four and five years, and around 20,000 words between six and seven years of age (Shipley, 2009). This vocabulary growth continues to expand over the school age years. Nagy and Herman suggest that students learn approximately 2,700 to 3,000 new words yearly (Nagy & Herman, 1987). These words are stored in what is known as an individual's lexicon, or "mental dictionary of word knowledge" (Vadasy & Nelson, 2012, p. 6).

Although students are expected to expand their vocabulary by 2,700 to 3,000 words yearly, not all students follow this trend. Some children fall behind their typically developing peers, and are challenged to keep up with their classroom curriculum. As children expand their vocabulary and word knowledge, they begin to read more, and therefore expose themselves to a greater number of words. Through exposure to unfamiliar words, these words can be learned incidentally and hence applied to better comprehend meaning of text (Anderson, Reynolds, Schallert, & Goetz, 1977). This process is known as the "Matthew Effect." Therefore, children

who have a limited vocabulary size when first beginning school will lag behind their peers' level of performance. A gap is formed, which continues to grow larger throughout the school years as academic demands increase (Biemiller & Slonim, 2001).

An understanding of a more extensive vocabulary is required as these academic demands increase, and as students begin learning vocabulary from specific subject areas, such as science and history. At this time students are required to synthesize vocabulary with background information to make meaning of the text. Without sufficient vocabulary knowledge students begin to struggle with comprehending the text and therefore continue to fall further and further behind their peers (Buly & Valencia, 2002). This lack of knowledge can continue to inhibit an individual's ability to comprehend materials and succeed as upper-grade level curriculum continues to become more challenging. Research has shown there to be specific words that aid in the comprehension of grade level materials, therefore, vocabulary has been divided into what is known as vocabulary tiers.

Vocabulary Tiers

Vocabulary can be classified into three different tiers based on a word's complexity and frequency of use. Beck and McKeown developed these vocabulary tiers as a way to help guide teachers in vocabulary instruction (Vadasy & Nelson, 2012). The researchers suggest that through this classification system, instructors are able to choose words that will most efficiently and effectively promote the success of students.

The first of these tiers is known as tier one. This tier consists of common, everyday words that are generally familiar to most English speakers, such as the words 'dog,' 'talk,' or 'swim' (Beck, McKeown, & Kucan, 2013). Both children and adults frequently use these basic words. They are typically acquired through hearing adults, peers, or educators use the words in everyday oral speech. These words rarely require specific vocabulary instruction to be learned (Blachowicz, Ogle, Fisher, & Taffe, 2013).

The second tier consists of words that are frequently occurring in the English language, but are not as commonly used. These words are high utility words, as they are typically used across a variety of subjects/topics. Words in this tier include those such as 'predict,' 'circumstance,' or 'contradict' (Beck et al., 2013). These words may be used in a variety of different contexts, and are generally more abstract, increasing their difficulty to be understood. Since these words tend to be more challenging for students, and are widely used across a variety of subjects, it is suggested that vocabulary instruction in the classroom focuses on words from this tier (Blachowicz et al., 2013).

The third tier consists of words that are specialized to specific subjects/topic areas. These words are typically only learned in specific content areas, such as in a biology or history class. Examples of tier three words include 'epidermis,' 'mitosis,' or 'peninsula' (Beck et al., 2013). Unlike tier two words, these words are not used across a variety of subjects. Therefore, they are not typically taught unless a specific need arises (Blachowicz et al., 2013).

Of the three tiers, children who have limited knowledge of tier two words will likely demonstrate greatest difficulty keeping up with grade level coursework. As students move through the challenges of academics, having an understanding of these words becomes increasingly more fundamental as they frequently occur in school texts (Vadasy & Nelson, 2012).

Vocabulary and Reading Comprehension

Without knowledge of frequently occurring tier two words, a reader's comprehension of materials can significantly decrease, making learning difficult, and ultimately affecting academic success. Researchers have shown this relationship between vocabulary knowledge and comprehension to be strong (Beck, Perfetti, & McKeown, 1982), with positive correlations typically ranging between .6 and .8 (Pearson, Hiebert, & Kamil, 2007). A study completed by Yildrim, Yildiz, and Ates (2011) closely examined the relationship between vocabulary and reading comprehension. Participants included 120 fifth grade students who attended a public school and came from middle socioeconomic statuses. Both expository and narrative texts were chosen as the reading comprehension materials. These texts were judged by classroom teachers with high inter-rater agreement to be appropriate for the fifth grade level. A total of 28 comprehension items and 45 vocabulary items were included in the tests. A correlational analysis yielded significant correlations between vocabulary and both measures of reading comprehension. The relationship between vocabulary and expository text, however, was stronger than that of narrative text (Yildrim, Yildiz, & Ates, 2011). This would

be expected, as the purpose of a narrative text is to tell a story, and entertain the reader (Weaver & Kintsch, 1991). In this case an individual would likely be able to infer information about the story based on context, even if an unknown vocabulary word was stumbled upon. The purpose of expository text, however, is to convey information and teach the reader about a specific subject/topic (Weaver & Kintsch, 1991). Therefore, children's textbooks would contain expository text, requiring a greater understanding of tier two and tier three words. Vocabulary deficits and difficulties with tier two and tier three words may also be present in children with language impairments.

Vocabulary and Language Impairment

Not only are vocabulary deficits predictors of reading difficulties, but can also be one indicator of language impairment. Children with specific language impairment (SLI) typically develop vocabulary at a slower rate than that of typically developing peers. By age two a normally developing peer would be expected to have more than a 200-word vocabulary, whereas a child with SLI will typically have around a 20 word vocabulary (Paul, 2007).

Dollaghan (1987) compared fast mapping abilities of children with language impairment and typically developing children. As previously mentioned, fast mapping is a way in which children rapidly learn words based on a single exposure to it. In this study Dollaghan found that both typically developing children and those with language impairment were able to infer connections between novel words and it's referent based on a single exposure; however, children with language

impairment demonstrated a much greater difficulty producing the novel word. After hearing the novel word twice, 64% of normal language children were able to reproduce all three phonemes in the novel word in the correct sequence, whereas only 9% of language impaired children were able to produce the word with all three phonemes in the correct sequence (Dollaghan, 1987).

This research is supported by findings of Gray, Plante, Vance, and Henrichsen (1999) who found that children with SLI typically score lower than children with normal language development on standardized vocabulary tests, including the Picture Vocabulary Test-III (PPVT-III) (Dunn & Dunn, 1997), Expressive Vocabulary Test (EVT) (Williams, 1997), Receptive One-Word Picture Vocabulary Test (ROWPVT) (Gardner, 1985), and the Expressive One-Word Picture Vocabulary Test-Revised (EOWPVT-R) (Gardner, 1990). Overall, the mean score for children with SLI was lower on all tests; however there was some overlap in scoring (Gray, Plante, Vance, & Henrichsen, 1999).

Vocabulary Assessment

Given the role of vocabulary in language deficits and academic success, educators need to evaluate vocabulary knowledge using valid and reliable assessment measures. Common core state standards have recently placed emphasis on students' vocabulary acquisition and instruction in the classroom (Blachowicz et al., 2013). This has created the need for vocabulary assessment in order to measure and report growth. Valid and reliable assessment measures would provide educators guidance as to where to place their focus of vocabulary instruction.

Assessment tools can be divided into two primary categories: clinical tools and academic tools.

Clinical Assessment Measures

Clinical tools are standardized assessments designed to measure a particular content area. These tests are typically administered by special education personnel to diagnose disabilities and determine eligibility for special education programs. Appropriately selecting a test for a diagnostic evaluation can provide valuable information and be critical to the future course of services (Plante & Vance, 1994).

Clinical tools typically have well established reliability and validity; however, they are not generally administered to all students in the regular education classroom. These tests can be time consuming and administering clinical assessment measures one on one to all students would not be a feasible task for educators. In addition to its limited feasibility in the regular education classroom, clinical vocabulary tests only assess knowledge of a limited number of words without focus on students' knowledge of academic vocabulary (Kearns & Biemiller, 2010). This loses effectiveness for regular educators, providing little information regarding course of vocabulary instruction to follow.

Two commonly administered clinical measures of vocabulary assessment are the Peabody Picture Vocabulary Test—Fourth Edition (PPVT-4) (Dunn & Dunn, 2007) and the Expressive Vocabulary Test—Second Edition (EVT-2) (Williams, 2007). These measures have strong, established validity, but as clinical tools are only administered to students referred for special education testing.

The PPVT-4 is a standardized test normed for children and adults ages 2;6 through 90 years and above. This test measures receptive vocabulary abilities and takes approximately 10-20 minutes to complete. An easel is used to present picture stimuli to the examinee. The stimulus book contains 228 test items and the entry point is determined from the examinee's age. The examinee is directed to point to a picture associated with a word presented verbally by the examiner, such as "Show me the pencil." In the event that the examinee cannot point, he/she can say the number associated with the target picture. Four choices are given for each word. The test contains two parallel forms: Form A and Form B. Throughout the administration, the examiner records responses on the test protocol by circling the answer given. Total errors are summed and converted to a standard score, which is compared against developmental norms. (Dunn & Dunn, 2007).

The EVT-2 is a standardized, expressive vocabulary test normed for children and adults ages 2;6 through 90 years and above. Similar to the PPVT-4, the EVT-2 takes approximately 10-20 minutes to complete and an easel is used to present picture stimuli. Unlike the PPVT-4, the examinee is required to verbally answer a question corresponding to each picture presented, such as "What is this animal?" when shown a picture of an octopus. A total of 190 test items are included in the stimulus book and the entry point depends on the examinee's age. Test items become increasingly more difficult throughout the administration. The test contains two parallel forms: Form A and Form B. The researcher records responses as

correct or incorrect, and standard scores are calculated and compared against developmental norms (Williams, 2007).

The PPVT-4 and EVT-2 were both standardized by age and grade on a total of 3,540 age normed cases, and 2,003 grade normed cases. According to the PPVT-4 and EVT-2 test manuals, an equal number of males and females were included. In the study, 60% of the individuals identified as Caucasian, 20% Hispanic, 17% American Indian, and 3% other. Socioeconomic status was based on education. Approximately 60% of the individuals completed one to four years of college. High school graduates and GED recipients accounted for 26% of the sample and the remaining 15% completed grade 11 or below. Individuals from the southern United States accounted for about half the sample and the rest of the individuals came from the west, north central and northeast regions of the United States (Dunn & Dunn, 2007; Williams, 2007).

Reliability was found to be high on both PPVT-4 and EVT-2 measures. Internal consistency, alternate form, and test-retest reliability were established on forms A and B of the PPVT-4 for all ages and grades, with M=.94-.95, M=.89, and M=.93, respectively (Dunn & Dunn, 2007). Internal consistency, alternate form, and test-retest reliability were also established for all ages and grades on forms A and B of the EVT-2, with M=.93-.94, M=.87, and M=.95, respectively (Williams, 2007).

The content validity of the PPVT-4 was determined through the selection of the vocabulary words from 20 content areas. The vocabulary words were derived from *Merriam-Webster's Collegiate Dictionary* and *Webster's New Collegiate*

Dictionary (Dunn & Dunn, 2007). EVT-2 item selection was based on the frequency and common usage of vocabulary (Williams, 2007).

Concurrent validity was established through correlations with a variety of frequently administered standardized assessment measures, including the Comprehensive Assessment of Spoken Language (CASL) (Carrow-Woolfolk, 1999), Clinical Evaluation of Language Fundamentals—4th Edition (CELF-4) (Semel, Wiig, & Secord, 2003), EVT-2, and Group Reading Assessment and Diagnostic Evaluation (GRADE) (Williams, 2001). When the PPVT-4 was correlated with multiple subtests of the CASL, correlation coefficients for children ages 8-12 years ranged from r=.63 to r=.79. Similar correlations were found with multiple subtests of the CELF-4, with correlations ranging from r=.68 to r=.75 for 9-12 year olds. When correlated with the GRADE, a correlation coefficient for fourth grade students resulted in r=.66. A larger correlation was found with the EVT-2, resulting in an average of r=.82 (Dunn & Dunn, 2007).

Similarly to the PPVT-4, the EVT-2 has established concurrent validity through correlations with the same assessment measures as previously stated. When correlated with multiple subtests of the CASL, correlation coefficients for ages 8-12 years ranged from r=.50 to .84. Less variance was found when correlated with multiple subtests of the CELF-4, with correlations ranging from r=.69 to r=.77 for 9-12 year olds. GRADE correlations resulted in r=.72 for fourth grade students (Williams, 2007).

Gray et al. (1999) continued with this research through a study investigating the diagnostic accuracy of four vocabulary tests administered to preschool-age children, including the Peabody Picture Vocabulary Test-III, Expressive Vocabulary Test, Receptive One-Word Picture Vocabulary Test, and the Expressive One-Word Picture Vocabulary Test-Revised. Participants of this study included 31 four and five year olds with a diagnosis of specific language impairment (SLI), and 31 four and five year olds with normal language (NL). All four vocabulary tests were administered to each of the children during a single session. Results found that the mean score for children in the normal language group was higher on all tests than the mean score for the SLI group of children. On the PPVT-III and EVT assessment measures, the normal language group achieved mean scores of 112 and 104, respectively. Children in the SLI group, however, achieved mean scores of 97 and 92, respectively, on these same measures. Overall, there was a statistically significant difference between the two group's scores; however, some overlap in scoring was present. These four tests were also correlated with the Structured Photographic Expressive Language Test-II (SPELT-II) (Werner & Krescheck, 1983) in order to provide further evidence for concurrent validity. The SPELT-II was administered to all participants one week prior to the vocabulary assessment measures. Moderate to high correlations were found between these testing measures, lending support to validity. It was found, however, that a considerate amount of variability did exist in the tests' ability to correctly identify children as having SLI verses normal language.

These researchers therefore do not suggest the use of these vocabulary assessment measures as an identification tool for SLI (Gray et al., 1999).

Although the PPVT-4 and EVT-2 were standardized on a variety of races and ethnicities, the majority of the norming population was Caucasian, as were the majority of participants in the former study completed by Gray et al., (1999). Restrepo et al. (2006) explains the possible variance in performance results between African Americans and European Americans on language and vocabulary assessments conceivably due to cultural differences. This study examined the validity of the PPVT-III and EVT when assessing African American and European American children. Participants included 210 four-year-old children whom attended a lottery-funded preschool in Northeast Georgia, with 57.5% of participants being African American and 42.4% being European Americans. All students were administered both assessment measures during the first 45 days of school. Results found that African American children did perform poorer on both assessment measures when compared to European American children. The differences between scores, however, were greater on the PPVT-III than the EVT. These results suggest that the PPVT-III would be more likely to incorrectly identify African American children as demonstrating a language disorder (Restrepo et al., 2006).

Academic Assessment Measures

Unlike clinical assessment measures, academic assessments are created or administered by general education faculty to all students, such as classroom tests and district or statewide assessments. Under the No Child Left Behind Act (NCBL) of

2001, annual assessments in core academic areas are required by schools to ensure that students are receiving adequate academic education. Under this act, schools are required to meet a certain level of proficiency, reaching adequate yearly progress (AYP) in core subject areas (Cortiella, 2005). If schools fail to reach AYP for two or more consecutive years, they will be subject to sanctions and improvements will need to be made (Kim & Sunderman, 2005). Schools are required under the NCLB act to assess academic areas including reading/language arts, science, and mathematics. Questions are chosen based on each state's academic standards (Cortiella, 2005). These statewide assessments also allow educators to compare students' performance over time, and gain a better understanding as to where progress needs to be made.

Two formats of academic assessment include paper-based and computer-based assessment. Paper-based assessment is a traditional paper and pencil testing approach, but with the growth of technology in today's society, the use of computer-based testing as a means of academic assessment has become increasingly common. With its ability to quickly gather data in a timely and costly manner, computerized-based testing has been advocated as a more efficient means of assessment (Olsen, 2005). Research has been completed investigating differences in outcomes of these two testing formats, and mixed results have been reported. Srivastava and Gray (2012) completed a study comparing reading comprehension scores on computer-based verses paper-based tests. This study consisted of a total of 46 students, 25 with typical language development and 14 with language learning disabilities. All

students were in the eighth grade. Four reading passages of appropriate length with appropriate vocabulary for eighth grade level students were chosen. Two forms of the test were created in both computer-based and paper-based formats, and comprehension questions were developed to complement the reading passages. Students completed one form of the test on paper and the other form on the computer. Results found no significant affects when comparing computer verses paper-based outcomes. Both typically developing adolescents and those with language learning disabilities received similar scores across testing measures (Srivastava & Gray, 2012).

Clariana and Wallace (2002) completed a similar study focusing on computer verses paper based assessment with older students. Participants included 105 freshman business undergraduates completing a Computer Fundamentals course. These students had all received course instruction prior to testing. Students were randomly assigned to either a computer-based or paper-based test. Both tests consisted of 100 multiple-choice questions, with four answer options provided. Students completing the computer-based test clicked on the correct response to answer the question, and had the option of going back to their questions to change answers if desired. Students completing the paper-based assessment read each question on paper and wrote the letter of the answer that was felt to be correct on a separate answer sheet. In order to compare assessment measures, a one-factor between-subjects ANOVA was calculated, with results finding differences in test score means to be statistically significant. The computer-based test group

demonstrated an overall mean score of 6.8 points higher than that of the paper-based test group (Clariana & Wallace, 2002).

It is possible that a students' attitude toward computer verses paper based assessment could also influence score results. A student who has a positive attitude will likely be more engaged and put forth greater effort in the testing process, therefore receiving a better score. According to a study completed by Lim, Ong, Wilder-Smith, and Seet (2006), the majority of students do tend to prefer computer-based testing opposed to paper-based testing. A total of 114 undergraduate students completed a survey indicating preference toward testing format for two examinations previously taken as a part of medical school assessments. For the first exam, 79.8% of students indicated they preferred computer-based testing, 9.6% preferred pen and paper, and 10.5% were unsure as to what their preference was for this examination. The majority of students preferred computer-based testing for the second examination as well, with 54.4% preferring computer-based testing, 26.3% preferring pen and paper, and 19.3% were unsure (Lim, Ong, Wilder-Smith, & Seet, 2006). Overall, computer-based assessment was perceived to be a positive form of testing for the majority of students.

One commonly used measure of academic assessment is the Measure of Academic Progress (MAP). The MAP is a computer based, dynamic assessment created by the Northwest Evaluation Association (NWEA). This test is typically taken twice a year by many school districts across the nation to assess students' performance in core areas, and compare students' achievements in academic areas

to expected grade level outcomes. The MAP is specifically designed to align with the state and national standards. Each year NWEA researchers and test developers study the most recent standards in order to ensure that the test is a reflection of the most current requirements (Northwest Evaluation Association, 2014a).

The MAP test consists of three sections, including reading, mathematics, and language. Vocabulary is one sub-section of the reading portion of this test. The Northwest Evaluation Association (2012b) describes this sub-section in their RIT Reference Chart for Reading as requiring the student to "decode words, recognize common words, understand word relationships and structures, and use context cues to decipher word meaning" (Northwest Evaluation Association, 2012b, "Word Recognition, Structure, and Vocabulary," para. 1).

Classroom teachers administer the MAP test in a quiet environment. Students are assigned to seats in a school computer lab, where they are monitored and given an unlimited amount of time to complete testing. The MAP test adapts specifically after each child's response to a question, providing a question of greater difficulty if the student responds correctly, or an easier question if the student responds incorrectly (Northwest Evaluation Association, 2014b). This allows the test to focus on the individual and enhance each child's engagement in the assessment process through alleviation of the boredom seen by high achieving students, and alleviation of the frustrations associated with lower achieving children who may struggle with questions typically easy for a child of his or her age (Olsen, 2005). The MAP uses the RIT (Rasch Unit) scale as it's scoring measure. This scale divides units into equal

intervals, and these units do not correlate to grade level. In contrast, this testing format allows for the ability to measure progress from year to year (Northwest Evaluation Association, 2014b). Instructors gain the ability to see where each student is struggling and utilize this information in the classroom to provide more effective learning strategies.

Nearly all students participate in MAP testing, however validity and reliability of this test is not well established. The majority of research has ben completed by the Kingsbury Center, a research division of the parent company, Northwest Evaluation Association. The Kingsbury Center regularly conducts linking studies that compare students' performance on the MAP to state standardized tests. This allows for the ability to predict how students will perform on the state standardized test based on their MAP RIT score. When changes are made to state standardized tests, new linking studies are completed (Northwest Evaluation Association, 2014a).

Linking studies are completed in each state due to the possibility of variance in state standardized tests. In the state of North Dakota, a linking study was completed in July of 2010. This study linked 11th graders RIT scores on the MAP to North Dakota State Assessments. The sample included 549 eleventh grade students from 16 different North Dakota schools. All students completed both the MAP test and state standardized test in the fall of 2007. Data showed that the MAP scores for eleventh grade students predicted their pass/fail status on the reading portion of the North Dakota State-Wide Assessment with 77.33% accuracy. There is no data to

illustrate validity at different age/grade levels or on specific subtests. A Pearson r correlation coefficient was computed to be r=.753 between these testing measures (The Kingsbury Center, 2010).

Similar findings were found in linking studies completed in a number of other sates. In the state of Kentucky, scores on both testing measures for 11, 577 students were used as a basis of the sample. This study looked at scores for students in grades three through eight. On the reading portion of this test, fourth grade students' scores predicted their pass/fail status on the reading portion of the Kentucky's Performance Rating for Educational Progress (K-PREP), Kentucky's state assessment measure, with 78.8% accuracy. Pearson r's correlation for these measures was computed to be r=.717 (The Kingsbury Center, 2012).

Merino and Beckman (2010) completed a study in the opposite form of the Kingsbury Center linking studies. In contrast to using MAP scores to predict pass/fail performance on statewide assessment, this study examined the ability of curriculum-based measures to predict performance on the MAP reading. Participants in this study included 376 elementary students in second through fifth grade. All students attended a public school in Nebraska. Curriculum based measures used in this study included the AIMSweb Oral Reading Fluency and Maze tests. These tests are used by many school districts to assess areas including reading accuracy and speed, reading comprehension, and vocabulary. All students completed these curriculum based assessments and the MAP test in the spring and fall of 2009. A multiple regression analysis was completed with results indicating

that the Oral Reading Fluency test alone, or in combination with the Maze test, is a significant predictor of the MAP reading composite score (Merino & Beckman, 2010).

In summary, in order to participate in grade level curriculum, students need adequate vocabulary knowledge to achieve expected academic success (Corson, 1997). As academic demands increase, fourth grade students are challenged to understand a more extensive vocabulary and to synthesize vocabulary with background information to make meaning of the text. Therefore, students' vocabulary skills should be carefully monitored and assessed as the demands of education increase. The PPVT-4 and EVT-2 are two clinical assessments with established reliability and validity; however, these tests are only administered to students referred for special education testing. The MAP is an academic assessment administered to all students in many school districts around the nation; therefore having potential to allow for quick identification of students with vocabulary deficits.

Given the importance of vocabulary and vocabulary assessment in academic success, this study was designed to investigate the potential usefulness of MAP vocabulary scores. Specifically, the following questions are addressed:

1. What is the correlation between MAP vocabulary test scores and PPVT-4 scores?
2. What is the correlation between MAP vocabulary test scores and EVT-2 scores?

CHAPTER III

METHOD

Materials

Three tests were individually administered to all participants. They included the Peabody Picture Vocabulary Test-Fourth Edition (PPVT-4) (Dunn & Dunn, 2007), Expressive Vocabulary Test-Second Edition (EVT-2) (Williams, 2007), and vocabulary subtest of the Measures of Academic Progress (MAP). All tests were administered and scored in accordance with instructions in the test manuals.

The PPVT-4 was administered to measure receptive vocabulary abilities. This test took approximately 10-20 minutes to complete. Colored and enlarged picture stimuli were presented to the participants through the use of a stimulus book, containing 228 test items. The entry point was determined based on the participant's age. For each item the researcher said a word to the participant, and he or she was then told to name the number, or point to, the picture that best represented the meaning of the word stated. A field of four choices was given for each word. Scoring was completed throughout the administration by circling given responses on the test protocol. The test was discontinued after a specified ceiling was reached. The total number of errors were summed and converted to standard scores according to the test manual procedures.

The EVT-2 was administered to measure expressive vocabulary and word retrieval. This test took approximately 10-20 minutes to complete. Enlarged and colored stimuli were presented to participants through the use of the stimulus book, containing 190 test items, with the entry point based on the participant's age. Test items were arranged in increasing levels of difficulty. The participants were required to verbally answer a question (e.g. "What is this?") corresponding to each picture presented. Answers were scored as a 1 for correct and 0 for incorrect. Administration of items was continued until a ceiling was reached. The total number of errors were summed and converted to standard scores according to the test manual procedures.

The MAP is a computer-based assessment that compares student achievement to expected grade level outcomes and measures progress over time. Students in Grand Forks, ND public school district took the math and reading portions of the MAP in May and September of the 2013 school year. The reading portion included reading tests that cover phonological awareness, phonics, concepts of print and vocabulary, word structure, comprehension, and writing. For the purpose of this study, the vocabulary subsection of the reading portion was examined. This section involved assessing the ability to decode and recognize common words, understand word structures and relationships, and determine word meanings through the use of context clues (Northwest Evaluation Association, 2012b).

The MAP is a dynamic assessment that adjusts to student responses throughout the test. This means that each student is presented with a unique set of questions. This adaptive method enables students to be assessed at their exact skill level. Each question is assigned a specific point value. A student's raw score is calculated by adding these points. To compare students' scores over time, and to compare students' performance to that of their peers, raw scores are converted to a standard score, or RIT score, which lie on an equal interval scale (Northwest Evaluation Association, 2012a). For the purpose of this study, RIT scores were collected.

The test was administered in a quiet environment to one classroom at a time. Students were assigned to specific seats at computers in a computer lab and are instructed to remain quiet during the testing period. Teachers read directions to all students prior to the start of the test and monitor students in the room throughout the assessment. The reading portion was estimated to take approximately 50 minutes; however students could use as much time as needed to complete the test. Once 90 percent of the class has finished testing, students were allowed to head back to their classroom (Northwest Evaluation Association, 2012a).

Participants

Twenty-eight participants (18 males, 10 females) were recruited from fourth grade classrooms in the Grand Forks, North Dakota Public School district. Participants ranged from 9;7 to 10;8 years of age, with a mean of 10;3. All participants were native English speakers. Individuals were neither included nor

excluded based on socio-economic status. Because the focus of the study is on vocabulary abilities in typically developing children or those with a language-based impairment, individuals who had been diagnosed with Autism Spectrum Disorder, cognitive deficits, or hearing impairments were excluded from the study. Individuals with language disorders, however, were included. All participants received $20 cash for their participation.

Procedure

Fourth grade students were recruited through flyers distributed to classrooms and from an advertisement published in *Kids Connections*, a monthly newsletter sent to all parents/guardians of students in the Grand Forks Public district. The advertisement contained the purpose of study, methodology, compensation details, and instructions on who to contact if interested (Refer to Appendix A).

Interested parents/guardians were instructed to contact the primary investigator via phone or email. The study was explained further and the parents/guardians were given an opportunity to ask questions. This initial contact also served as a screening to determine participant eligibility (e.g. age, native language, any existing medical or educational diagnosis). A member of the research team later contacted the parents/guardians to schedule a time for participant testing. Research was conducted at one of two locations, at the participant's school, either before or after school hours or during weekend or evening hours on the University of North Dakota campus.

A research team consisting of three graduate assistants administered the research protocol to all participants. Prior to the participant's arrival on site, the researcher set up the materials necessary to carry out the assessment by arranging the tests, manuals, informed consent form, writing utensils, and a video recorder in a quiet room with minimal distractions.

At the beginning of each testing session, the researcher obtained the parent/guardian's signature on a consent form (See Appendix B) and the participant's signature on an assent form (see Appendix C). Through the assent form, the purpose of the study was explained and the participant was assured that he/she did not need to participate in the study and could cease participation at any time. The participants were encouraged to do their best and to expect that some questions would be easy and some would be difficult. As needed, the participant could take breaks. Parents/guardians were given the option to stay in the testing room, or a nearby waiting area, whatever the participant was most comfortable with. Each testing session was video recorded for the purpose of obtaining inter-rater reliability.

The testing protocol was a part of a larger research study and consisted of the Gray Oral Reading Test—5th ed. (GORT-5) (Wiederholt & Bryan, 2012), the Peabody Picture Vocabulary Test—4th ed. (PPVT-4) (Dunn & Dunn, 2007) the Expressive One-Word Vocabulary Test—2 ed. (EVT-2) (Williams & Williams, 2007), and a researcher-designed vocabulary assessment based on the student's current weekly spelling list. The participants completed MAP testing at their school, as part

of a district-wide requirement. This test was taken on an Apple Notebook computer in a classroom monitored by a school teacher. The order of test administration was counterbalanced to control for any order effects according to a pre-determined schedule. The entire testing session took about one hour to complete.

The administered tests were scored online according to the procedures in their test manuals. The data was entered into a password-protected spreadsheet, kept on the primary investigator's computer, and later transferred into the Statistical Package for the Social Sciences (SPSS) program for analysis.

Videos and test protocols were coded with a subject number to ensure participant privacy. All hand-written data sheets, test protocols, and videos (on a flash drive) were stored in a locked file cabinet and kept separate from the consent forms. All research materials will be kept for a period of three years before being destroyed according to University of North Dakota policy. The primary investigator and the members of the IRB audit team will be the only individuals with access to the filing cabinet.

Data Analysis

This investigation used descriptive statistics and a Pearson r correlation as measures of statistical analysis. Descriptive statistics including the mean, range, and standard deviation were calculated for MAP, PPVT-4, and EVT-2 scores. The Pearson r correlation coefficient was calculated to measure the degree of the association between measures. All data are presented in graphs and tables as well as in a written summary.

CHAPTER IV

RESULTS

Description of Variables

Twenty-eight fourth grade students between the ages of 9;7 and 10;8 completed this study. Participants were from seven of thirteen elementary schools in the Grand Forks Public School District. Raw scores on the EVT-2 and PPVT-4 were calculated and converted to standard scores according to the EVT-2 and PPVT-4 test manuals. Standard scores between 85-115 are considered to be within the average range. The range of scores for participants was 90-132 on the EVT-2 and 91-149 on the PPVT-4. These scores are slightly higher than the published norms. The mean score, however, fell within average for both standardized tests.

Raw scores on the vocabulary subsection of the MAP test were calculated and converted to RIT (Rasch Unit) scores, which lie on an equal interval scale. These scores are independent of grade level, therefore allowing for the ability to track progress made from year to year. According to the 2011 NWEA RIT Scale report, the mean RIT reading score for students beginning their fourth grade year is 199.8 (Northwest Evaluation Association, 2011). Normative data specific to the vocabulary subsection, however, is not available. The range of scores obtained for participants on the MAP vocabulary was 192-261 with a mean of 215.32.

The range, mean, and standard deviation of scores on all assessment measures were calculated and presented in Table 1. Descriptive statistics were generated through the use of the Statistical Package for the Social Sciences (SPSS-Version 21) program.

Table 1. Range, Mean, and Standard Deviation of Scores Obtained on All Assessment Measures (N=28)

	Range	Mean	Std. Deviation
EVT-2	90-132	111.54	12.03
PPVT-4	91-149	115.18	14.85
MAP Vocab	192-261	215.32	13.98

Correlational Analyses

To determine the relationship between the MAP vocabulary scores and the EVT-2/PPVT-4 scores, Pearson r correlation coefficients were computed. The Pearson r correlation is used to measure the strength and direction of a relationship between any two variables. The relationship is assigned a value ranging from -1.0 to +1.0. A value of 0 would indicate no relationship between variables, but the closer the value is to reaching -1.0 or +1.0, the stronger the relationship (Taylor, 1990).

Two correlational analyses were performed. Pearson r correlations were generated through the use of the Statistical Package for the Social Sciences (SPSS Version 21.) The first analysis compared participants' performance on the EVT-2 to MAP vocabulary scores. This analysis yielded a moderate correlation of r=.45, with r^2=.2025 (Gray, 2004).

The r is an effect size measure. This value can be expressed in a percent, which quantifies the variance explained in the data (Taylor, 1990). Therefore, 20.25% of variation in EVT-2 scores can be explained by variation in MAP vocabulary scores. This correlation was statistically significant at the p= .046 level (2-tailed). See Figure 1.

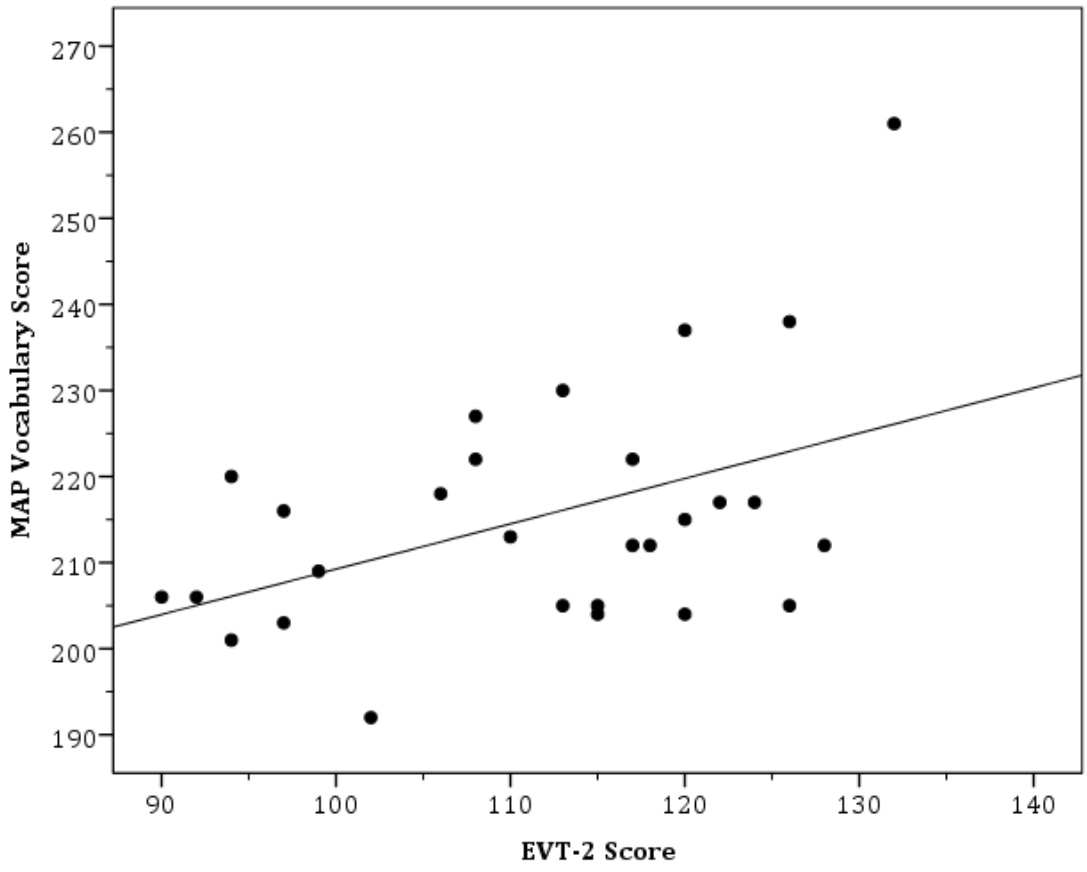

Figure 1. A scatterplot of MAP vocabulary scores as a function of EVT-2 scores.
N=28, r =.45*, r^2 =.2025, p = .046
A linear trend line is displayed.

The second analysis compared students' performance on the PPVT-4 to the MAP Vocabulary score. This analysis yielded a fair to moderate correlation of r = .38, with r^2=.1444 (Gray, 2004). This correlation is significant at the *p*=.046 level (2-tailed). See Figure 2.

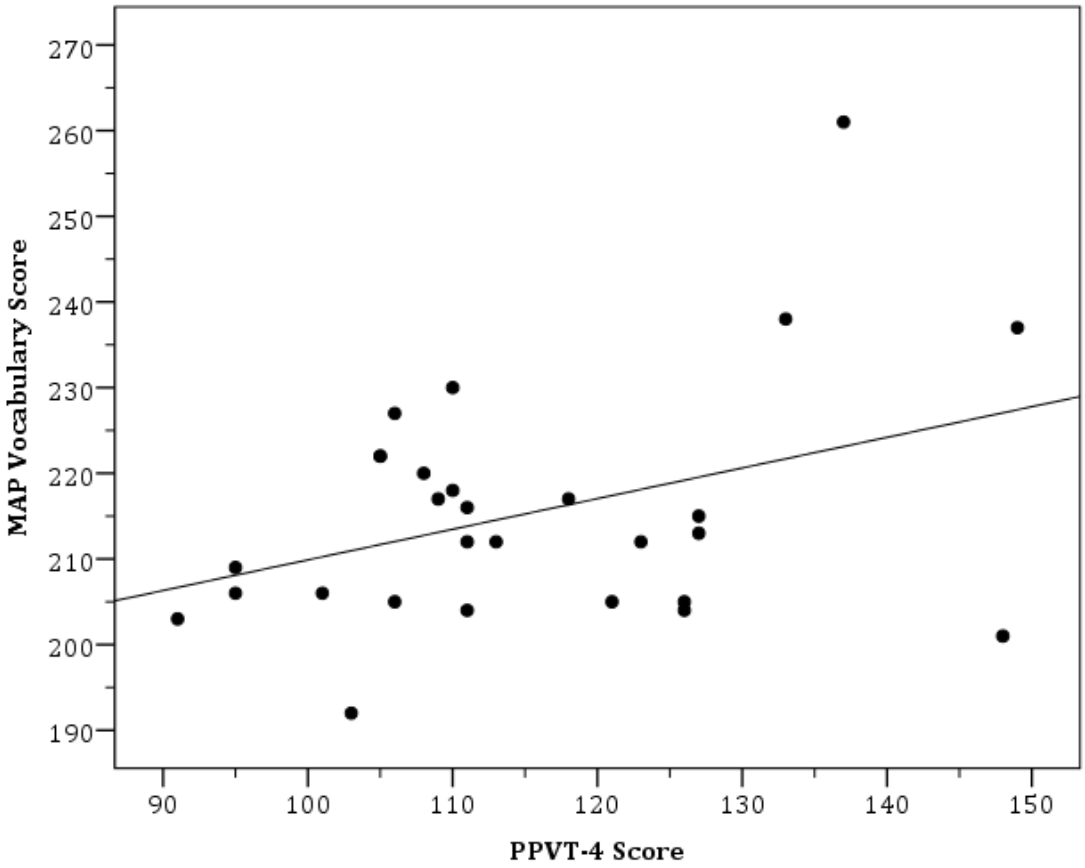

Figure 2. A scatterplot of MAP vocabulary scores as a function of PPVT-4 scores.
N=28, r =.38, r^2= .1444, *p* = .046
A linear trend line is displayed.

CHAPTER V

DISCUSSION

The present study examined the relationship between vocabulary scores obtained on the MAP test with receptive vocabulary scores obtained on the PPVT-4 and expressive vocabulary scores obtained on the EVT-2. A moderate correlation of r=.45 and r^2 =.2025 was found between the MAP vocabulary and EVT-2 scores, while a fair to moderate correlation of r=.38 and r^2= .1444 was found between MAP vocabulary and PPVT-4 scores (Gray, 2004).

The test scores were correlated, reflecting that they measured the same general construct, which is vocabulary, but not in the same manner. These correlations were lower than anticipated. It is possible that the relatively low correlations were due to differences in the format of the tests. The MAP test assesses vocabulary with a given context (i.e. narrative, recipe), giving the test taker an opportunity to deduct the meaning. An example of a MAP vocabulary question as taken from the RIT Reference Chart for Reading (2012b) is as follows:

> My friend Chris always does what he promises to do. If he says he'll meet me after school, he is always there waiting for me. Chris is a reliable friend. What does reliable mean? (Northwest Evaluation Association, 2012b, "Word Recognition, Structure, and Vocabulary," para. 5)
>
> 1. friendly
> 2. bright
> 3. dependable
> 4. capable

In this case the correct answer would be dependable (Northwest Evaluation Association, 2012b). If the test taker did not immediately know the meaning of the word reliable, the context clues may help the reader determine the correct answer.

In contrast, the PPVT-4 assesses receptive vocabulary in a decontextualized format. For example, a test item on the PPVT-4 is one such as "Show me the pencil," in which the examinee would be asked to point to the picture of a pencil from a field of four choices. In such cases, no context is provided and the examinee does not have the option of deciphering the word's meaning from context clues. Likewise, the EVT-2 also assesses expressive vocabulary in a decontextualized format by requiring the examinee answer a question when looking at a picture, such as answering the questions "What is this musical instrument?" when shown a picture of only a guitar.

With these variations in testing format, the underlying ability required to complete tasks on the MAP test verses the PPVT-4 and EVT-2 may be different. Targeting vocabulary in a contextualized format is a more naturalistic form of assessment opposed to assessing vocabulary in a decontextualized format (Paul, 2012). In our everyday life environment, vocabulary words tend to be used within a context to some degree (i.e. reading a book, interacting with friends), therefore the way the MAP assesses vocabulary more closely reflects how vocabulary is functionally used in our daily lives. Researchers Nagy, Herman, and Anderson (1985) agree that vocabulary words are frequently learned in our natural environment through reading of text. In their 1985 study of 57 eighth graders,

participants read a 1,000 word expository or narrative text, and completed vocabulary assessment measures after completion of the readings. Results found that students did increase their word knowledge following the task, suggesting that students do in fact learn vocabulary incidentally through support of context (Nagy, Herman, and Anderson, 1985). Therefore, it would be anticipated that kids would score better when provided a context.

A second difference that was noticed during testing was specific to the types of words assessed on the MAP test verses PPVT-4 and EVT-2. The MAP vocabulary sub-test has the ability to test abstract words, such as those that represent concepts or ideas (e.g. the word reliable). These words do not have a physical referent that can be referred to. In addition, this format allows for the ability to test academic terms, as these are often abstract. Due to the nature of the PPVT-4 and EVT-2 assessments (e.g. student identifies one of four pictures with the target meaning, or student expressively states the meaning of a picture), these tests can only use "picturable" words, that is, words that relate to more concrete objects and actions, rather than abstract concepts (Kearns, & Biemiller, 2010).

In the present study, interesting patterns between scoring relationships were found. It would be anticipated that that students who scored high on the PPVT-4 would also score high on the MAP vocabulary, and likewise students who scored high on the EVT-2 would also score high on the MAP vocabulary. This was not necessarily the case. Some participants did follow this trend, however a significant number did not. When examining correlations between scores obtained on the

EVT-2 and MAP vocabulary, it was found that the majority of participants who received high scores on the EVT-2 achieved average MAP vocabulary scores in comparison to the remaining participants. A somewhat opposite pattern was found between the PPVT-4 and MAP vocabulary. When examining correlations between these scores, it was found that several participants who scored high on the MAP vocabulary achieved average PPVT-4 scores in comparison to the remaining participants. These observed patterns likely have influence on the low correlations.

A third notable difference between testing measures is the text medium display. The MAP test is a computer-based assessment, whereas the PPVT-4 and EVT-2 are paper-based assessments. With the advance in today's technology, there has been a recent shift toward the use of computer-based measures as a form of annual assessment. Research has been shown to be inconsistent as to whether this different form of testing has an effect on performance. The findings of this study are consistent with Srivastava and Gray (2012) who completed research comparing reading comprehension scores on computer-based verses paper-based assessments, with results indicating no significant effects across testing measures (Srivastava & Gray, 2012).

Although differences in testing format are present, MAP scores do provide beneficial information regarding students' vocabulary abilities. The MAP seems to measure a different subset of skills, but it may still be beneficial as a part of a comprehensive speech-language assessment to consider MAP scores, along with other academic assessments to supplement standardized clinical tests in order to

provide a more in depth view of students' performance. This allows for the ability to gain a better understanding as to whether students are able to comprehend those more abstract concepts, and if they are able to utilize context clues to infer meaning. In addition, MAP scores are a better reflection of academic vocabulary skills.

Some limitations to this study should be noted. First, this study was completed with the small sample of 28 participants. In order to provide more valid and reliable results, and for the benefits or limitations of the MAP to be more clearly identified, this study should be replicated with a much larger sample size. It was also noted that published Pearson r correlation coefficients between the PPVT-4 and EVT-2 did not match correlations found in this study. According to the PPVT and EVT test manuals (Dunn & Dunn, 2007; Williams, 2007), an average correlation of r=.82 was found; however, this study found the correlation to be r=.52. This may in part be due to the nature of the small sample size. A further limitation found was that the participants of this study in included individuals who scored higher than the average sample of students. This questions the generalizability of results to students who may be at or below the average.

In conclusion, the goal of this study was to examine the relationship between MAP vocabulary scores and PPVT-4 scores, and MAP vocabulary and EVT-2 scores. If strong relationships were found, special educators could further use the MAP annual assessment test results when making clinical decisions and identifying children who may need further speech-language evaluations. Based on the analysis of this small sample, data did not provide strong enough results to support the use

of the MAP vocabulary test as a basis for making clinical decisions; however, with correlations in the fair to moderate and moderate range, MAP scores do offer some insight into students' vocabulary and should be used as a part of a comprehensive speech-language assessment to gain a better understanding of overall vocabulary abilities.

APPENDICES

APPENDIX A
RECRUITMENT LETTER

UNIVERSITY OF UND NORTH DAKOTA

DEPARTMENT OF COMMUNICATION SCIENCES AND DISORDERS
SPEECH, LANGUAGE AND HEARING CLINIC
MONTGOMERY HALL ROOM 101
290 CENTENNIAL DRIVE STOP 8040
GRAND FORKS, NORTH DAKOTA 58202-8040
(701) 777-3232
FAX (701) 777-4578

Dear Parents/Guardians,

I am a speech-language pathologist and researcher at the University of North Dakota. My research team is conducting a study comparing scores on the MAP test to other clinically administered tests. The MAP test is a computer-based test that is administered by your child's school. If your child chooses to participate, I will need your permission to access these scores. The clinical tests will include one test of reading comprehension and three tests of vocabulary knowledge, one of which entails using your child's current weekly spelling list. For your convenience, my research team can schedule testing sessions at your child's school during after school hours. Weekend and evening sessions are available at UND, as well. Each testing session should last approximately 1 hour. Your child will be compensated with $20 for their participation.

If your child is interested in participating in this study, please contact me via email: sarah.robinson@und.edu or by phone 701-777-1490.

Thank you,
Sarah Robinson, PhD, CCC-SLP

APPENDIX B
CONSENT FORM

PARENTAL CONSENT TO PARTICIPATE IN RESEARCH

TITLE:	A comparison of students' reading and vocabulary performance on MAP testing to performance on clinical measures
PROJECT DIRECTOR:	Sarah Robinson
PHONE #	777-3723
DEPARTMENT:	Communication Sciences and Disorders

STATEMENT OF RESEARCH
A person who is to participate in the research must give his or her informed consent to such participation. This consent must be based on an understanding of the nature and risks of the research. This document provides information that is important for this understanding. Research projects include only subjects who choose to take part. Please take your time in making your decision as to whether to allow your child to participate. If you have questions at any time, please ask.

WHAT IS THE PURPOSE OF THIS STUDY?
We invite your child to take part in a research study conducted by Dr. Sarah Robinson from the Department of Communication Sciences and Disorders at the University of North Dakota. The purpose of the study is to compare your child's score on sections of the MAP test (which s/he takes at school) to tests we are going to give him/her today.

HOW MANY PEOPLE WILL PARTICIPATE?
Approximately 80 fourth grade students will be selected to participate in this study. All of the students selected will need to complete the MAP testing in May (at school).

HOW LONG WILL MY CHILD BE IN THIS STUDY?
The testing session for this study will take approximately one hour. There will be only one testing session.

WHAT WILL HAPPEN DURING THIS STUDY?
There are two parts to the study.
1. The first part is the testing session. We will administer two standardized tests to evaluate your child's vocabulary abilities and one test to evaluate his/her reading abilities. This testing session will be video recorded.
2. The second part of the study is the MAP testing. The Grand Forks Public Schools administer the MAP testing to all students at school. With your permission, we will access your child's score for the May testing session.

WHAT ARE THE RISKS OF THE STUDY?
Participation in this study involves the following risks.

1. Your child may become uninterested, fatigured or frustrated during the testing session. We will offer appropriate breaks to use the restroom, get a drink of water, or walk around as needed. The tests that we are administrating are routinely used by speech-language pathologists during assessments.
2. It is possible that your child may become embarrassed if s/he does not know some of the items being tested. All participants will be assured that the items increase in difficulty and they will not know some or many of the words. They will be encouraged to guess if they are not sure or they will be told to respond "I don't know."
3. Your child may feel uncomfortable being video recorded during the testing session. Students will be assured that only the researcher and the research asssitants will have access to the video recordings. They will also be assured that we record sessions so that we can make sure that the evaluator has not made any mistakes.

WHAT ARE THE BENEFITS OF THIS STUDY?
Your child may benefit by knowing that s/he has helped in the research process. You will also have access to your child's vocabulary and reading comprehension scores. In the future, others may benefit by learning about what MAP scores tell educators.

ARE COSTS INVOLVED IN THIS STUDY?
You will not have any costs for allowing your child to participate in this research study. Upon completion of the testing session, your child will receive a $20 gift card.

WHO IS FUNDING THE STUDY?
The University of North Dakota and the research team are receiving no payments from other agencies, organizations, or companies to conduct this research study.

CONFIDENTIALITY
Confidentiality will be maintained to the extent allowed by law. We will make every effort to ensure that a loss in confidentiality does not occur. We will store all written records in a locked cabinet. We will store computer files related to your child's data under password protection. When the research program is complete, we will write up the results of the study as a research report. Your child will not be identified in any way except as a subject number. Our research records may be reviewed by Government agencies and the University of North Dakota Institutional Review Board.

IS THIS STUDY VOLUNTARY?
Your child's participation is voluntary. You or your child may choose not to participate or to discontinue participation at any time without penalty. Your decision whether or not to participate will not affect your current or future relations with the University of North Dakota.

INJURY DUE TO PARTICIPATION
If your child is injured as a direct result of being in this study, neither the University of North Dakota nor the principal investigator, Sarah Robinson, will pay for any care, lost wages, or provide other financial compensation. Please refer to the "Risks of the Study" section above for a list of possible risks of participating in the study.

CONTACTS AND QUESTIONS?
Sarah Robinson is the researcher conducting this study. You may ask any questions you have now. If you later have questions, concerns, or complaints about the research, please contact Sarah Robinson at 777-3723 during the day.

If you have questions regarding your rights as a research subject, or if you have any concerns or complaints about the research, you may contact the University of North Dakota Institutional Review Board at (701) 777-4279. Please call this number if you cannot reach research staff, or you wish to talk with someone else.

AGREEMENT
The University of North Dakota Institutional Review Board has approved this consent form as signified by the committee's stamp. This consent form must be reviewed at least once each year and expires on the date indicated on the stamp. Your signature below indicates that you have read the information in this document and have had a chance to ask any questions you have about the study. Your signature also indicates that you have decided to let your child participate, and have been told that you can change your mind and withdraw your consent for your child's participation at any time. You have been given a copy of this consent form to keep. You have been told that by signing this consent form you are not giving up any of your child's legal rights.

_____ _____ _____
NAME OF CHILD PARTICIPANT (please print) AGE DATE

_____ _____
SIGNATURE OF PARENT OR GUARDIAN DATE

_____ _____
SIGNATURE OF INVESTIGATOR DATE

APPENDIX C
ASSENT FORM

TITLE: A comparison of students' reading and vocabulary performance on MAP testing to performance on clinical measures

PROJECT DIRECTOR: Sarah Robinson

PHONE # 777-3723

DEPARTMENT: Communication Sciences and Disorders

I am doing a research study. A research study is a special way to find out about something. I want to find out if kids score the same or different on two tests. If you want to be in this study, you will have to take a vocabulary test where you will first point at pictures of the words that I say and then you will tell me what some words mean. It is OK if you don't know the answers. Some questions are very difficult and it is OK to guess if you aren't sure or just say "I don't know". Next we will talk about your spelling words. I will ask you to tell me what some of the words mean. You will also take a reading test. Some of the things that I ask you to read will be easy for you and other things will be hard. It is OK to guess or say that you don't know. Just try your best. We will video record the testing session so that I can make sure that we have scored your answers correctly.

I want to tell you about some things that may happen to you if you are in this study. You may get tired of answering my questions. Or you may get tired of sitting for a long time. We will take a break in between the tests so that you can stretch, walk around or get a drink of water. If you want to take a break at any other time, you can tell me.

Not everyone who is in this study will benefit. A benefit means that something good happens to you. If you decide to be in the study and take the tests, you will get $20 cash. You will also be helping with research. I hope that other people will be able to learn something from what we find out in this study.

When we are done with the study, I will write a report about what we find out. I will not use your name in the report. You do not have to be in this study. It is up to you. If you want to be in the study, but change your mind later, you can stop being in the study.

If you want to be in this study, please sign your name.

Your name (printing is OK) Date

I certify that this study and the procedures involved have been explained in terms the child could understand and that he/she freely assented to participate in the study.

Signature of person obtaining assent Date

REFERENCES

Anderson, R.C., Reynolds, R.E., Schallert, D.L, & Goetz, E.T. (1977). Frameworks for comprehending discourse. *American Educational Research Journal, 15,* 433-440.

Beck, I., McKeown, M., & Kucan, L. (2013). *Bringing words to life* (2nd ed.). New York, NY: The Guilford Press.

Beck, I., Perfetti, C. A., & McKeown, M. G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology*, *74*(4), 506-521. doi:10.1037/0022-0663.74.4.506

Biemiller, A., & Slonim, N. (2001). Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology*, *93*(3), 498-520. doi: I0.1037//0022-0663.93.3.498

Blachowicz, C., Ogle, D., Fisher, P., & Taffe, S. (2013). *Teaching academic vocabulary K-8: Effective practices across the curriculum*. New York, NY: The Guilford Press.

Buly, M. R., & Valencia, S. W. (2002). Below the bar: Profiles of students who fail state reading assessments. *Educational Evaluation and Policy Analysis*, *24*(3), 219-239. Retrieved from http://epa.sagepub.com/content/24/3/219.full.pdf+html

Carrow-Woolfolk, E. (1999). *Comprehensive assessment of spoken language*. Circle Pines, MN: American Guidance Service.

Chall, J. & Jacobs, V. (2003). The classic study on poor children's fourth grade slump. *American Educator* (Spring 2003 ed.). Retrieved from http://www.aft.org/newspubs/periodicals/ae/spring2003/hirschsbclassic.cfm

Clariana, R., & Wallace, P. (2002). Paper–based versus computer–based assessment: Key factors associated with the test mode effect. *British Journal of Educational Technology*, *33*(5), 593-602.

Corson, D. (1997). The learning and use of academic English words. *Language Learning, 47(4),* 671-718.

Cortiella, C. (2005). No child left behind: Understanding assessment options for IDEA-eligible students. *National Center for Learning Disabilities.* Retrieved from http://www.cehd.umn.edu/NCEO/onlinepubs/NCLD/AssessmentOptions.pdf

Dollaghan, C. A. (1985). Child meets word: "Fast mapping" in preschool children. *Journal of Speech and Hearing research*, *28*(3), 449-454. doi:10.1044/jshr.2803.454

Dollaghan, C. A. (1987). Fast mapping in normal and language impaired children. *Journal of Speech and Hearing Disorders, 52*, 218-222. doi:0022-4677/87/5203-0218501.00/0

Dunn, L, & Dunn, D. (1997). *Peabody picture vocabulary test, third edition* (3rd ed.). Minneapolis, MN: Pearson Assessments

Dunn, L., & Dunn, D. (2007). *Peabody picture vocabulary test, fourth edition* (4th ed.). Minneapolis, MN: Pearson Assessments.

Gardner, M. (1985). *Receptive one word picture vocabulary test*. Norvato, CA: Academic Therapy Publications.

Gardner, M. (1990). *Expressive one word picture vocabulary test-revised*. Norvato, CA: Academic Therapy Publications.

Gray, D.E. (2004). *Doing research in the real word.* Thousand Oaks, CA: Sage Publications Inc.

Gray, S., Plante, E., Vance, R., & Henrichsen, M. (1999). The diagnostic accuracy of four vocabulary tests administered to preschool-age children. *Language, Speech, and Hearing Services in Schools*, *30*(2), 196-206. doi:10.1044/0161-1461.3002.196

Honig, B., Diamond, L., & Gutlohn, L. (2008). *Teaching reading sourcebook: For all educators working to improve reading achievement* (2nd ed.). Novato, California: Arena Press.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27*(2), 236-248.

Jalongo, M. R., & Sobolak, M. J. (2010). Supporting young children's vocabulary growth: The challenges, the benefits, and evidence-based strategies. *Early childhood education journal* , *38*, 421-429. doi: 10.1007/s10643-010-0433-x

Kearns, G., & Biemiller, A. (2010). Two-questions vocabulary assessment: Developing a new method for group testing in kindergarten through second grade. *Journal of Education*, *190*(1/2), 31-42.

Kim, J. S., & Sunderman, G. L. (2005). Measuring academic proficiency under the No Child Left Behind Act: Implications for educational equity. *Educational Researcher*, *34*(8), 3-13. doi: 10.3102/0013189X034008003

Li, P., Xiaowei, Z., & MacWhinney, B. (2007). Dynamic self-organization and early lexical development in children. *Cognitive Science, 31*, 581-612.

Lim, E. C., Ong, B. K., Wilder-Smith, E. P., & Seet, R. C. (2006). Computer-based versus pen-and-paper testing: Students' perception. *Annals-Academy of Medicine Singapore*, *35*(9), 599. Retrieved from http://annals.edu.sg/PDF/35VolNo9Sep2006/V35N9p599.pdf

Merino, K., & Beckman, T. O. (2010). Using reading curriculum based measurements as predictors for the measure of academic progress map standardized test in nebraska. *International Journal of Psychology: A Biopsychosocial Approach/Tarptautinis psichologijos zurnalas: Biopsichosocialinis poziuris*, *6,* 85-98.

Nagy, W. E., & Herman, P. A. (1987). Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction. In M.G. McKeown & M.E. Curtis (Eds.), *The Nature of Vocabulary Acquisition* (pp. 19-35). Hillsdale, NJ: Lawrence Erlbaum Associates.

Nagy, W.E., Herman, P.A., & Anderson, R.C. (1985). Learning words from context. *Reading Research Quarterly, 20(2),* 233-253.

Northwest Evaluation Association. (2011). *2011 normative data*. Retrieved from http://www.nwea.org/sites/www.nwea.org/files/resouces/2011_Normative_Data_Overview.pdf

Northwest Evaluation Association. (2012a). *Proctor handbook for client-server map users.* Retrieved from http://www.nwea.org/support/article/1150/proctor-handbook-map®

Northwest Evaluation Association. (2012b). *RIT reference chart for reading.* Retrieved from http://www.nwea.org/node/4863

Northwest Evaluation Association. (2014a). *Linking MAP to state tests: Proficiency cut score estimation procedures.* Retrieved from http://www.nwea.org/our-research/state-information?state=ND

Northwest Evaluation Association. (2014b). *Map overview.* Retrieved from http://www.nwea.org/products-services/assessments/map®

O'Grady, W. (2005). *How children learn language*. Cambridge, UK: Cambridge University Press.

Olson, A. (2005). Improving schools one student at a time. *Educational Leadership*, *62*(5). Retrieved from http://content.ebscohost.com/pdf9/pdf/2005/EDL/01Feb05/15969221.pdf?T=P&P=AN&K=15969221&S=R&D=keh&EbscoContent=dGJyMNLe80SeqLU4y9fwOLCmr0yep7NSr6i4Ta6WxWXS&ContentCustomer=dGJyMPGvrkiwra5OuePfgeyx44Dt6fIA

Paul, R. (2007). *Language disorders from infancy through adolescence: Assessment and intervention* (3rd ed.). St. Louis, MO: Mosby Elsevier.

Paul, R. (2012). *Language disorders from infancy through adolescence: Listening, speaking, reading, writing, and communicating* (4th ed.). St. Louis, MO: Mosby Elsevier.

Pearson, P. D., Hiebert, E. H., & Kamil, M. L. (2007). Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly*, *42*(2), 282-296. doi:10.1598/RRQ.42.2.4

Plante, E., & Vance, R. (1994). Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools*, *25*(1), 15-24.

Read, J. (2000). In Alderson J., Bachman L. (Eds.), *Assessing vocabulary*. Cambridge, UK: Cambridge University Press.

Restrepo, M. A., Schwanenflugel, P. J., Blake, J., Neuharth-Pritchett, S., Cramer, S. E., & Ruston, H. P. (2006). Performance on the PPVT–III and the EVT: Applicability of the measures with African American and European American preschool children. *Language, Speech, and Hearing Services in Schools*, *37*(1), 17-27.

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical evaluation of language fundamentals, fourth edition*. San Antonio, TX: Psychological Corporation.

Shipley, K., & McAfee, J. (2009). *Assessment in speech-language pathology* (4th ed.). Clifton Park, NY: Delmar Cengage Learning.

Srivastava, P., & Gray, S. (2012). Computer-based and paper-based reading comprehension in adolescents with typical language development and language-learning disabilities. *Language, Speech, and Hearing Services in Schools*, *43*(4), 424-437.

Taylor, R. (1990). Interpretation of the correlation coefficient: A basic review. *Journal of Diagnostic Medical Sonography. 1,* 35-39.

The Kingsbury Center. (2010). *North Dakota linking study: A study of the alignment of the NWEA RIT scale with the 11th grade North Dakota state assessments*. Retrieved from http://www.nwea.org/sites/www.nwea.org/files/resources/ ND%20Linking%20Study%20August%202010%20rev%202010_1129.pdf

The Kingsbury Center. (2012). *Kentucky linking study: A study of the alignment of the NWEA RIT scale with Kentucky's performance rating for educational progress.* Retrieved from http://www.nwea.org/sites/www.nwea.org/files/resources/ KY_2012_LinkingStudy.pdf

Vadasy, P., & Nelson, J. (2012). In Harris K., Graham S. (Eds.), *Vocabulary instruction for struggling students*. New York, NY: The Guilford Press.

Weaver, C. A. I., & Kintsch, W. (1991). Expository text. In R. Barr, M. L. Kamil, P. B. Mosenthal & P. D. Pearson (Eds.), *Handbook of reading research volume II* (pp. 230-245). Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Werner, E. O., & Kresheck, J. (1983). *Structured photographic expressive language test (2nd ed.).* Sandwich, IL: Janelle Publications.

Wiederholt, L., & Bryant, B. (2012). *Gray oral reading test, fifth edition.* Austin, Tx: Pro-ed.

Williams, K.T. (1997). *Expressive vocabulary test.* Minneapolis, MN: Pearson Assessments.

Williams, K. T. (2001). *Group reading assessment and diagnostic evaluation*. Circle Pines, MN: American Guidance Service.

Williams, K. T. (2007). *Expressive vocabulary test, second edition* (2nd ed.). Minneapolis, MN: Pearson Assessments.

Yildirim, K., Yildiz, M., & Ateş, S. (2011). Is vocabulary a strong variable predicting reading comprehension and does the prediction degree of vocabulary vary according to text types. *Educational Sciences: Theory & Practice*, *11*(3), 1541-1547.