2017

The Relationship Between the Gray Oral Reading Test, Fifth Edition (GORT-5) and Woodcock Johnson Test of Achievement, Fourth Edition (WJ IV ACH) for Referred Children

Megan Edwards
edwards149@marshall.edu

THE RELATIONSHIP BETWEEN THE GRAY ORAL READING TEST, FIFTH EDITION (GORT-5) AND WOODCOCK JOHNSON TEST OF ACHIEVEMENT, FOURTH EDITION (WJ IV ACH) FOR REFERRED CHILDREN

A thesis submitted to
the Graduate College of
Marshall University
In partial fulfillment of
the requirements for the degree of
Education Specialist
In
School Psychology
by
Megan Edwards
Approved by
Dr. Lanai Jennings, Committee Chairperson
Dr. Sandra Stroebel
Dr. Peter Prewett
Dr. Conrae Lucas-Adkins

Marshall University
May 2017

May 2017

APPROVAL OF THESIS

We, the faculty supervising the work of Megan Suzanne Edwards, affirm that the Thesis, *The Relationship Between the GORT-5 and WJ IV ACH for Referred Children,* meets the high academic standards for original scholarship and creative work established by the School Psychology Program and the College of Education. This work also conforms to the editorial standards of our discipline and the Graduate College of Marshall University. With our signatures, we approve the manuscript for publication.

Dr. Lanai Jennings, Department of School Psychology Committee Chairperson Date 5-3-17

Dr. Sandra Stroebel, Department of School Psychology Committee Member Date 5-3-17

Dr. Peter Prewett, Department of School Psychology Committee Member Date

Dr. Conrae Lucas-Adkins, Department of School Psychology Committee Member Date 5-3-17

ACKNOWLEDGMENTS

I would first like to thank Jon Caleb, my family, and friends who provided the much needed support through this process. I would also like to thank Dr. Lanai Jennings for her dedication to this study as well as the long nights she spent tirelessly helping me revise each section. You saw the potential from the beginning. Also, thank you to my committee members Dr. Sandra Stroebel, Dr. Conrae Lucas-Adkins, Dr. Linda Winter, and Dr. Peter Prewett who took the time to help me complete this thesis. Furthermore, this study would not have happened without the incredible help of Mary Toler, Rachel James, Rebecca Wendell, Mallory Frampton, and Paula King. Their hard work and time put forth completing assessments built this study.

TABLE OF CONTENTS

LIST OF TABLES

ABSTRACT

Most children will learn to read when provided early literacy opportunities and reading instruction in early elementary school. However, approximately 20 percent of children will struggle to read and necessitate supplemental reading intervention. A smaller subset of struggling readers, moreover, will require sustained, intensive reading intervention, as well as accommodations provided through special education. Although universal screeners and progress monitoring instruments can be useful in guiding instructional strategies and gauging a student's response to an evidenced-based reading intervention, norm-referenced screening tools can additionally be beneficial in identifying struggling readers who likely have specific learning disabilities. The purpose of the present study was to determine the utility of the GORT-5 as a screening instrument for reading disabilities, as well as its relationship with the WJ IV ACH. In a naturalistic sample of 41 referred participants, findings indicate the GORT-5 Oral Reading Index (ORI) score positively correlates with the WJ IV ACH. The areas of comprehension from both the GORT-5 and the WJ IV ACH correlated well with each other but not with the other areas of reading. Also, the majority of students identified as at-risk on the GORT-5 were also indicated to be at-risk on the WJ IV ACH. Overall, the GORT-5 was shown to be an accurate screener in the area of reading.

CHAPTER 1

LITERATURE REVIEW

As Multi-tiered Systems of Support (MTSS) continue to evolve in the schools, school teams should intuitively be able to better identify and serve struggling readers at an early age (Decker, Englund, & Albritton, 2012; Flanagan & Alfonso, 2011). However, the growth of the assessment technology, which supports the identification of students for reading intervention in MTSS, can engender unintended outcomes for other core elements of MTSS. The need for the current study, examining the relationship between the Gray Oral Reading Test, Fifth Edition (GORT-5) and the Woodcock Johnson Test of Achievement Fourth Edition (WJ IV ACH) in reading, surfaced as a result of the unintended outcomes of the use of computer-based assessment in one district. While the use of a commercial, computer-based assessment, administered in whole group, efficiently allowed teachers to obtain universal screening data three times per year, as well as progress monitoring data, the results often proved to be highly variable for students at Tier III.

The variability in computer-based progress monitoring data proved difficult for student assistance teams to effectively discriminate struggling readers from those who were suspected of specific learning disabilities in reading. As a consequence, diagnostic staff started to routinely administer the GORT-5 to better gauge the level of reading difficulty of referred students or those being considered for a special education referral. The WJ IV ACH was additionally administered in the district to most students referred for a special education evaluation. The scores on both the GORT-5 and the WJ IV ACH

were analyzed to see if student's scores on the screener are similar to the scores on the WJ IV ACH.

The findings of the study help to indicate if the GORT-5 is an accurate assessment tool for identifying students who are struggling readers. However, it is first important to review the literature and describe the difference between struggling readers and students with specific learning disabilities in reading, as well as the processes used to differentiate the two in schools. In the literature review, the overall concepts of MTSS will be covered to show its role in the schools and process of identifying students as having a specific learning disability. Second, we will discuss some screeners that have been reviewed in the literature. Thirdly, the purpose of this study is reviewed.

Struggling Readers

Students who are struggling readers, defined as readers who have a deficit in reading independently, can best be helped if they are identified at an early age (Beers, 2003). These students are slightly, a year and a half and less, behind their peers when it comes to reading independently. They typically need extra support developing vocabulary and phonemic awareness, the ability to hear and manipulate the sounds in spoken language (Osewalt, n.d.). Also, as children reach grade 3, the curriculum relies more on reading skills overall, and students have to use those skills in math, science and social studies. Therefore, if they are not identified, their deficits will grow and the students will fall further behind as they continue with their education. This is the point when reading disabilities can occur. According to McCray (2001), students who lack reading skills are more likely to carry those deficiencies throughout schooling and are at a greater risk of dropping out than their peers who are proficient readers. With so many

children in the United States struggling with reading, more concise and specific screeners as well as achievement assessments are needed in order to accurately identify these students for possible support services.

Students with Specific Learning Disabilities in Reading

According to the United States Department of Education, in 2014, 8.7 percent (5,944,241) of students ages 6 to 21, nationally, were identified as having a disability and received special education services under IDEA, Part B (U.S. Department of Education, 2016.) Of this total, 39.2 percent of students were classified under the category of specific learning disabilities, making it the largest disability category under IDEA, Part B. Under the area of specific learning disabilities, 80% of these students have a reading disorder. The second largest disability category was speech or language impairments (17.6%) (U.S. Department of Education, 2016.)

Children and adolescents with reading disabilities differ from students who are struggling readers in that they read significantly, more than one and a half years, below their age level peers. Although there are a variety of identification methods used to determine SLD eligibility (e.g., Response to Intervention, Patterns of Strengths and Weaknesses, and Ability-Achievement Discrepancy approaches), as well as disagreement among researchers and practitioners alike, as to the validity of each model (Burns, 2016; Feifer & Della Toffalo, 2007; Flanagan & Alfonso, 2011), the Individuals with Disabilities Education Act of 2004 (IDEA 2004) defines SLD as a disability involving "basic psychological processes" that is marked by "the imperfect ability to listen, think, speak, read, write, spell, or to do mathematical calculations" (IDEA, 2006, p.46757). These students usually show a pattern of strengths and weaknesses in their cognitive and

achievement scores, which align to highlight a disability. This differs from students who exhibit overall depressed cognitive abilities, which align to low academic achievement, including the area of reading.

According to Feifer and Della Toffalo (2007), there are four distinct types of reading disorders. The first, dysphonetic dyslexia, involves the inability to sound out words phonologically. This includes the skills to hear the individual sounds in a word and put them together, break down a word into the individual sounds, and blend sounds together to make words (Osewalt, n.d.). Students with dysphonetic dyslexia tend to memorize words and learn through visual and orthographic cues (Feifer & Della Toffalo, 2007). The second type is surface dyslexia, which is the opposite of the first reading disorder. These students can easily sound out words, but lack the ability to fluently recognize words in print. They do not pick up on visual cues, and have to read words by breaking them down by individual sounds. The third reading disability is called mixed dyslexia. Students with this disability have trouble with decoding words, picking up on visual cues, reading fluency, and verbal memory. Lastly, reading comprehension is the fourth reading disability subtype. These students have good decoding skills, but cannot hold the information long enough to understand what they have read (Feifer & Della Toffalo, 2007). These disabilities cause students to fall further behind their peers than students who have these skills.

Systems of Support

Before testing for a learning disability can occur, a child must go through a process referred to as Multi-tiered Systems of Support (MTSS) in many districts and states. MTSS is a multi-tiered framework to identify and help support students with

academic or behavioral disabilities (Flanagan & Alfonso, 2011). This process is used to ensure sufficient intervention is being given to each student and to help show their rate of learning and growth overall. It begins with universal screening, Tier I, to ensure all students are receiving a beneficial education and are learning sufficiently. Tier I, or core instruction, involves the daily provision of scientifically-based reading instruction that is typically differentiated within the general education setting to support all students. These screenings, which are generally administered three to four times per year, also help teachers by pointing out academic strengths and weaknesses in students, which help them tailor instruction to students (Decker, et al., 2012; Flanagan & Alfonso, 2011; Merrell, Ervin, & Peacock, 2011; Shinn & Walker, 2010).

Although districts have numerous types of universal screeners to choose from for their MTSS, district staff must consider several important components before formally adopting commercially developed assessments or making their own. First they must consider what the assessments measure and how they connect to the curriculum (Merrell & Tymes, 2007). Is the content assessed consistent with the scope and sequence of the curriculum at each grade level, or are the skills required to meet end of year goals? Secondly, cost and efficiency of test administration are other important variables, as well as the ease at which results are reported. Although paper-pencil formats were more common in the past, schools are increasingly using technology to implement the universal screening (Merrell & Tymms, 2007). While computer-based assessments increase ease of whole group administration, such screeners do have limitations. For example, they require knowledge of computers in order to use them. It also may be more difficult for students with disabilities to maneuver the mouse on a computer than it would

be to answer questions verbally after they were read aloud. Also, while the computer assessments are standardized, it is easier for a student to click mindlessly through the questions without showing how much they really know and understand. Finally, with human interaction, it is easier to observe if a student truly understands the question and is answering to his or her best ability. Observations done through written assessments help to indicate the extent to which attention and motivation impact reading results (Merrell & Tymms, 2007).

Last and most importantly, school teams must consider the psychometric properties like cut scores, sensitivity and specificity. After administering these assessments, school teams often use percentile ranks to determine cut scores to indicate if students are at-risk and will require supplemental instruction at Tier II or III (Jenkins, Hudson, & Johnson, 2007; Nese, Park, Alonzo, & Tindal, 2011). For example, if a school sets the beginning of year cut score at the 25th percentile rank, this means any student at or below the 25th percentile would be considered at-risk and likely moved to Tier II for supplemental instruction. However, when setting cut scores or establishing decision points for transitioning students to Tier II school teams should consider test sensitivity and specificity. Test sensitivity and specificity are concepts critical to the establishment of cut scores and subsequent provision of intervention. Sensitivity is defined as the accuracy rate of a measure in identifying students who truly have reading difficulties. In other words, it is the likelihood a measure will identify true positives. Conversely, specificity is referred to as the accuracy of a measure in identifying if an individual is not at-risk, when they will truly perform average or above average. This is called a true

negative. This would indicate if a test accurately identifies students who do not have a reading problem (Wiederholt & Bryant 2012a).

Furthermore, both false positives and false negatives create difficulties for school teams. For example, a certain percentage of students may score just below the cut score percentile and are classified being at-risk and in need of intervention. In these instances, the students are still placed in Tier II intervention programs and, yet, may actually have made adequate progress without the added intervention. In this situation, the students were misidentified as being at-risk for reading difficulties and received intervention that other students with true reading difficulties, in fact, needed (Jenkins et al., 2007). To illustrate this point, STAR uses a cut score at the 40th percentile. When Clemens et al. (2015) tested students who met that criterion, they found high specificity and low sensitivity, meaning there was a large number of students who were not identified as at-risk that needed to be. They raised the cut scores to what they considered to be an acceptable level, but had high specificity. This resulted in students who were not at-risk being identified as at-risk. Ultimately, Clemens et al. (2015) found the use of a word-reading fluency cut score at the 15th percentile brought about the most accurate results with fewer false positives. It is the balance between specificity and sensitivity that makes a test accurate in defining a cut score (Clemens et al., 2015).

After school teams complete assessments and set cut scores, they utilize their grade level and/or problem-solving teams and decide which students are at-risk or struggling based on the totality of the data presented. Students designated at-risk are then provided increasingly targeted interventions in Tier II of the MTSS due to the insufficient response to the core instruction, as evident through the universal screenings and

benchmarks for Tier I. The targeted intervention provided at Tier II is matched to the area of need in the essential areas of reading in order reach end of grade level reading goals. Tier II involves short-term strategic instruction for support in addition to core instruction, generally in small groups for a minimum of nine weeks. The targeted instruction is typically provided for three to five 30-minute sessions each week. Progress monitoring is completed every two to three weeks to monitor gains or slippage and adjust instruction to the students' ongoing needs (Decker et al., 2012; Flanagan & Alfonso, 2011; Merrell K. et al., 2011; Shinn & Walker, 2010). The National Center on Intensive Intervention publishes tools and charts, which provide information to consumers on the quality and validity of many screening and progress monitoring tools (National Center on Intensive Intervention, n.d.).

If the student continues to struggle after substantial targeted instruction, tertiary intervention, or Tier III, can be implemented to provide more sustained and intensive reading instruction. Tier III intervention involves intensive instruction in a small group or individual setting for three to five 30-60 minute sessions per week. Although districts can define their MTSS session length and intervention duration according to their district needs, many implement Tier III for a minimum of nine to twelve weeks. Teachers or interventionists continue to obtain regular progress monitoring data. In the case of inconsistent or poor progress, school teams may opt to initiate a second or third round of Tier III to provide access to sustained, intensive instruction (Decker et al., 2012; Flanagan & Alfonso, 2011; Merrell K. et al., 2011; Shinn & Walker, 2010).

Despite access to sustained, intensive intervention, some children and adolescents fail to meet end of intervention goals, even when these goals were developed based on

modest growth trajectories. In these situations, particularly wherein strong, supplemental interventions are provided over multiple consecutive months, teachers, parents, and other interventionists may begin to suspect an underlying reading disability. At this point, an initial referral to special education may be requested, especially if progress-monitoring data indicate low, persistent achievement and the need for long-term support (Merrell K. et al., 2011; Shinn & Walker, 2010). To qualify for special education, IDEA 2004 requires a formal comprehensive evaluation to assess all areas of suspected disability and all areas of suspected need, thus generally requiring an individual achievement test (IDEA, 2006).

However, if a student's response is variable over time, the need for a referral may be less clear. In this case, norm-referenced reading screeners or other diagnostic assessments may be used to quantify a student's achievement levels. Specific screeners are available for reading, mathematics and writing. An example is the new Wide Range Achievement Test Fifth Edition (WRAT5), which will give a quick assessment of a student's academic levels and help identify possible learning disabilities (Wilkinson & Robertson, 2017). Norm-referenced screeners can better gauge how low a student is achieving compared to same-age peers in a normative sample and can be beneficial in determining if special education testing is warranted. Screeners help discriminate students who may only need extra supports in the classroom from those who are better candidates for special education referrals. Because those students who don't meet the screening cut scores are typically not moved along for referrals, referral accuracy is likely to increase, rendering a higher placement rate among those receiving full special education evaluations (Merrell K. et al., 2011; Shinn & Walker, 2010; VanDerHeyden,

Witt, & Naquin, 2003). As such, the use of a screener also decreases the amount of time spent testing students for special education who may not qualify under the strict criteria set by each state, while increasing the amount of time that can be spent focusing on different interventions for children (VanDerHeyden, et al., 2003).

Although many studies have examined the predictive or concurrent validity of screeners, a high percentage of these studies involve curriculum based measurements at the benchmark periods and their relationship to statewide summative testing (Ditkowsky & Koonce, 2010; Hunley, Davies, & Miller, 2012; Marchand, & Furrer, 2014; McGlinchey and Hixson, 2004; Miller, Bell, & McCallum, 2015; Patton, Reschly, & Appleton, 2014; Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2008). Fewer recent studies have described the relationship between more traditional, norm-referenced measures. Prewett (1995) examined the concurrent validity of two cognitive screeners, the Matrix Analogies Test- Short Form (MAT) and the Kaufman Brief Intelligence Test (K-BIT) with the Wechsler Intelligence Scale for Children-Third Edition (WISC-III). Results indicated the two cognitive screeners gave mean standard scores that were significantly higher than the intelligence assessment (Prewett, 1995). A separate study, completed by Erford, O'Brocki, and Moore-Thomas (2007), analyzed the construct-related validity of a screener, the Reading Essentials Skills Screener-Upper Elementary Version (RESS-U). The RESS-U is a quick screener, given in approximately 10 minutes with 35 reading items (Erford, et al., 2007). Researchers compared pre-test and post-test scores of struggling readers and proficient readers with three other commonly used reading assessments: Wide-Range Achievement Test- Revised (WRAT-R), Woodcock-Johnson Tests of Achievement- Revised (WJ-R Broad Reading cluster), and Slosson Oral

Reading Test- Revised (SORT-R). Their results indicated high construct-related validity when the RESS-U was compared to the three reading assessments, meaning the RESS-U was a reliable screener.

VanDerHeyden, et al. (2003) completed a study comparing Curriculum Based Assessments (CBAs) in reading and math to scores on psychoeducational assessments. They also reviewed the accuracy of teacher referrals versus CBA referrals. More specifically, they first assessed a pre-test and post-test of the CBA data using the Basic Skill Builders and Comprehensive Inventory of Basic Skills-Revised (CIBS-R) reading and math probes. Next, they evaluated the predictive validity using students identified as at-risk from the CBAs by giving them the Woodcock-Johnson psychoeducational battery-Revised (WJ-R) and the Iowa Test of Basic Skills (ITBS), individualized achievement assessments. The investigators administered CBAs and the CIBS-R to first and second grade students. Next, they were given tests in the WJ-R that measure reading and math which included Letter Word Identification, Passage Comprehension, Calculation, and Applied Problems. The assessments were used to compare to a process called Problem Validation Screening (PVS). This consists of class wide academic assessments, performance/skill deficit assessments, and a brief instructional session. Their findings indicated that the CBMs identified only five more students overall than the WJ-R that needed special education (VanDerHeyden, et al., 2003).

GORT-5

Although there are many reading screeners available for use, no published studies to date in the peer-reviewed literature have examined the utility of the GORT-5 by examining its relationship to the WJ-IV (ACH) reading tests. The GORT-5 is an assessment that is widely used throughout schools to measure oral reading. This assessment was first published in 1963 by Dr. William S. Gray and was most recently revised in 2012 in its fifth edition. This tool can be used to identify students who exhibit oral reading difficulties and additionally help pinpoint what specific reading areas or behaviors are problematic. This assessment is relatively brief and is normed for students age six through 23. The GORT-5 has two equivalent forms, both with 16 passages that increase in difficulty as the student moves from one to the next (Hall & Tannebaum, 2012; Wiederholt & Bryant, 2012a). A grade-level table, provided in the manual, indicates the entry points. The student is given a passage to read aloud while the examiner keeps time and checks for errors. Examiners can also record other reading behaviors such as substitutions, omissions, and additions along with a reader's prosody (Hall & Tannebaum, 2013; Wiederholt & Bryant, 2012b) Testing continues until a basal and ceiling is reached. The basal score is met when a student's fluency score, sum of rate and accuracy for one passage, is 9 or 10. The ceiling is reached when a student's fluency score is 2 or fewer (Hall & Tannebaum, 2013). After the student reads the passage aloud, he or she is read five comprehension questions that they have to answer in an open-ended format (Hall & Tannebaum, 2013; Wiederholt & Bryant, 2012b).

The GORT-5, consequently, measures reading rate, accuracy, fluency, and comprehension. Rate refers to how quickly a student can read, as well as automaticity.

Accuracy tests measure the extent to which a student can read the word correctly. Fluency combines rate and accuracy to measure how fast and accurately a student can read. Comprehension, on the other hand, refers to the extent to which a student can obtain meaning from the text. However, research demonstrates that fluency influences comprehension. If a student is a slow reader, this will decrease their ability to understand what they have read. The same goes for accuracy (Flanagan and Alfonso, 2011). If a student is unable to blend sounds together or break words down, they are likely to focus more on decoding and thereby extract less meaning from the text.

The comprehension portion of the GORT has frequently been criticized. Prior to the fifth edition of the GORT, the comprehension section was made up of questions each student could read for oneself with a list of multiple-choice answers. Many examiners using the prior version observed the multiple-choice items were problematic as examinees were able to successfully answer the comprehension questions due to prior knowledge despite having extreme difficulty reading the passage. This phenomenon is known as "passage independence" (Wiederholt & Bryant, 2012a, p. 58). Keenan and Betjmann (2006) conducted a study in 2006 that measured the passage dependence of the Gray Oral Reading Test, Fourth Edition (GORT-4) using 77 university students. The students were asked to read the story prompt at the beginning of each passage and then answer the GORT-4 multiple-choice answers. Their research revealed that 86% of the questions could be answered correctly without reading the passage. To combat this issue, the comprehension section of the fifth edition was changed to include open-ended questions that were read aloud to the student. The authors also removed the passage prompts at the beginning of each story. After these were changed, the authors completed

a survey using 77 university students to measure the amount of passage dependence in the fifth edition. Their results indicated that only "eighty-eight percent of the comprehension questions were highly passage dependent due to only 10% of students being able to answer the questions correctly by guessing" (Wiederholt & Bryant, 2012a, p. 58), thereby increasing the overall validity of the comprehension component of the latest edition (Hall & Tannebaum, 2013).

In order to assess the extent to which the GORT-5 measures what it was designed to measure, Wiederholt & Bryant (2012a) conducted criterion-validity studies by comparing the scores on the GORT-5 to five similar reading assessments: the Nelson-Denny Reading Test (NRDT); the Reading Observation Scale (ROS); the Test of Silent Contextual Reading Fluency (TOSCRF); the Test of Silent Reading Efficiency and Comprehension (TOSREC); and the Test of Silent Word Reading Fluency (TOSWRF). Findings from these analyses suggest large, positive correlations. Coefficients yielding scores of .70-.89 are considered very large indicating they are very similar, while coefficients between .90 and 1.00 are considered almost perfect (Hopkins, 2002). The scores that yielded a .66 correlation and above are indicated in Table 1. It is important to note, no scores yielded a coefficient lower than .54 (Hopkins, 2002; Wiederholt & Bryant, 2012a, p. 64). These findings indicate that the GORT-5 scores when correlated with the NDRT, TOSCRF, and TOSREC scores yielded very large, positive coefficients (Hopkins, 2002; Wiederholt & Bryant, 2012a).

Need for Study

The purpose of this study is to identify how well the GORT-5 scores correspond with the WJ IV ACH scores in reading. This investigation is necessary to see how useful

the GORT-5 is as a screening measure for children and adolescents suspected of having reading disabilities. While there has been research conducted with previous versions of the GORT, no peer-reviewed studies are available within the literature that examine the relationship between the GORT-5 and the WJ IV ACH. Quantifying the relationship between the two instruments is an important step in validating the use of the GORT-5 as a robust screening tool, which can accurately identify students with potential reading disabilities.

Research Question 1: What is the correlation between the GORT-5 Oral Reading Index score and the WJ IV ACH Broad Reading score? The investigator predicts there will be a very large, positive correlation given the similarity in constructs measured.

Research Question 2: What is the correlation between the GORT-5 Oral Reading Index score and the other WJ IV ACH reading tests and cluster scores? The investigator predicts there will be a medium, positive correlation between the reading scores.

Research Question 3: Is the GORT-5 ORI standard score equivalent to the WJ IV ACH Broad Reading cluster standard score? The investigator predicts the mean scores will be comparable and no significant differences will emerge.

Research Question 4: What percentage of students, who perform at or below the tenth percentile (WVBE Policy 2419) on the GORT-5 ORI standard score, perform at or below this same criterion on the WJ IV ACH Broad Reading cluster? The investigator predicts there will be a large percentage of students who perform at or below the tenth percentile on both assessments.

CHAPTER 2

METHOD

Participants

Forty-one students in grades second through eighth participated in the investigation in a rural school district in the Mid-South region of the United States. All students were referred for a special education evaluation. Of the 41 participants, 19 (46.3%) were females and 22 (53.7%) were males. In this study, 24 (58.5%) of the participants were in elementary school, whereas 17 (41.5%) were in middle schools. The majority of students, 37 (90.2%), were White, non-Hispanic. Another 4.9% (2) and 4.9% (2) reported Multiple Race/Ethnicities and Hispanic Race/Ethnicities, respectively. Only students assessed for a suspected disability category under the Individuals with Disabilities Education Act (IDEA) of 2004 were included in the sample. Students referred for gifted only exceptionalities were excluded.

Materials

The current investigation incorporated multiple assessment measures from the GORT-5 and the WJ IV ACH as outlined below.

GORT-5. The GORT-5 is a tool that measures oral reading. The individual scales include rate, accuracy, fluency, and comprehension. An overall ORI is additionally calculated from the combined fluency and comprehension scales. The assessment contains two equivalent forms, A and B, wherein each form has 16 increasingly difficult stories. A student reads the story aloud while the examiner keeps time and marks for errors or miscues. Comprehension is measured afterwards in the form of five open-ended questions. Raw scores, percentile ranks, and scaled scores are based on a mean of 10 with

a standard deviation of 3. The ORI scores are based on a normal distribution containing a mean of 100 and a standard deviation of 15.

GORT-5 Reliability and Validity. According to the authors, Wiederholt and Bryant (2012a), the GORT-5 has average internal consistency reliability and the reliability coefficients exceed .90 for both forms. Average test-retest coefficient for the ORI for form A and form B exceed .85, as do test-retest for different forms (form A to B and form B to A). "The average corrected ORI coefficient was .93" (Hall & Tannebaum, 2012, p. 517.). Binary classification studies completed indicate that the GORT-5 has a sensitivity of .82, specificity of .86, ROC/AUC of .92, cut-score rate of 90 and shows low false positive rates. This indicates it is able to accurately identify students who have reading disabilities. The standard error of estimate is the GORT-5 ORI standard score, plus or minus 10.

WJ IV ACH. The WJ IV ACH is an individualized achievement test, which measures reading, written language, mathematics, and academic knowledge (LaForte, McGrew, & Schrank 2014). It is one of the most widely used achievement tests for special education eligibility determinations. This test identifies individual strengths and weaknesses, and the results drive intervention recommendations for those assessed. This test is conducted in order to show the strengths and weaknesses an individual has in certain areas in order to tailor education to those needs. This aids in discovering learning disabilities and developing interventions for individuals.

For the purpose of this study, only the domains that include reading were used. Tests in the reading areas include Letter-Word Identification, Passage Comprehension, Word Attack, Oral Reading, Sentence Reading Fluency, Reading Recall, Word Reading

Fluency, and Reading Vocabulary (Mather & Wendling, 2015). These tests can be used independently or can be used to create cluster scores in the areas of Reading, Broad Reading, Basic Reading Skills, Reading Comprehension, Reading Fluency, Reading Rate, and Phoneme-Grapheme Knowledge (LaForte, et al., 2014; Schrank, et al., 2014). Broad Reading encompasses the areas of comprehension, rate of reading, and accuracy. These combine to show an overall view of skills needed to read. This test can be given to ages 2-80+. Any one of three equivalent standard battery forms can be used. An extended battery form can also be tacked onto any of the three standard battery forms for additional information (Schrank, et al., 2014). In this study, only form A was given.

WJ IV ACH reliability and validity. The median reliability coefficient for the WJ IV ACH of each cluster was .90 or higher. Reliability tests were completed for each test in the WJ IV ACH and of the 39 completed, 38 of these were .80 or higher and 17 were .90 or higher. When testing between forms was analyzed, Passage Comprehension revealed a slight decrease in difficulty of items on form C than forms A and B. Raw score-to-W-ability was completed on the multiple forms of each test. The W-ability for each form on tests 1-7 was very similar, meaning they are equivalent across the ranges of abilities. Further tests completed also showed the tests are equivalent in difficulty across the forms.

Intercorrelation validity was high between the clusters from the same achievement domain and low between clusters from different domain. Correlations between "Basic Reading Skills and the two Reading Comprehension clusters were .74 and .78" (LaForte, et al., 2014, p. 144). When the Kaufman Test of Educational Achievement-Second Edition (KTEA-II) and the Wechsler Individual Achievement Test-Third Edition (WIAT-

III), were correlated with the WJ IV ACH, correlation coefficients ranged from .78 to .91 in the WJ IV Reading Cluster (LaForte, et al., 2014).

Procedure

Trained school psychologists, diagnosticians, and/or interventionists routinely administered the GORT-5 and the WJ IV ACH to students referred for special education testing or reevaluation in the school district. The assessments were completed within the 80-day timeline set forth by the West Virginia Board of Education *Regulations for the Education of Students with Exceptionalities* (WVBE Policy 2419) for initial evaluations. Counterbalancing of assessments was not possible due to the naturalistic setting and the caseload of practitioners. Furthermore, the evaluators may not have administered both assessments to the same student. The assessors administered the GORT-5 and the WJ IV ACH to each student under standard conditions and in accordance with the test procedures outlined in the GORT-5 and WJ IV ACH manuals. Twenty students were given form A of the GORT-5, whereas 19 students were administered form B and two forms were unknown. On the WJ IV ACH, form A was administered to all students. All GORT-5 protocols were hand scored, while the WJ IV ACH protocols were scored using the online scoring and reporting system at wjscore.com.

School psychologists and diagnosticians collected data from the 2014-2015 and 2015-2016 school years. Other data collected include sex, school, grade, age at testing, race/ethnicity, and all WJ IV ACH reading scores and GORT-5 scores. More specifically, the standard scores from each WJ IV ACH reading test and cluster were recorded along with the percentile ranks. Some students were not given enough reading tests to calculate a Reading Comprehension Extended score; instead they received a Reading

Comprehension score. The Reading Comprehension cluster includes the tests Passage Comprehension and Reading Recall. Reading Comprehension Extended includes those tests along with Reading Vocabulary. These scores were placed into the same category, Reading Comprehension Extended, in order to run the analyses. School psychologists and diagnosticians additionally recorded GORT-5 scaled scores and percentile ranks for reading rate, fluency, and comprehension, as well as the ORI and its corresponding percentile rank.

Data Analysis

District personnel provided the data with no identifying information in a Microsoft Excel worksheet. The investigator transferred the data to the IBM Statistical Package of Social Sciences (SPSS) software for analysis (IBM Corp, 2013). After scanning for possible data errors and outliers, the investigator ran basic descriptive statistics (i.e.; mean, median and standard deviation) for each reading standard or scaled score. Next, Pearson *r* correlation coefficients were generated. Paired samples correlations were run in order to find how the average GORT-5 ORI scores compared with each WJ IV reading cluster mean score (Broad Reading, Basic Reading, Reading Comprehension Extended, and Reading Fluency). Finally, the GORT-5 ORI and WJ IV ACH Broad Reading percentile ranks were recorded into dichotomous variables: participants who performed at or below the tenth percentile on the respective tests and those who performed above the tenth percentile on each test. The investigator generated 1) cross tabs or contingency tables to illustrate the match between the dichotomous performance of each participant and 2) a Fisher's Exact test, a Chi-Square like test appropriate for nominal, dichotomous data. A Fisher's Exact test, a statistical test, is

useful for all sample sizes and indicates if there is a significant association between binary classifications. A t-test was also conducted between form A and B of the GORT-5. No significant differences were noted; therefore, the data was combined.

CHAPTER 3

RESULTS

Research Questions 1: What is the correlation between the GORT-5 Oral Reading Index score and the WJ IV ACH Broad Reading scores?

As predicted, participants' performance on the GORT-5 ORI was significantly related to reading performance on the WJ IV ACH Broad Reading cluster as shown in Table 2 and Table 3. The Pearson *r* correlation coefficient for the GORT-5 ORI and WJ IV Broad Reading cluster was very large, $r = .81$, $p = .001$.

Research Question 2: What is the correlation between the GORT-5 Oral Reading Index score and the other WJ IV ACH reading tests and cluster scores?

As illustrated in Table 2 and Table 3, Pearson *r* correlation coefficients for the GORT-5 ORI and the WJ IV Basic Reading, Reading Fluency, and Reading Comprehension similarly yielded *r* values within the very large range that were significant at the .001 level, which correspond with our predictions. Correlations between both the GORT-5 ORI and the WJ IV Basic Reading and Reading Comprehension clusters yielded Pearson's *r* correlation coefficients of .78, whereas the correlation between the GORT-5 ORI and WJ IV Reading Fluency cluster was .70.

Toward the same end, when correlated with the WJ IV reading clusters, all GORT-5 scaled scores excluding Comprehension, additionally rendered large-to-very large correlation coefficients. Although the GORT-5 Comprehension measure was significantly related to the WJ IV reading measures, the strength of the associations (aside from the GORT-5 Comprehension and WJ IV Reading Comprehension

comparison) was generally weaker than most comparisons in Table 2. In particular, the GORT-5 Comprehension and the WJ IV Reading Fluency Cluster produced only a modest correlation ($r = .46$, $p = .003$)

Research Question 3: Is the GORT-5 ORI standard score equivalent to the WJ IV ACH Broad Reading Cluster standard score?

Next, the comparability of the GORT-5 ORI and the WJ IV ACH cluster scores was evaluated. A t-test was conducted to look at the standard score means (Table 5). Mean standard scores on the GORT-5 and WJ IV Broad Reading score were significantly different ($t= 2.87$, $p= .007$). As illustrated in Table 4, the GORT-5 ORI score, on average, was 3.58 points higher than the WJ IV Broad Reading cluster score, whereas standard deviation for the WJ IV Broad Reading score was 3.4 points higher than the GORT-5 ORI. This corresponds to observations of the individual test, wherein the lowest possible standard score on the GORT-5 was 52 (scaled score of 1) where as for the WJ the Broad Reading cluster score could go as low as a standard score of 40. The third pairing on Table 4, GORT-5 ORI and WJ IV Reading Comprehension Extended showed very similar results, with a mean difference of 2.85 ($t= 2.69$, $p= .010$). With the sample's confidence interval of 95 out of 100, we can say that the true mean difference lies between .71 and 4.99. Pair four (GORT- 5 ORI and WJ IV Reading Fluency) had a mean difference of 1.83 ($t= 1.27$, $p= .211$), but this pairing was not significant. Unlike the other pairings, in pair two (GORT-5 ORI and WJ IV Basic Reading) there was a four-point significant difference in the opposite direction ($t= -2.25$, $p= .030$). Pairings one, two and three were unlike our predictions due to the significant difference in mean scores. Pairing

four has a smaller mean difference and was not significant, which aligned more with our predictions.

Research Question 4: What percentage of students who perform at or below the tenth percentile on the GORT-5 ORI standard score, perform at or below this same criterion on the WJ IV ACH Broad Reading Cluster?

Next, the researchers wanted to discover how many students who scored at or below the tenth percentile on the GORT-5 would similarly score at or below the tenth percentile on the WJ IV ACH, indicating a true positive. The crosstabulations and Fisher's Exact test are displayed in Table 6 and Table 7. Twelve students scored at or below the tenth percentile on both assessments, which show true positives. Twenty-four students scored above the tenth percentile on both assessments, indicating true negatives. Three students scored at or below the tenth percentile on the GORT-5 and scored above the tenth percentile on the WJ IV ACH. This would indicate the GORT-5 identified them as being at-risk for a reading disability, while the WJ IV ACH did not identify them as at-risk. In other words, the GORT-5 gave a false positive score. Only two students scored above the tenth percentile on the GORT-5, but at or below the tenth percentile on the WJ IV ACH, showing a false negative. This aligned with our predictions made. There were a high number of students who were identified at being at-risk on both assessments.

CHAPTER 4

DISCUSSION

The GORT-5 and WJ IV ACH were given to forty-one students in grades second through eighth in a rural school district in the Mid-South region of the United States in order to indicate if the GORT-5 is an accurate screener. Participants were tested through the referral process for a special education evaluation. The GORT-5 was indicated to be an accurate screener when scores were compared to WJ IV ACH Broad Reading cluster scores, due to a large, positive correlation. The GORT-5 also had correlation coefficients in the very large range when compared to the WJ IV ACH Basic Reading, Reading Fluency, and Reading Comprehension Extended clusters. The individual scaled scores on the GORT-5 also rendered large-to-very large correlation coefficients when compared to the WJ IV ACH cluster scores. The GORT-5 Comprehension score produced more modest correlations to all of the WJ IV ACH clusters except for Reading Comprehension Extended.

Next, the means between the GORT-5 ORI and the WJ IV ACH cluster standard scores were compared and significant differences were noted between all of these except for ORI and Reading Fluency. The ORI score was found to be, on average, 3.58 points higher than the Broad Reading cluster score, representing a significant difference. This, in part, was believed to be a result of the lower floor provided by the WJ IV reading tests. The lowest possible standard score for the GORT-5 ORI is a 52, whereas the WJ IV lowest possible reading standard score is a 40. Therefore, although the GORT-5 is a powerful reading instrument, it has a higher floor than the WJ IV. Consequently, the GORT-5 ORI is less able to discriminate the readers with the most severe deficits from

those who are very low in the same manner as the WJ IV ACH. Secondly, the third pairing of the ORI and WJ IV Reading Comprehension Extended yielded similar results. The ORI and WJ IV Reading Fluency resulted in a smaller mean difference that was insignificant. Lastly, unlike the other pairings, there was a four-point significant difference in the opposite direction between the ORI and WJ IV Basic Reading. This could be due to the fact that the individual tests that make up the Basic Reading cluster (Letter-Word Identification and Work Attack) measure reading skills in isolation, such as individual word reading and blending, whereas the ORI score measures oral reading of entire passages and connected text comprehension.

When specificity and sensitivity were tested, the GORT-5 had twelve students with scores indicating a true positive and twenty-four students with scores indicating a true negative. Significantly higher expected counts of students were observed with true positives and negatives, than would be expected due to chance. Only three students showed a false positive, meaning they scored at or below the tenth percentile one the GORT-5, while scoring above the same criterion on the WJ IV ACH Broad Reading cluster. Two students were indicated to have false negative scores. This means they scored above the tenth percentile on the GORT-5, but below that criterion on the WJ IV ACH Broad Reading cluster. While there were five non-corresponding scores in total, each score that fell above the criterion was close to the tenth percentile. It is important to note that the tenth percentile was used as an important cut score in this particular district. However, other districts may use higher or lower percentile ranks for their cut scores.

Limitations

The present investigation includes several limitations. First, because the investigation was conducted in a school setting, multiple examiners administered the assessments. Therefore, many external factors could have contributed to the scores. Second, due to the fact the assessments were given in a random order, they were not counterbalanced. Although this is not ideal for a controlled study, this type of administration order is typical in most school systems.

Next, in many districts the Broad Reading cluster score is not used in eligibility determination due to the broad range of reading skills it measures. This cluster score was used as the primary correlation factor when comparing the WJ IV ACH to the GORT-5. Thus, additional comparisons are needed between the ORI and the WJ IV Reading Fluency, Reading Comprehension, and Basic Reading.

Also, time in between administrations was not controlled, although duration did not exceed the 80-day initial evaluation timeline requirement. Another limitation is due to the findings being based on students from one rural school district with predominantly low SES communities with little diversity in race/ethnicity. Furthermore, due to different evaluators giving the assessments, some did not administer all individual reading tests needed on the WJ IV to render a Reading Comprehension Extended score. Therefore, some students had that score while others only had Reading Comprehension. These scores were collapsed into the same category, Reading Comprehension Extended, in order to run the analyses. Reading Comprehension alone includes the Passage Comprehension and Reading Recall tests. Reading Comprehension Extended includes the tests stated prior, along with Reading Vocabulary. The addition of the reading vocabulary

may have increased students' scores, which would have led to potentially unreliable results. Also, since the WJ IV ACH measures reading at a more in-depth level, it can be difficult to pin point what single reading skill is measured on the GORT-5 in comparison to the WJ IV ACH.

Finally, the ORI score on the GORT-5 is a standard score, whereas all scores on the WJ IV ACH are standard scores. To conduct validity testing, the tests have to produce similar means as well as have a high correlation (Wiederholt & Bryant, 2012a). Scaled scores, not standard scores, represent the GORT-5 individual tests. It can be inferred that the individual scores from each assessment line up when looking into specific ranges of scores (below average, average, above average, etc.), but not when looking individually between the mean standard scores.

Future Research

Since this study focused on students referred for special education, it would be beneficial to replicate this study with a more diverse population of students, including students who do not struggle academically and those who excel in reading. This would increase the external validity. Also, it would be beneficial to code the number of tests given by each examiner in order to rule out examiner effects. It would also be beneficial to increase the *N* size to see if the same results are produced with a larger and more diverse sample. Additionally, since the Reading Comprehension Extended and Reading Comprehension scores were both used under one category, future research would benefit from analyzing the scores from each cluster in order to see if the addition of the vocabulary tests affected the scores at all.

Moreover, since the Broad Reading cluster is not used in some districts for eligibility purposes, it would be beneficial to compare the GORT-5 scores and the specific reading tests in the WJ IV ACH used in most districts for those eligibility determinations. Those scores can also be used along with the GORT-5 when analyzing the cut score percentiles. Finally, a t-test was conducted between form A and B of the GORT-5 and no significant differences were found. Future research should focus on the relationship between form A and form B separately with the WJ IV tests of reading achievement.

The results from this study indicate that the GORT-5 is a valid screener used to determine which students should be referred for a special education evaluation, as compared to those students who should not. This will help to enhance the proficiency of the referral process in school systems. In MTSS, the GORT-5 can be used as a screener to decrease the amount of time spent testing students who may not qualify for special education under the strict criteria set by each state. The GORT-5's short administration time can also decrease the amount of time the student spends out of the classroom. In contrast to computer-based assessments, the GORT-5's pencil and paper method allows for direct observation of testing behaviors, along with reading patterns, which help during psychoeducational evaluations. Additionally, these observations can then help guide specific recommendations for students in the classrooms. Overall, the accuracy of the GORT-5 as a screener can greatly improve the ways school teams run the referral process through MTSS. This screener can help to increase the accuracy of the referral process while also guiding recommendations for the classroom for students who qualify for special education and those who do not.

References

Beers, K. (2003). When kids can't read: What teachers can do (2nd ed.). Portsmouth, NH: Heinemann.

Brown, V.L., Fishco, V.V., & Hanna, G.S. (1993). Nelson-Denny reading test. Austin, TX: PRO-ED.

Burns, M. K. (2016, January). Effect of cognitive processing assessments and interventions on academic outcomes: Can 200 studies be wrong? *National Association of School Psychologists*, *44*(5). Retrieved from http://www.nasponline.org/publications/periodicals/communique/issues/volume-44-issue-5/effect-of-cognitive-processing-assessments-and-interventions-on-academic-outcomes-can-200-studies-be-wrong.

Clemens, N. H., Hagan-Burke, S., Lou, W., Cerda, C., Blakely, A. Frosch, J… Jones, M. (2015). The predictive xalidity of a computer-adaptive assessment of kindergarten and first-grade reading skills. *School Psychology Review*, *44*(1), 76-97.

Decker, S.L., Englund, J., & Albritton, K. (2012). Integrating multi-tiered measurement outcomes for special education eligibility with sequential decision-making methodology. *Psychology in the Schools, 49(4).*

Ditkowsky, B., & Koonce, D. A. (2010). Predicting performance on high-stakes assessment for proficient students and students at-risk with oral reading fluency Growth. *Assessment for Effective Intervention*, *35*(3), 159-167.

Erford, B. T., O'Brocki, C. M., & Moore-Thomas, C. (2007). Technical analysis of the reading essentials skills screener-upper elementary version (Master's thesis). April.

Every Student Succeeds Act (ESSA) (2010, December 10). In *U.S Department of Education*. Retrieved from http://www.ed.gov/essa?src=rn.

Feifer, S.G., & Della Toffalo, D. (2007). Integrating RTI with cognitive neuropsychology: A scientific approach to reading. Middletown, MD: *School Neuropsych Press, 27.*

Flanagan, D. P., & Alfonso, V. C. (2011). Essentials of specific learning disability identification (pp. 1-23). Hoboken, NJ: John Wiley & Sons Inc.

Hall, A.H., & Tannebaum, R.P (2012). Test Review: J. L. Wiederholt & B. R. Bryant. (2012). Gray Oral Reading Tests—Fifth edition (GORT-5). Austin, TX: Pro-Ed.

Hammill, D.D., Wiederholt, J.L., & Allen, E.A. (2006). Test of Silent Contextual Reading Fluency. Austin, TX: PRO-ED.

Hopkins, W.G. (2002). A scale of magnitudes for effect sizes. In *A New View of Statistics.* Retrieved April 18, 2017, from http.//sportssci.org.resource.stats/effectmag/html.

Hunley, S. A., Davies, S. C., & Miller, C. R. (2012). The relationship between curriculum-based measures in oral reading fluency and high-stakes tests for seventh grade students. Research in Middle Level Education Online, *36*(5), 1-8.

IBM Corp. (2013). IBM Statistical Package of Social Sciences (SPSS) Statistics for Windows. 22.0. Armonk, NY:IBM Corp.

Individuals With Disabilities Education Act, 20 U.S.C. 1400 (2004)

Individuals With Disabilities Education Act, 34 C.F.R. 300. 301-300.311 (2006). Retrieved from http://www.ecfr.gov/cgi-bin/text-idx?tpl=/ecfrbrowse/Title34/34cfr300_main_02.tpl.

Jenkins, J.R., Hudson, R.F., Johnson, E.S. (2007). Screening for at-risk readers in a response to intervention framework. *School Psychology Review,* 36(4), 582-600.

Keenan, J.M., & Betjemann, R.S. (2006). Comprehending the gray oral reading test without reading it: Why comprehension tests should not include passage-independent items. *Scientific Studies of Reading,* 10, 363-380.

LaForte, E. M., McGrew, K. S., & Schrank, F. A. (2014). Woodcock Johnson IV assessment service bulletin number 2 (4thth ed.). Rolling Meadows, IL: Riverside http://www.assess.nelson.com/pdf/WJIVASB2.pdf.

Marchand, G.C., & Furrer, C.J. (2014). Formative, informative, and summative assessment: The relationship among curriculum-based measurement of reading, classroom engagement, and reading performance. *Psychology in the Schools, 51*(7), 659-676. Doi:10.1002/pits.21779.

Mather, N., & Wendling, B. J. (2015). Essentials of psychological assessment: Essentials of tests of achievement (1). Somerset, US: Wiley. Retrieved from http://www.ebrary.com.

McCray, A. D. (2001, November). Middle school students with reading disabilities. *The Reading Teacher*, *55*(3), 298-300.

McGlinchey, M. T., & Hixson, M. D. (2004). Using curriculum-based measurement to predict performance on state assessments in reading. *School Psychology Review*, *33*(2), 193-203.

Merrell, C., & Tymms, P. (2007). Identifying reading problems with computer-adaptive assessments. *Journal Of Computer Assisted Learning*, *23*(1), 27-35. doi:10.1111/j.1365-2729.2007.00196.x.

Merrell, K.W., Ervin, R.A., & Peacock, G.G. (2011). School psychology for the 21st century, second edition: Foundations and practices. The Guilford Press.

Miller, K.C., Bell, S.M., & McCallum, R.S. (2015). Using reading rate and comprehension CBM to predict high-stakes achievement. *Journal of Psychoeducational Assessment, 33*(8), 707-718.

National Center on Intensive Intervention. (n.d.). In *American Institutes for Research*. Retrieved from http://www.intensiveintervention.org/tools-chart-resources.

Nese, J. T., Park, B. J., Alonzo, J., & Tindal, G. (2011). Applied curriculum-based measurement as a predictor of high-stakes assessment: Implications for researchers and teachers. *The Elementary School Journal*, *111*(4), 608-624. doi:10.1086/659034.

Osewalt, G. (n.d.). Phonological awareness, phonemic awareness and phonics: What you need to know. In *Understood: For Learning and Attention Issues*. Retrieved April 1, 2017, from https://www.understood.org/en/learning-attention-issues/child-learning-disabilities/reading-issues/phonological-awareness-phonemic-awareness-and-phonics-what-you-need-to-know.

Patton, K.L.S., Reschly, A.L., & Appleton, J. (2014). Curriculum-based measurement as a predictor of performance on a state assessment: Diagnostic efficiency of local norms. *Educational Assessment, 19*(4) 284-301. doi:10.1080/10627197.2014.964117.

Prewett, P.N. (1995). A comparison of two screening tests (the Matrix Analogies Test-Short Form and the Kaufman Brief Intelligence Test) with the WISC-III. *American Psychological Association, Inc. 7(1), 69-72.*

Roehrig, A. D., Petscher, Y., Nettles, S. M., Hudson, R. F., & Torgesen, J. K. (2008). Accuracy of the DIBELS oral reading fluency measures for predicting third grade reading comprehension outcomes. *Journal of School Psychology*, *46*, 343-366.

Schrank, F.A., Mather, N., & McGrew, K.S. (2014a). Woodcock-Johnson IV Test of Achievement. Rolling Meadows, IL: Riverside.

Shinn, M.R., Walker, H.M. (2010). Interventions for achievement and behavior problems in a three-tier model including RTI. *National Association of School Psychologists.*

U.S. Department of Education. (2016). 38th Annual report to Congress on the implementation of the Individuals with Disabilities Education Act, 2016. In *Office of Special Education and Rehabilitative Services, Office of Special Education Programs*. Retrieved April 1, 2017, from https://www2.ed.gov/about/reports/annual/osep/2016/parts-b-c/38th-arc-for-idea.pdf.

VanDerHeyden, A. M., Witt, J. C., & Naquin, G. (2003). Development and validation of a process for screening referrals to special education. *School Psychology Review*, *32*(2), 204-227.

Wagner, R.K., Torgesen, J.K., Rashotte, C.A., & Pearson, N.A. (2010). Test of silent reading efficiency and comprehension. Austin, TX: PRO-ED.

West Virginia Board of Education Policy 2419: *Regulations for the Education of Students with Exceptionalities (2014)*. Retrieved from https://wvde.state.wv.us/osp/policy2419.html.

Wiederholt, J. L., Bryant, B. R. (2012a). Gray Oral Reading Test: Examiner's manual (5th ed.). Austin, TX: Pro-Ed.

Wiederholt J.L., Bryant B.R. (2012b). Gray Oral Reading Test: Examiner's record booklet; form-A (5th ed.). Austin, TX: Pro-Ed.

Wilkinson, G. S., & Robertson, G. J. (2017). Wide Range Achievement Test, Fifth Edition (WRAT5) (5th ed.). N.p.: Pearson.

APPENDIX A

Table 1
GORT-5 Coefficients Correlation to Other Criterion Tests

		GORT-5 Score				
Criterion Test	Score Used	Rate	Accuracy	Fluency	Comprehension	ORI
NDRT	Total Score	.77	.76	.78	.80	.81
NDRT	Comprehension	.71	.71	.72	.74	.76
NDRT	Vocabulary	.76	.79	.79	.85	.85
TOSCRF	Total Score	.79	.67	.79	.75	.81
TOSREC	Index Score	.76	.66	.75	.74	.79

Note. ORI= Oral Reading Index; NDRT= Nelson-Denny Reading Test (Brown, Fischo, & Hanna, 1993); TOSCRF= Test of Silent Contextual Reading Fluency (Hammill, Wiederholt, & Allen, 2006); TOSREC= Test of Silent Reading Efficiency and Comprehension (Wagner, Torgesen, Rashotte, & Pearson, 2010).

Table 2

Pearson Correlation

		GORT-5 Score				
Criterion Test Score Used		Rate	Accuracy	Fluency	Comprehension	ORI
WJ IV	Broad Reading	.78* (40)	.75* (40)	.83* (40)	.59* (40)	.81* (40)
	Basic Reading	.72* (39)	.82* (39)	.83* (39)	.58* (39)	.78* (39)
	Fluency	.76* (40)	.65* (40)	.76* (40)	.46** (40)	.70* (40)
	Comprehension Extended	.55* (39)	.66* (39)	.68* (39)	.69* (39)	.78* (39)

Pearson's r Correlation Between GORT-5 and WJ IV Reading Tests
Numbers in parentheses are N
*p<.001
**p <.01.

Table 3
Paired Samples Correlation

	N	Correlation	Significance
Pair 1: GORT-5 ORI Standard Score and WJ IV Broad Reading Standard Score	40	.809	.000
Pair 2: GORT-5 ORI Standard Score WJ IV Basic Reading Standard Score	39	.782	.000
Pair 3: GORT-5 ORI Standard Score WJ IV Reading Comprehension Extended Standard Score	39	.782	.000
Pair 4: GORT-5 ORI Standard Score WJ IV Reading Fluency Standard Score	40	.703	.000

Table 4
Paired Samples Statistics

	Mean	N	Standard Deviation	Std. Error Mean
Pair 1:				
GORT-5 ORI Standard Score	85.83	40	9.829	1.554
WJ IV Broad Reading Standard Score	82.25	40	13.210	2.105
Pair 2:				
GORT-5 ORI Standard Score	85.82	39	9.957	1.594
WJ IV Basic Reading Standard Score	89.97	39	17.521	2.806

Pair 3:				
GORT-5 ORI Standard Score	85.08	39	8.728	1.398
WJ-IV Reading Comprehension Ext Standard Score	82.23	39	10.564	1.692
Pair 4:				
GORT-5 ORI Standard Score	85.83	40	9.829	1.554
WJ IV Reading Fluency Standard Score	84.00	40	12.700	2.008

Table 5
Paired Samples T-test continued

	95% Confidence Interval of the Difference			
	Lower	Upper	t	Sig. (2-tailed)
Pair 1:				
GORT-5 ORI Standard Score WJ IV Broad Reading Standard Score	1.057	6.093	2.872	.007
Pair 2:				
GORT-5 ORI Standard Score WJ IV Basic Reading Standard Score	-7.895	-.413	-2.248	.030
Pair 3:				
GORT-5 ORI Standard Score WJ-IV Reading Comprehension Ext Standard Score	.707	4.985	2.693	.010
Pair 4:				
GORT-5 ORI Standard Score WJ IV Reading Fluency Standard Score	-1.079	4.729	1.271	.211

Table 6

GORT-5 ORI and WJ IV Broad Reading Crosstabulation

	N		WJ IV Broad Reading		Total
			1.00	2.00	
GORT-5 ORI	1.00	Count	12	3	15
		Expected Count	5.1	9.9	15.0
		Adjusted Residual	4.7	-4.7	
	2.00	Count	2	24	26
		Expected Count	8.9	17.1	26.0
		Adjusted Residual	-4.7	4.7	
Total		Count	14	27	41
		Expected Count	14.0	27.0	41.0

Table 7

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	22.117[a]	1	.000		
Continuity Correction[b]	19.018	1	.000		
Likelihood Ratio	23.530	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	21.578	1	.000		
N of Valid Cases	41				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 5.12.

b. Computed only for a 2x2 table

APPENDIX B

MARSHALL UNIVERSITY
www.marshall.edu

Office of Research Integrity

April 18, 2017

Megan Edwards
School Psychology Program
Marshall University

Dear Ms. Edwards:

This letter is in response to the submitted thesis abstract entitled *"The Relationship Between the GORT-5 and WJIV ACH for Referred Children."* After assessing the abstract it has been deemed not to be human subject research and therefore exempt from oversight of the Marshall University Institutional Review Board (IRB). The Code of Federal Regulations (45CFR46) has set forth the criteria utilized in making this determination. Since the information in this study does not involve human subjects as defined in the above referenced instruction it is not considered human subject research. If there are any changes to the abstract you provided then you would need to resubmit that information to the Office of Research Integrity for review and a determination.

I appreciate your willingness to submit the abstract for determination. Please feel free to contact the Office of Research Integrity if you have any questions regarding future protocols that may require IRB review.

Sincerely,

Bruce F. Day, ThD, CIP
Director