Marshall University
Marshall Digital Scholar

Theses, Dissertations and Capstones

2018

Relationship between the WJ-IV ACH reading tests and the GORT-5

Emily D. Nestor
nestor16@marshall.edu

RELATIONSHIP BETWEEN THE WJ-IV ACH READING TESTS AND THE GORT-5

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
Emily D. Nestor
Approved by
Dr. Lanai Jennings, Committee Chairperson
Dr. Sandra Stroebel
Dr. Conrae Lucas-Adkins

Marshall University
August 2018

APPROVAL OF THESIS

We, the faculty supervising the work of Emily D. Nestor, affirm that the thesis, the Relationship Between the WJ-IV Achievement Reading Tests and the GORT-5, meets the high academic standards for original scholarship and creative work established by the School Psychology Program and the College of Education and Professional Development. 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 7-25-2018
Dr. Sandra Stroebel, Department of School Psychology	Committee Member	Date 7-25-2018
Dr. Conrae Lucas, Department of School Psychology	Committee Member	Date 7-25-2018

ACKNOWLEDGMENTS

I would first like to thank the Woodcock Institute for the Advancement of Neurocognitive Research and Applied Practice for their grant support throughout this study. With their support and interest in research, additional materials were able to be purchased to be used throughout. I would also like to thank the diagnosticians and school psychologists at Jackson County Schools, who dedicated their time to assessing countless students to be used in the study. I would also like to thank Dr. Lanai Jennings for the passion and dedication she shares in everything that she does, especially regarding the study; without her relentless commitment it would not have been possible. I would also like to thank my committee members Dr. Sandra Stroebel and Dr. Conrae Lucas-Adkins, who also dedicated their time to make this thesis successful. Finally, I would not have been able to complete this process without the continuous support and love from my family and friends.

TABLE OF CONTENTS

LIST OF TABLES

ABSTRACT

The present study examined the relationship between the Gray Oral Reading Test – Fifth Edition and the Woodcock Johnson-IV Achievement Tests in reading in a sample of 104 school age participants between the ages of 7 and 18. Pearson correlations tests indicated large correlations (*r=.87, p=.01)* between the GORT-5 ORI and the WJ IV ACH Broad Reading cluster. Additional comparisons for fluency and comprehension yielded comparable results (*r=.85, p=.01; r=.84, p=.01)*. Similarly, a Fisher's Exact Test illustrated that the odds of scoring at or below the tenth percentile on the WJ IV ACH was very high when a student scored at or below the tenth percentile on the GORT-5. The findings suggest that the use of this relatively brief oral reading measure can be beneficial as a diagnostic assessment in a district's MTSS because its results are closely aligned to the WJ IV –ACH, a popular achievement test frequently used in combination with other assessments to help determine if struggling readers have academic deficits significant enough to warrant special education services.

CHAPTER ONE

Literature Review

According to the Cortiella & Horowitz (2014), approximately five percent of the public school population is identified as having a learning disability and an additional 15% or more are struggling academically and may have unidentified or unaddressed learning and attention difficulties. As a result, schools throughout the United States have a variety of strategies and programs to address the needs of struggling readers. To accompany these reading supports, schools have a variety of universal screening, progress monitoring, and diagnostic instruments in reading at their disposal for implementation of Multi-Tiered Systems of Support (MTSS).

The selection of the MTSS assessment instruments by school and district leaders is contingent upon a number of variables including financial resources and technology access (Eagle, Dowd-Eagle, Snyder, & Holtzman, 2015). Ease of administration and scoring is another key consideration given the demands placed on already overburdened teachers and interventionists. For this reason according to Eagle et al. (2015), computer-based assessments, which afford class-wide reading screening and progress monitoring opportunities in under an hour, are becoming increasingly popular.

Despite the shift to commercially developed computer-based progress monitoring systems (Greif, Martin, & Spinath, 2014), computer-based assessments pose certain limitations. They, for example, can be problematic for students who have attention deficits or a tendency to randomly respond to test items during group administration (Clemens, et. al., 2015). Computer-based assessments, furthermore, have less utility in assessing progress for young children who rapidly respond to test items, fail to recheck answers, and ultimately sacrifice accuracy in favor of speed. Computer-based assessments, consequently, can yield a series of highly variable data points for certain subgroups of examinees. The inconsistent data can be difficult for educators to interpret in a MTSS framework and subsequently recommend appropriate interventions and special education referrals (Clemens, et al., 2015). Such limitations are likely to result in inappropriate reading referrals, which can be particularly costly since the majority of students who are referred for special education and later qualify under the specific learning disability category have reading difficulties (Flanagan & Alfonso, 2011).

The current project first originated from difficulties school psychologists and other diagnostic staff experienced in one West Virginia district while attempting to interpret the results of one computer-based reading assessment, STAR Reading, for a subgroup of students with significant variability and/or borderline scores on benchmark and/or progress monitoring data. In response to the lack of clear data for students whose Student Assistance Team (SAT) pressed for a special education referral, the school psychologists or diagnostic staff administered the Gray Oral Reading Test – Fifth Edition (GORT-5), a widely used test of oral reading (Wiederholt & Bryant, 2012). This test enabled the examiner to directly assess both oral reading fluency and comprehension of the students in question and proved to be a relatively quick and useful tool in affirming or disconfirming the need to pursue a special education referral.

Although the GORT-5 proved useful for data-based decision making at the individual student level for a group of referred students and in one district (Edwards, 2017), the purpose of the current investigation is to determine if the GORT-5 has 1) strong criterion validity when compared to the Woodcock Johnson Tests of Achievement

– Fourth Edition (WJ-IV ACH) and 2) broad utility as a secondary, diagnostic screener to accurately identify and separate students who potentially qualify for a specific learning disability in reading for a more diverse group of participants. However, prior to outlining the current study, it is first important to provide an overview of the Individuals with Disabilities Education Act (IDEA) of 2004 requirements with respect to reading assessment for specific learning disabilities identification in a Response to Intervention (RTI) or MTSS approach. This discussion will entail the different types of reading assessments including universal screenings and curriculum-based measurements, as well as more diagnostic instruments like the GORT-5 and WJ IV ACH.

Individuals with Disabilities Education Act

In 2004, the United States Department of Education reauthorized the IDEA 2004. IDEA 2004 included many changes for the nation's school systems; however, those related to the identification of students with specific learning disabilities (SLD) were some of the most extensive. One of the main areas of emphasis of IDEA 2004 was a shift to providing intense instruction and intervention supports to rule out the lack of instruction as a factor in the students' academic difficulties. For states and districts opting to use RTI as an identification method, they were now mandated to provide increasingly sustained and intensive scientifically-based reading instruction and interventions for struggling readers suspected of having reading disabilities. Students' responsiveness to these interventions were to be measured over a period of time to assess efficacy. In reaction to the paradigm shift, many districts adopted RTI for SLD identification although policy implementation of data collection procedures, time in intervention, etc. varied significantly by state (Hudson & McKenzie, 2016) and from district-to-district.

When reauthorized, IDEA 2004 expressly indicated that states could not preclude districts from implementing RTI models, but could opt to eliminate "ability-achievement discrepancy model" as a means to identify students as having a SLD. Out of fifty states, 11 prohibited the use of the discrepancy model (Schultz & Stephens, 2009). The discrepancy model was the most frequently used SLD method prior to the advent of RTI in IDEA 2004. It primarily utilized a student's intellectual ability and achievement scores for a comparison. The intellectual component according to Kavale (2005), required a student to fall in the "normal" (above 80) or "average" (90-110) range.

Decker, Hale & Flanagan (2013) contend that the IQ-achievement discrepancy model alone is an invalid approach to determining SLD, and led to an over identification of SLD because it did not discriminate between low achieving students and students who, in fact, actually have SLDs. In order to determine if a child can be identified as having a SLD, according to the discrepancy model, their IQ and achievement scores must be considered significantly discrepant. A significant difference implies that if the student's IQ score is generally two standard deviations more than the score from one of the eight areas on the achievement assessments that a student could be found eligible for services (Reschly & Hosp, 2004). While two standard deviations is a typical significant difference, each state reserved the right to set the criteria for meeting SLD requirements; some used only one and one-half standard deviations (Restori, Katz, & Lee, 2009). In addition to different standard deviation thresholds, discrepancy models can be operationalized through a variety of methods. For example, Peterson & Shinn (2002) discuss the three different approaches to the discrepancy models: Intra-Individual Achievement Discrepancy (IAD), Absolute Achievement Discrepancy (AAD), and

Relative Achievement Discrepancy (RAD). The IAD approach utilizes a regression approach to determine the variability and the magnitude of the discrepancy to determine the severity. The AAD approach is based on a universal distribution and the discrepancy between a student's actual achievements on a norm-referenced test, in comparison to the national average. The RAD approach examines not only their performance on a standardized assessment, but also achievement discrepancies within their classroom performance and school district. For example, the lowest performing students in a high-achieving school district were identified as SLD.

Proponents of Response to Intervention (RTI) as an SLD identification method have identified three main concerns with the IQ-achievement discrepancy model. Speece (2002) noted the first problem with the discrepancy model is that it represents a "wait-to-fail" approach. It fails to identify students in primary grade levels as having a SLD as they do not meet the eligibility requirements until the gap within their achievement widens (Restori, Katz, & Lee, 2009). Supported by Fuchs, Fuchs, & Zumeta (2008), the identification of students with a SLD is delayed until grades 3-5 when using this model. Another major criticism of this model is students with below average intellectual abilities and low achievement scores experience long term problems without ever receiving special education services because they do not meet the discrepancy gap between IQ and achievement, despite clearly being unable to perform grade-level tasks (Fuchs et al., 2008). Finally, the third criticism of this model is the disproportionality of students with diverse backgrounds, including: race/ethnicity, language, and socioeconomic status (Shifrer, Muller, & Callahan, 2011). Skiba et al. (2008) note significant variability in a diagnostic model that is regression based, versus a standard score, of students of racial or ethnic minorities, who are more likely to attend high poverty schools.

As outlined in IDEA 2004, a student's eligibility determination meeting is comprised of a multi-disciplinary team of professionals (i.e. school psychologist, teacher, principal, speech and language pathologist, etc.) and the child's parent. The student must demonstrate a low level of performance with respect to same-age peers and be making insufficient progress or rate of learning after the most stringent interventions have been implemented. IDEA 2004 maintained some components of the original act such as the exclusionary factors. For example, for the student to be classified as having SLD, the youth's learning difficulties must not be due primarily to visual, hearing or motor disabilities, cultural or environmental factors, economic disadvantage, or intellectual disabilities. The parent or guardian of the student must also be provided with any evidence of assessments of their child's achievement throughout the intervention process and must be made aware of their right to request an evaluation of their child under IDEA. The United States Department of Education explicitly states that a child's determination for special education cannot rely on any single procedure and must include a variety of assessments (IDEA, 2004).

Response to Intervention

Under IDEA (2004), states no longer were required to use a discrepancy model for identifying students with disabilities. The shift introduced RTI as a new method of SLD identification. RTI is an integrative approach to assessment and reading intervention, with an overall goal of maximizing student achievement and maintaining a structured approach to the systematic design of the classroom (National Center on

Response To Intervention, 2010). RTI is primarily used to identify students who are "at-risk." An "at-risk" student has a high probability of not meeting the designated learning outcomes for his or her grade level. RTI is also used to monitor students' progress and provide appropriate instruction to students based on their specific needs (National Center on Response To Intervention, 2010). Following the passage of IDEA 2004, RTI became the leading alternative to the discrepancy model in SLD identification.

Instruction and Intervention Tiers

RTI is a multi-tiered intervention model that is typically composed of three primary tiers of intervention and assessment. The three levels of instruction and intervention provide a framework for increasing intensity in academic support and monitoring of progress for students as they move from tier one to three. While frequency and intensity of intervention and progress monitoring are typically increased as a student moves from the core or primary level of support to the secondary and tertiary levels, group size, conversely, decreases.

Tier one is the universal level of RTI and is implemented within the whole-group classroom setting. The curriculum within tier one is to implement core curriculum with the goal of the majority of the class reaching proficiency. Tier one is structured to function from system-wide data; if the majority of students are not making adequate progress then a system-wide change needs to be made within the curriculum. Fuchs. Fuchs, and Zumeta (2008) indicated that approximately 80-85% of students solely receive instruction from this primary tier; which is often accomplished within the classroom utilizing differentiated instruction strategies, such as small groups. Within the universal level of RTI continuous assessment and progress monitoring is used to identify students who may need more interventions. Assessments given in tier one typically occur at three different times during the school year (September, January, May) and are administered in the general education classroom. These assessments are used to determine the effectiveness of the curriculum and channel students to additional intervention when insufficient progress is evident (Della Toffalo & Feifer, 2007).

Tier two is the next level in the RTI model. Tier two is comprised of students who have been identified as "at-risk" in tier one and are known as the targeted group (Berkeley, Bender, Peaster, & Saunders, 2009). A student who is eligible for tier two is performing in the bottom 25 percent of the class. Tier two is designed to include approximately 20-30 percent of students (Della Toffalo & Feifer, 2007). Tier two consists of small group instruction in addition to the classroom curriculum that students are already receiving. The small groups in tier two are designed to have three to five students in them working with more intense interventions on the specific skill deficits. In tier two, a reading or intervention specialist typically delivers interventions, although schools may alternatively rely on general educators and flexible school scheduling to deliver the secondary intervention. Tier two students also receive more frequent assessments. Typically progress-monitoring assessments are given every one to four weeks.

Tier three is the most intense level of interventions and is comprised of students who did not make adequate progress in tier two, thus demonstrating the need for more sustained interventions. Tier three cannot typically be adequately implemented in a general education classroom and often requires support from special education teachers or reading interventionists and is an intensive and individualized intervention specific to students (Berkeley, et. al., 2009). Tier three is designed to address significant academic needs, often without the student being enrolled in special education. In tier three instructional time is increased significantly and progress monitoring occurs weekly. Ideally less than five percent of the student population requires tier three (Della Toffalo & Feifer, 2007) although this percentage can vary substantially by location.

Components of An Effective RTI Model. A multi-tiered system of support includes four primary components. These include: evidence-based curricular and instructional practices for all students; a data based framework used to make decisions; use of a problem solving system across all levels of the support system; and a structural team approach to planning, evaluating, and implementing interventions. Within the multi-tiered system of supports, RTI provides a data based framework to make decisions regarding academic performance. Within the RTI model, there are six critical components for it to be the most effective. These components are: universal screening, baseline data, measurable terms, accountability plan, progress monitoring plan and data based decision-making (Freeman, Miller, & Newcomer, 2015).

Universal screening as described above occurs with the entire student population, at least three times per year and is used to benchmark all students. Baseline data is to be collected for all students by assessing their present levels and to monitor students' responses to instruction and growth over time. Measurable terms are used to clearly define the academic problem areas for individual students. The next three components involve planning. The first is an accountability plan. An accountability plan is developed based on the needs of specific students and considers the ways that student's progress will be monitored, as well as specific details to the interventions like duration and intensity. The next type of plan is the progress-monitoring plan; it is a predetermined plan for evaluating intervention efficacy. It is single-case design in nature and includes how frequently progress monitoring will occur and what type of progress monitoring data will be recorded. The final component of effective RTI is data based decision-making. Data-based decision making requires the school team use the continuous evaluation data to directly inform future classroom instruction, intervention, and the movement of students between tiers. Consequently, data based decision making also involves the initial identification of at-risk students for appropriately matched interventions (Della Toffalo & Feifer, 2007).

When students are screened at the beginning of the school year to assess their basic skills, student performance below a percentile rank of 25 indicates some need for more intense instruction (Hudson & McKenzie, 2016). The at-risk students aren't responding satisfactorily to their present level of instruction and are, therefore, placed in the second tier (Berkeley et al., 2009). Finally, students who do not show adequate progress in tier two, are moved to the third tier. Approximately 5% of the student population receives tier three services, in which they receive more individualized instruction and more frequent progress monitoring. Students who despite sustained, intensive intervention continue to demonstrate insufficient progress are considered non-responsive (NR) by Fuchs & Deschler (2007). The study completed by Hudson & McKenzie (2016) revealed that over 90% of states do not regulate or recommend a specific number of days that a NR student should spend in the different RTI tiers prior to a referral to special education.

Types of Assessment in RTI/MTSS

Curriculum-based measurement. Curriculum-based measurements (CBM) are integral to measuring student progress in an MTSS framework. According to Patton, Reschly and Appleton (2014), CBM is a brief tool of standardized measurement used to screen student performance in many different academic areas such as reading, mathematics, written expression and spelling. More specifically, Deno (2003) describes CBM as “the practice of using what is learned as the basis for assessing what has been learned” (p. 5). When repeated over time, these brief assessments, which are directly tied to a student’s curriculum, aid in the evaluation of the effectiveness of instruction. CBM is a vital component of the overall MTSS, as RTI puts an emphasis on educators collecting multiple data points to gauge student progress throughout the year, specifically for the referral of students to special education services. This demand is met by the use of CBM. CBM is used in the classroom to measure present levels of performance and as an early screening; identify potential “at-risk” students; and predict academic progress and achievement on statewide assessments (Miller, Bell, & McCallum, 2015). CBM has many benefits for use in the schools. These include the abundance of information that can be obtained from the measures, the brief amount of time that they take to administer, score and interpret, and the cost efficiency of them to the schools (Miller et al., 2015). CBM provides a variety of alternate test forms, which allows for repeated sampling (Fuchs, Fuchs, Hosp & Jenkins, 2001). Not only does CBM allow teachers to generate quantitative data, but they are also able to generate qualitative descriptions of performance for each student (Fuchs, et al., 2001). Typically, a struggling student is in the lowest 10 to 25 percent of the class (Miller et al., 2015). Within the RTI model, CBM is used most commonly in the reading setting. The Dynamic Indicators of Basic Early Literacy Skills (DIBELS) is one of the most frequently used CBMs (Miller et al., 2015).

Oral reading fluency (ORF), which is universally assessed through the use of reading CBM, is considered to be a “vital sign” of student achievement (Miller et al, 2015). Fuchs, Fuchs, Hosp & Jenkins (2001) describe ORF as a measure of phonological segmentation, recoding skills, and rapid word recognition. ORF is related to a student’s overall reading proficiency and assesses the students’ ability to accurately and fluently read a grade-level appropriate passage and can be assessed in isolation via word lists or text (Fuchs, et al., 2001). Jenkins, Fuchs, Espin, van den Broek and Deno (2000) completed a study examining the differences between reading words in context or isolation. In their study, 113 total students ranging from skilled readers to readers with disabilities read two different measures, a 400-word folk tale and a list of randomly ordered words from the folktale in randomly assigned groups. They also completed the reading comprehension portion of the Iowa Test of Basic Skills. The criterion validity coefficient for the text fluency was .83 and .53 for the list fluency. ORF because of its significant implications has become the most commonly assessed CBM and can accurately assess reading proficiency (Fuchs et al., 2001; Miller et al., 2015).

Because reading fluency implies the ability to read with automaticity, comprehension is considered to be related to a student’s reading fluency. Reading comprehension CBM can take the form of many different types of assessment. In a study completed by Fuchs, Fuchs and Maxwell (1988), they examined three different measures to directly measure reading comprehension, by using the reading comprehension subtests of the Stanford Achievement Test. Question answering is the first type of reading comprehension assessment and is most commonly used in classrooms. Students are typically asked to read passages and then answer short questions about what they have read. The second type of comprehension measure is passage recall. Passage recall requires students to read and then retell the story with limited prompts. The third direct measure is cloze. Cloze assessments require a reasoning process to generate appropriate words to complete the sentences. Seventy students with reading disabilities were given all three alternative measures to measure reading comprehension. Results indicated criterion validity coefficients for the three: question answering (.82), the recall (.70), and the cloze measure (.72).

Other Types of Progress Monitoring Instruments in MTSS. In addition to CBMs, computer based assessments, also known as e-assessments, are another form of progress monitoring tool that is used in the classrooms. E-assessments, for example STAR Reading, were developed to evaluate student progress more efficiently. In order for this to be done, the system should be able to accurately identify the student's actual reading skills and their progress, as well as create an accurate prediction of their achievement (Greif, Martin & Spinath 2014). Computer based assessments have both positive and negative characteristics associated with them.

Positive characteristics of e-assessments include the opportunity to be individualized. With computer-based assessments, tests can be adapted to meet the specific ability levels of the students. In addition, test items have the ability to adapt based on student responses. If a student is answering questions correctly then questions will get progressively more difficult and conversely, if they are missing questions then items will get easier. Another advantage of e-assessments is that it provides immediate feedback from student's assessments. For example, programs can provide item analysis, like how long students spent answering questions. Finally, computer based assessments provide a more standardized administration of assessments (Blazer, 2010).

Typical CBMs focus on measuring oral reading fluency; however, with the e-assessment this component is lost when a student is completing an assessment with a computer or other piece of technology. In addition, there is a significant cost associated with using e-assessments in order to have access for all students and access to the technology. Training the staff in computerized administration is also a limitation. There are many more components associated with the use of computers for assessment than the standard paper and pencil assessments. Finally, scoring open-ended questions is more difficult when interpreted by a computer, as opposed to an individual scorer (Blazer, 2010).

Use of Norm-Referenced Diagnostic Screeners in MTSS

Empirically valid screening tools, which require measures to have strong psychometric properties, are a critical assessment component within a MTSS model. Many studies have examined the specific relationship between oral reading fluency and norm referenced assessments. For example, research conducted by Reschly, Busch, Betts, Deno & Long (2009), found moderate correlations for oral reading fluency (ORF) with norm-referenced tests (r= .60-.70). Additionally, Wayman, Wallace, Wiley, Ticha and Espin (2007) found high reliability coefficients on ORF. In further studies, conclusions have been drawn to include ORF data in universal screening (Kilgus, Methe, Maggin, & Tomasula, 2014). Klingbeil, McComas, Burns & Helman (2015) examined different screening measures and their ability to predict students' future performance. They administered the Measures of Academic Progress-Reading (MAP) and the ORF probes from AIMSweb. The Fountas & Pinnell Benchmark Assessment System (BAS) was used as the formative assessment through the tiers of MTSS. The correlations between the predictors were moderate to strong (*r=.52 to r=.70).*

Fuchs, Fuchs, & Compton (2011) discuss a two stage screening process which utilized the Word Identification and Word Attack subtests of the Woodcock Reading Mastery Tests in the first stage. During these stages, 485 children were assessed in the fall of first grade and completed subtests to obtain a composite score that included both timed and untimed performance on the subtests. Results illustrated that measuring a response to classroom instruction with dynamic assessment reduced false positives, whereas using running records did not reduce the number of false positives.

Tucker & Jones (2010), used the GORT-4 in a study to measure the effects of reading instruction in the general education classroom versus instruction with supplementary intervention over a 10-week period. All students were given Form A of the GORT-4 as a pretest to assess fluency, rate and accuracy. At the end of the 10 weeks, the GORT-4 Form B was administered to again measure rate, fluency and accuracy. An independent measures t test was used to compare the mean differences of pretest and posttest scores in the areas of rate, accuracy, and fluency between the experimental and control group. Results yielded t scores in the critical region for all three areas, indicating that the scores of the experimental group are significantly higher than the score of the control group. In a previous study completed by Edwards (2017), 41 students referred for special education evaluations were administered the GORT-5 and the WJ-IV Tests of Achievement by trained school psychologists and diagnosticians to determine if the

GORT-5 could be used as a diagnostic screener. Of the 41 students, 19 were female and 22 were male. Most students were White, non-Hispanic (90.2%). All assessments were completed within the state mandated 80-day timeline. Both assessments were given to all participants. Results of this study indicate a large, positive correlation between the GORT-5 ORI and the WJ IV ACH Broad Reading Cluster (r =.81). Individual scaled scores form the GORT-5 also had large to very large correlation coefficients when aligned with the corresponding WJ IV ACH cluster scores. Additionally, the study examined the relationship between the ORI standard score and the Broad Reading cluster standard score. Results found that on average the GORT-5 ORI score was 3.58 points higher than the WJ IV Broach Reading score. Similar results were found in the areas of comprehension and fluency. The study also examined the number of students who performed at or below the 10^{th} percentile on the ORI and their performance on the WJ Broad Reading cluster, using a Fischer's exact test and crosstabulations. The results yielded: twelve students (29.7%) at or below the tenth percentile on both assessments (true positives), twenty-four (58.5%) scored above the tenth percentile on both assessments (true negatives), three (7.3%) scored at or below on the GORT-5 but above on the WJ IV which indicates that the GORT would give a false positive, and two (4.9%) students scored at or above on the GORT-5, but at or below on the WJ ACH illustrating a false negative.

Need for Study

The purpose of the current research was to evaluate the concurrent criterion validity of the GORT-5 and the WJ-IV ACH reading tests. Understanding the relationship between the two assessments is essential to determine if the GORT-5 could be used as an effective tier three diagnostic instrument, one which can help educators determine which students do, in fact, need referred for a comprehensive special education evaluation due to a suspected reading disability. The present study was necessary to extend the work of Edwards (2017) to include a larger and more diverse sample of participants with and without disabilities. In addition, this study reduced the number of days between test administrations to a maximum of thirty days, tightening the testing window. Finally, aside from Edwards (2017) study, an exhaustive search of scholarly literature revealed no comparisons between the two instruments to date.

Research Questions

Research Question 1:

What is the correlation between the GORT-5 ORI, Fluency and Comprehension *scaled scores* and the WJ-IV ACH Broad Reading, Fluency and Reading Comprehension scores? *The investigator predicts there will be very large, positive correlations given the similarity in constructs measured.*

Research Question 2:

Does GORT-5 ORI reading performance at or below the tenth percentile, predict WJ-IV ACH performance at or below the tenth percentile for cluster scores commonly used in eligibility determinations: Reading Comprehension, Basic Reading, and Reading Fluency? *The investigator predicts GORT-5 performance at or below the tenth percentile can sufficiently predict performance on the select WJ-IV cluster scores, although significant levels will vary by pairings.*

CHAPTER TWO

Method

Participants

A total of 104 school age children participated in this study. All participants were enrolled in public elementary, middle, and high schools in north central and southern West Virginia, as well as southeastern Ohio. Thirty-seven participants (35.6%) were females and 67 (64.4%) were male. Seventy-four (71.2%) participants were ages 7 through 12, whereas 30 participants (28.8%) were ages 13-18. Participants, representing 22 schools, attended grades two through eleven. The highest number of participants were in fifth (18 students), second (17), and sixth grades (15). All participants were English language proficient. Ninety-five (91.3%) of participants were white, non-Hispanic. The remaining students identified as Asian (1; 1.0%), Black (1; 1.0%), Hispanic (2; 1.9%); Multiple Races (5; 4.8%). The majority of the students were referred for psychoeducational evaluations or re-evaluations by their schools due to learning, social-emotional, behavior difficulties or suspected giftedness. However, some children without disabilities volunteered to participate with written informed consent of their parents (see Appendix A for Human Subjects Interim Review Board Permissions). Participants outside of the school system were recruited through IRB approved fliers posted in community libraries. In all, the sample was comprised of 84 (80.8%) students with IEPs and 20 (19.2%) students with no identified disabilities under IDEA 2004.

Materials/Apparatus

Gray Oral Reading Tests – Fifth Edition. The most recent edition of the Gray Oral Reading Test is the fifth edition that was published in 2012. The test is normed for students ranging from 6 years 0 months to 23 years 11 months and assesses reading rate, accuracy, fluency and comprehension. The GORT-5 was normed with a sample of 2,556 students in 33 different states, representing the four major U.S. geographic regions. The average size for each tested age level was 147 and aligned with national expectations for gender, ethnicity and geographic region (Hall & Tannebaum, 2012).

The GORT-5 is composed of sixteen reading passages that can be administered to students. Each passage increases with difficulty and contains comprehension questions at the completion of each one. In addition to the comprehension component, the examiner also records the total time spent reading, any omissions or substitutions from the passage, and prosody. The GORT-5 is intended to be given to individual students and administration time can take anywhere from 15-45 minutes. Weiderholt and Bryant (2012), recommend that the examiner predetermine the student's entry passage based on their grade level or prior knowledge of the student's reading ability and then continue testing until the ceiling is met. After the test has been given, the raw scores are recorded for the four subscales. The four subscales include rate, accuracy, fluency and comprehension. Using these four raw scores their Reading Oral Index (ORI) is calculated. Raw scores are translated into scaled scores for each area and the ORI standard score, allowing student performance to be interpreted in comparison to same age peers. The five normative scores produced are: grade and age equivalents, percentile ranks, scaled scores and the ORI (Mullis, 2012). Score ranges include a scale of very poor to very superior. An index score less than 70 is at the bottom of the scale, very poor, and ranges through poor (70-79), below average (80-89), average (90-110), above average (111-120), superior (121-129), and very superior (greater than 130).

According to Wiederholt & Bryant (2012), multiple studies were completed to assess the reliability of the GORT-5, including test-retest, alternate forms (delayed administration) and interscorer reliability. The test-retest studies for immediate administration and delayed administration both produced correlation coefficients that correlate to high reliability (.82-.90). Different studies were completed to assess the interscorer reliability using multiple examiners; all studies produced correlations of .99 or larger indicating that there is a high level of interscorer agreement, concluding that results produced by the GORT-5 can be obtained with a high level of confidence. The manual includes strong evidence for content, construct, and criterion-related validity. Content validity was established by linking the formatting and scoring procedures of the GORT-5 to other reading tests, such as the Gilmore Oral Reading Test. In order to ensure content validity passages focus on topics that would eliminate bias. Construct validity was developed using a three-step procedure. This procedure included: constructs expected for test performance, a set of hypotheses based on the constructs and using empirical methods to verify the hypotheses. A variety of relationships between the GORT-5 and performance were examined and expected patterns were found in all except correlating abilities to secondary students. Criterion-related validity was established by comparing the GORT-5 to five previously developed reading tests as illustrated in Table 1, like the Nelson-Denny Reading Test (NDRT), the Reading Observation Scale (ROS), the Test of Silent Contextual Reading Fluency (TOSCRF), and the Test of Silent Word Reading Fluency (TOSWRF) (Hall & Tannebaum, 2012). Conclusive evidence of validity was established with these tests and the abilities that they measure with average correlation coefficients of .68-.77, as illustrated in Table 1 (Wiederholt & Bryant, 2012; Hall & Tannebaum, 2012).

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).

Woodcock Johnson IV – Achievement. The Woodcock Johnson IV- Achievement is comprised of two different batteries of assessment: the standard battery includes tests 1-11 and the extended battery, which includes tests 12-20. Within the Woodcock Johnson IV, the various subtests examine the following areas of reading emphasis: reading, broad reading, basic reading skills, reading comprehension, reading fluency, and reading rate.

With the revision of the Woodcock Johnson IV, new subtests were created. These new subtests include: Verbal Attention, Letter-Pattern Matching, Phonological Processing, Nonword Repetition, Segmentation, Oral Reading, Reading Recall, and Word Reading Fluency. The new subtests have a significant impact on the assessment of reading and a more focused approach to assessing reading. Phonological Processing is one of the newest subtests and was specifically developed to examine phonological processing abilities (McGrew, LaForte, & Schrank, 2014). In order to obtain a complete assessment of phonological processing, the Woodcock Johnson IV developed three subtests to completely create a phonological processing score. The three subtests include: word access, word fluency, and substitution. In addition to Phonological Processing, the Woodcock Johnson IV recently added Oral Reading and Reading Recall that directly correlates to Response to Intervention and a more classroom comparative approach to reading fluency. The revision of the Woodcock Johnson IV is the most authentic reading assessment because it assesses the students' accuracy and fluency and then the recall.

The Woodcock Johnson IV implemented basic principles during the development of the test, to ensure that this assessment would eliminate any bias against individuals with any specific needs or disabilities. The Woodcock Johnson IV employs selective testing procedures which allows students a flexible experience, by the administrator being able to choose the order of the subtests. The flexibility of choosing the order of subtest administration is beneficial to an examiner when knowing the student and their strengths or weaknesses and can keep the student from becoming frustrated by letting them take frequent breaks or shifting tasks.

The Woodcock Johnson IV produces two different types of score reports: variations and comparisons. Variations are the more descriptive of the two score reports. The pattern of strengths and weakness, PSW, is the variation report generated from the results of the students' assessments. The PSW indicates the students' areas of potential strengths and then their areas of weakness. The comparison report that the score report generates is used to make a hypothesis about the student's performance and predict future performance. Each cluster of the WJ IV ACH yielded a median reliability coefficient of

.90 or higher. According to McGrew, LaForte and Schrank (2014), the WJ IV ACH Reading Cluster correlates to the Kaufman Test of Educational-Achievement-Second Edition (KTEA-II) and the Wechsler Individual Achievement Test-Third Edition (WIAT-III). The correlation coefficients ranged from .78 to .91.

Procedure

Trained examiners including school psychologists, school psychology interns, diagnosticians, and special educators administered the WJ-IV and GORT-5 to participants within school settings. Counterbalanced order was implemented to the greatest extent possible and executed within each examiner's test settings. However, due to multiple examiners and time between assessments perfect adherence to counterbalancing was not always possible. When feasible the GORT-5 and the WJ-IV ACH reading tests were administered in the same test session. However, in most cases the two instruments were administered over a series of two sessions. No more than thirty days between administrations was deemed acceptable for the purpose of the study.

Participants were given the GORT-5 and WJ IV ACH reading subtests under standard conditions utilizing the specific start points, ceiling and basal rules as outlined by their manuals. Following the completion of the assessments, the GORT-5 scores were totaled and hand scored to obtain scaled scores for rate, accuracy, fluency and comprehension. The scaled fluency and comprehension scores were used to obtain the ORI standard score. Each participant completed the following eight subtests of the WJ IV ACH: Letter-Word Identification, Passage Comprehension, Word Attack, Oral Reading, Sentence Reading Fluency, Reading Recall, Word Reading Fluency, and Reading

Vocabulary. The WJ IV ACH subtests were totaled and then entered using the online program: wjscore.com to obtain cluster scores.

Data Analysis

A comprehensive list of data was maintained in a Microsoft Excel worksheet with no identifying information by a study investigator. The data was transferred into the IBM Statistical Package of Social Sciences (SPSS) software by the investigator in preparation for data analysis (IBM Corp, 2013). Pearson r correlation coefficients were generated between the GORT-5 ORI and the WJ IV ACH Broad Reading. Pearson r correlation coefficients were also generated for comparisons of the GORT-5 reading fluency and comprehension to the WJ IV ACH reading fluency and comprehension extended clusters.

In order to complete the second analysis, the percentile ranks of the GORT-5 and WJ IV ACH were converted into two dichotomous variables: the first being above the tenth percentile and the second being at or below the tenth percentile. Two analyses were generated following the conversion. The investigator ran a cross tabs, which produced contingency tables to illustrate dichotomous performance between each participant (Edwards, 2017). Secondly, the investigator conducted a Fisher's exact test, a Chi-Square like test, to examine the association between the binary classifications.

CHAPTER THREE

Results

Research Question 1: What is the correlation between the GORT-5 ORI, Fluency and Comprehension *scaled scores* and the WJ-IV ACH Broad Reading, Fluency and Reading Comprehension scores? The investigator predicts there will be very large, positive correlations given the similarity in constructs measured.

Research Question 1 was answered by generating a Pearson r correlation in SPSS for each aforementioned comparison between the GORT-5 and the WJ IV ACH. As predicted by the researcher, the participants' performance on the GORT-5 was closely associated with their performance on the WJ IV ACH tests overall, as shown in Tables 2 and 3. The Pearson r correlation coefficient for the GORT-5 ORI and WJ IV Broad Reading cluster is large, $r= .87, p= .01.$ Other comparisons yielded similar large correlations. The GORT-5 Fluency score and the WJ IV Fluency cluster score yielded the following coefficient, r= .85, p= .01; GORT-5 Comprehension and WJ IV Comprehension Extending, $r= .84, p= .01.$

Table 2 Pearson's *R* Correlation Between GORT-5 and WJ IV Reading Tests

		GORT-5 Score				
Criterion Test Score Used	SS	Rate	Accuracy	Fluency	Comprehension	ORI
WJ IV	Broad Reading	.85* (104)	.84* (104)	.87* (104)	.75* (104)	.87* (103)
	Basic Reading	.76* (92)	.86* (92)	.84* (92)	.74* (92)	.84* (91)
	Fluency	.86* (101)	.80* (101)	.85* (101)	.66* (101)	.81* (101)
	Comprehension Extended	.67* (97)	.77* (97)	.78* (97)	.84* (97)	.86* (96)

Numbers in parentheses are N
*p<.01

Table 3 Paired Samples Correlation

	N	Correlation	Significance
Pair 1: GORT-5 ORI Standard Score and WJ IV Broad Reading Standard Score	103	.868*	.000
Pair 2: GORT-5 ORI Standard Score and WJ IV Basic Reading Standard Score	91	.842*	.000
Pair 3: GORT-5 ORI Standard Score and WJ IV Reading Comprehension Ext. Standard Score	96	.862	.000
Pair 4: GORT-5 ORI Standard Score and WJ IV Reading Fluency Standard Score	100	.814*	.000

Research Question 2: Does GORT-5 ORI reading performance at or below the tenth percentile predict WJ-IV ACH performance at or below the tenth percentile for cluster scores commonly used in eligibility determinations: Reading Comprehension, Basic

Reading, and Reading Fluency? The investigator predicts GORT-5 performance at or below the tenth percentile can sufficiently predict performance on the select WJ-IV cluster scores, although significant levels will vary by pairings.

Research Question 2 is answered by generating Pearson product-moment correlation coefficients in SPSS for the aforementioned comparisons. The principal researcher utilized a Chi square-like statistic (i.e.*, Fishers Exact Test)* appropriate for 2x2 classifications to determine if the odds of scoring at or below the tenth percentile rank on the WJ-IV clusters is significantly increased when students fall at or below the tenth percentile rank on the GORT-5 ORI. The crosstabulations and Fisher's Exact Test are illustrated in Tables 4, 5, 6, and 7. Students who score at or below the tenth percentile on the GORT-5 similarly scored at or below the tenth percentile on the WJ IV ACH. Twenty-nine students or 27.8% performed at or below the 10^{th} percentile on both assessments, which illustrates a true positive. A total of 55 students (52.8%) performed above the 10^{th} percentile on both assessments, which indicates a true negative. A total of 6 students (5.8%) performed at or below the tenth percentile on the GORT-5 ORI but not in the Broad Reading cluster of the WJ IV ACH. The results indicate a false positive and would not identify those students as at-risk readers with the GORT-5 alone. Six students (5.8%) scored above the tenth percentile on the GORT-5 but at or below on the WJ ACH, which is indicative of a false negative. Remaining participants had a missing score for at least one of the measures and therefore were unable to be included in the calculations.

Table 4 GORT-5 ORI and WJ IV Broad Reading Percentile Rank Crosstabulation

	N		GORT-5 ORI Total Percentile Rank		
			1.00	2.00	
WJ IV Broad Reading PR	1.00	Count	29	6	35
		Expected Count	12.8	22.2	35.0
		Adjusted Residual	7.2	-7.2	
	2.00	Count	6	55	61
		Expected Count	22.2	38.8	61.0
		Adjusted Residual	-7.2	7.2	
Total		Count	35	61	96
		Expected Count	35.0	61.0	96.0

When examining the results of the crosstabulations for the GORT-5 ORI and the WJ IV Basic Reading cluster, the following comparisons were made. Twenty-four students performed at or above the 10th percentile in both areas assessed (23.1%); 51 scored at or below the 10th percentile, which indicated a true negative (49.0%). Nine students (9.7%) performed at or below the 10th percentile on the GORT-5, but not on the WJ IV, which is a false positive for identifying them as having a learning disability in the areas of Basic Reading skills. Only 4 students (3.8%) scored above the tenth percentile on the GORT-5 but at or below on the WJ IV, indicating a false negative.

Table 5 GORT-5 ORI and WJ IV Basic Reading Percentile Rank Crosstabulation

	N		WJ IV Basic Reading Total Percentile Rank		
			1.00	2.00	
GORT-5 ORI PR	1.00	Count	24	9	33
		Expected Count	10.5	22.5	33.0
		Adjusted Residual	6.4	-6.4	
	2.00	Count	4	51	55
		Expected Count	17.5	37.5	55.0
		Adjusted Residual	-6.4	6.4	
Total		Count	28	60	88
		Expected Count	28.0	60.0	88.0

In regard to the results of the crosstabulations of the GORT-5 Reading Comprehension and the WJ IV Reading Comprehension Extended a total of 30 students (28.8%) scored at or below the tenth percentile on both assessments (true positive); 47 scored at or above on both assessments to represent a true negative (45.2%). However, 10 students (9.6%) performed at or below the tenth percentile on the GORT-5, but not the WJ IV, which is a false positive. Solely using these scores would incorrectly identify them as an at-risk reader. Only 8 students (7.7%) scored above the tenth percentile on the GORT-5 but at or below the tenth percentile on the WJ IV, which would indicate a false negative.

Table 6 GORT-5 Reading Comp. and WJ IV Reading Comp. Ext Percentile Rank Crosstabulation

	N		WJ IV Reading Comp Ext. Percentile Rank		Total
			1.00	2.00	
GORT-5	1.00	Count	30	10	40
Reading Comp		Expected Count	16.0	24.0	40.0
PR		Adjusted Residual	5.9	-5.9	
	2.00	Count	8	47	55
		Expected Count	22.0	33.0	55.0
		Adjusted Residual	-5.9	5.9	
Total		Count	38	57	95
		Expected Count	38.0	57.0	95.0

The results of the GORT-5 Reading Fluency and WJ IV Reading Fluency crosstabulations yielded similar results as predicted and described above for the other areas. Twenty-seven students (26.0%) performed at or below the tenth percentile on both assessments indicating a true positive. Fifty-six students (53.8%) scored above the tenth percentile on both assessments indicating true negatives. Nine students (8.7%) performed at or below the tenth percentile on the GORT-5 but not the WJ IV, which indicates that the GORT-5 alone would identify them as an at-risk reader in the area of Reading

Fluency. Six students (5.8%) scored above the tenth percentile on the GORT-5 but at or below on the WJ IV, indicating a false negative.

Table 7 GORT-5 Reading Fluency and WJ IV Reading Fluency Percentile Rank Crosstabulation

	N		WJ IV Reading Comp Ext. Percentile Rank		Total
			1.00	2.00	
GORT-5 Reading Comp PR	1.00	Count	27	9	36
		Expected Count	12.1	23.9	36.0
		Adjusted Residual	6.6	-6.6	
	2.00	Count	6	56	62
		Expected Count	20.9	41.1	62.0
		Adjusted Residual	-6.6	6.6	
Total		Count	33	65	98
		Expected Count	33.0	65.0	98.0

CHAPTER FOUR

Discussion

In the present study, 104 school age children were administered the GORT-5 and WJ IV ACH in grades two through eleven. Participants represented 22 different elementary, middle and high schools in West Virginia and southeastern Ohio. Most students (80.2%) had IEP's, while 19.8% of participants had no diagnosed disability under IDEA 2004. All participants were English language proficient. Overall, students' scores on the GORT-5 proved to be closely aligned to their scores on the WJ IV. Very large, positive correlations were yielded between the GORT-5 ORI and WJ IV Broad Reading cluster, as well as the fluency and comprehension comparisons between each instrument.

When examining the sensitivity and specificity for each, the GORT-5 ORI and the WJ ACH Broad Reading cluster had a total of 29 (27.9%) students whose scores indicate a true positive, 55 (52.9%) indicate a true negative and only 6 showed false positives (5.8%) and false negatives (5.8%). With the scores of true positives and negatives being significantly higher than the false positive and negative, it indicates that the results are valid and not due to chance. Crosstabulations for comparisons between the GORT-5 ORI and the WJ IV Basic Reading Skills cluster, the GORT-5 Fluency and WJ IV Fluency cluster, and the GORT-5 Comprehension and the WJ IV Comprehension extended cluster yielded similar results. These results are comparable to the results found by Edwards (2017), with a smaller sample size of only students who were referred for special education. Her results yielded similar percentages of false positives and negatives.

Therefore, the results indicate that the GORT-5 is accurate at predicting a student's performance on the WJ IV reading tests.

Utilizing an instrument such as the GORT-5 that has proven validity with instruments such as the WJ-IV ACH can be beneficial to students, teachers and school psychologists. The GORT-5 can be administered in approximately 15-45 minutes and provides information regarding a student's rate, fluency, accuracy and comprehension. The brief time that it takes to administer this tool is not only more economical for reading teachers, diagnosticians, and school psychologists, but it also reduces the amount of time that a student is pulled away from classroom instruction, if they are needlessly referred for a full special education battery. Moreover, the GORT-5 can also provide the examiner with specific qualitative information regarding student performance. While the optional error analysis portion of the GORT-5 was not used in the study, it can be useful to individuals administering the GORT-5 as part of tier 2 or tier 3 to assist in instructional planning. Therefore, the results from the overall study indicate that within the MTSS process, the GORT-5 could be used as a valid screener for determining if a student should be referred for a subsequent special education evaluation.

Limitations

One primary limitation of the study is the difficulty to counterbalance the assessments because of the nature of the study design. Assessments were often administered in a random order by different examiners, which makes it difficult to control in the school setting. Each examiner was instructed to counterbalance the students which they managed. However, examiners had varying amounts of students which they tested which effected the overall counterbalance of the study.

Another limitation of the study is the researchers' inability to control the behavior of a participant. Each participant displayed varying levels of motivation, therefore it is difficult to differentiate between students who were highly motivated and giving their best effort to students who were disinterested in their performance. The investigator used no motivational or behavior measure to quantify the examinee's persistence and engagement in task completion during testing.

The final limitation size lies in the overall diversity of the sample. While the researchers worked to recruit more students who were not referred for special education, there was more access to participants in the school setting who were being referred for testing. Moreover, fewer participants were recruited in grades 11 and 12 and from racially and ethnically diverse backgrounds given the lack of diversity in the overall population.

Future Research

With the majority of students in this study (80.8%) having an IEP, future research could examine the differences within the different disability categories recognized under IDEA 2004. In addition to examining the differences amongst students with disabilities, the study could also continue to recruit more participants with no identified disabilities and those who are assumed to be of average or above average reading abilities to better balance the overall sample and provide more heterogeneity within the sample to better generalize the results.

In the future, the nature of the comprehension questions within the GORT-5 could be examined. Many students that struggle to read throughout the passage are still able to correctly answer certain comprehension questions. Further analysis of the relationship between a student's accuracy and comprehension scores on the GORT-5 could be beneficial to determine if there is a correlation between the two scores. Finally, examining the relationship between the GORT-5 and other widely used achievement instruments, such as the Weschler Individual Achievement Test, Third Edition, would further validate the use of the GORT-5 as a screening tool to predict a student's performance.

REFERENCES

Berkeley, S., Bender, W. N., Peaster, L. G., & Saunders, L. (2009). Implementation of response to intervention: a snapshot of progress. *Journal of Learning Disabilities,* 42, 85-95.

Blazer, C. (2010). Information Capsule. *Research Services.* 9(18).

Brown, J. I., Fishco, V. V., & Hanna, G. (1993). *Nelson-Denny reading test: Manual for scoring and interpretation, forms G & H*. Riverside Publishing Company.

Clemens, N., Hagan-Burke, S., Luo, W., Cerda, C., Blakely, A., Frosch, J., Gamez-Patience, B., & Jones, M. (2015). The Predictive Validity of a Computer-Adaptive Assessment of Kindergarten and First Grade Reading Skills. School Psychology Review: March 2015, 44(1), 76-97.

Decker, S. L., Hale, J. B., & Flanagan, D. P. (2013). Professional Practice Issues in the Assessment of Cognitive Functioning for Educational Applications. *Psychology in the Schools,* 50(3).

Deno, S. L. (2003). Curriculum-Based Measurement: Development and Perspectives. *Assessment for effective intervention,* 28(3-4), 3-12.

Della Toffalo, D. A., & Feifer, S. G. (2007). *Integrating RTI With Cognitive Neuropsychology: A Scientific Approach to Reading.* Middletown, MD: School Neuropsych Press, LLC.

Eagle, J. W., Dowd-Eagle, S. E., Snyder, A. & Holtzman, E. G. (2015). Implementing a Multi-Tiered System of Support (MTSS): Collaboration Between School Psychologists and Administrators to Promote Systems-Level Change. *Journal of Educational and Psychological Consultation.* 25(2-3), 160-177.

Edwards, M., (2017) The Relationship Between the Gray Oral Reading Test, Fifth Edition (GORT-5) and the Woodcock Johnson Test of Achievement, Fourth Edition (WJ IV ACH) For Referred Children. Marshall University.

Flanagan, D. P., & Alfonso, V. C. (2011). *Essentials of Specific Learning Disability Identification*. Hoboken, NJ: John Wiley & Sons.

Freeman, R., Miller, D., Newcomer, L. (2015). Integration of Academic and Behavioral MTSS at the District Level using Implementation Science. *Learning Disabilities: A Contemporary Journal.* 13(1), 59-72.

Fuchs, D., & Deschler, D. D. (2007). What we need to know about responsiveness to intervention (and shouldn't be afraid to ask). *Learning Disabilities Research & Practice,* 22, 129-136.

Fuchs, D., Fuchs, L. S., & Compton, D. L., (2011) Smart RTI: A Next-Generation Approach to Multilevel Prevention. *Exceptional Children*, 78(3). 263-279.

Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal measures of reading comprehension. *Remedial and Special Education,* 9(2), 20-28.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral Reading Fluency as an Indicator of Reading Competence: A Theoretical, Empirical, and Historical Analysis. *Scientific Studies of Reading*, 5(3), 239-256.

Fuchs, L.S., Fuchs, D., & Zumeta, R. O. (2008). Response to Intervention: A Strategy for the Prevention and Identification of Learning Disabilities. In *Educating Indviduals with Disabilities*: IDEIA 2004 and Beyond. 115-135.

Greif, S., Martin, R., & Spinath, B. (2014). Introduction to the special section on computer-based assessment of cross-curricular skills and processes. *Journal of Educational Psychology*, 106(3), 605-607.

Hall, A. H., & Tannebaum, R. (2012). Test Review. *Journal of Psychoeducational Assessment*, 31(5). 516-520

Hammill, D. D., Wiederholt, J. L., & Allen, E. A. (2006). *TOSCRF: Test of silent contextual reading fluency*. Pro-Ed.

Hudson, T., & McKenzie, R. (2016). Evaluating the Use of RTI to Identify SLD: A Survey of State Policy, Procedures, Data Collection, and Administrator Pereptions. *Contemp School Psychology*.20:31-45.

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

Individuals with Disability Education Act Amendments of 1997 [IDEA]. (1997). Retrieved from http://thomas.loc.gov/home/thomas.php

The IRIS Center. (2012). Response to Intervention. Retrieved from http://iris.peabody.vanderbilt.edu/module/rti01-overview/you-have-completed-this-module/#content

Jenkins, J. R., Fuchs, L. S., Espin, C., van den Broek, P., & Deno, S. L. (2000). Effects of task formant and performance dimension on word reading measures: Criterion validity, sensitivity to impairment, and context facilitation. In *Pacific Coast Research Conference, San Diego, CA.*

Kavale, K. A. (2005). Identifying specific learning disability: Is responsiveness to intervention the answer?. *Journal of Learning Disabilities,* 38(6), 553-562.

Kilgus, S. P., Methe, S. A., Maggin, D. M., & Tomasula, J. L. (2014). Curriculum-based measurement of oral reading (R-CBM): A diagnostic test accuracy meta-analysis of evidence supporting use in universal screening. *Journal of School Psychology*, 52, 377-405.

Klingbeil, D. A., McComas, J. J., Burns, M. K., & Helman, L. (2015). Comparison of predictive validity and diagnostic accuracy of screening measures of reading skills. *Psychology in the Schools,* 52(5), 500-514.

McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). *Technical Manual. Woodcock Johnson IV*. Rolling Meadows, IL: Riverside.

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.* doi: 10.1177/0734282915574028.

Mullis, R. (2012). Test intro: Gray oral reading test- fifth edition. Retrieved from http://lcpsdiags.com/lcpsdiags/debsays/79-debs-tests/88-gray

Cortiella, C., & Horowitz, S. H. (2014). *The state of learning disabilities: Facts, trends and emerging issues. New York: National center for learning disabilities,* 2-45.

National Center on Response to Intervention (2010). Essential Components of RTI – A - Closer Look at Response to Intervention. Washington, DC: Department of Education, Office of Special Education Programs, National Center on Response to Intervention.

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

Peterson, K. M. H., & Shinn, M. R., (2002). Severe Discrepancy Models: Which Best Explains School Identification Practices for Learning Disabilities? *School Psychology Review.* 31(4). 459-476.

Reschly, A. L., Busch, T. W., Betts, J., Deno, S. L., & Long, J. D. (2009). Curriculum-based measurement oral reading as an indicator of reading achievement: A meta-analysis of the correlational evidence. *Journal of School Psychology*, 47, 427–469.

Reschly, D. J., & Hosp, J. L. (2004) State SLD Identification Policies and Practices. *Learning Disability Quarterly*, 27, 197-213.

Restori, A. F., Katz, G. S., & Lee, H. B. (2009). A critique of the IQ/achievement discrepancy model for identifying specific learning disabilities. *Europe's Journal of Psychology*, 5(4), 128-145.

Schultz, E. K., & Stephens, T. L. (2009). SLD Identification: An Analysis of State Policies. *Academic Exchange Quarterly*, 13(4).

Shifrer, D., Muller, C., & Callahan, R. (2011). Disproportionality and Learning Disabilities: Parsing Apart Race, Socioeconomic Status, and Language. *Journal of Learning Disabilities,* 44(3), 246-257.

Skiba, R. J., Simmons, A. B., Ritter, S., Gibb, A. C., Rausch, M. K., Cuadrado, J. M., Chung, C. G. (2008) Achieving Equity in Special Education: History, Status and Current Challenges. *Exceptional Children,* 74, 264-288.

Speece, D. (2002). Classification of learning disabilities: Convergence, expansion, and caution. *Identification of learning disabilities: Research to practice.* 467-519

Tucker, C., & Jones, D. (2010). Response to Intervention: Increasing Fluency, Rate, and Accuracy for Students at Risk for Reading Failure. *National Forum of Educational Administration and Supervision Journal.* 28(1), 28-47.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2010) TOSREC: *Test of silent reading efficiency and comprehension.* Pro-Ed.

Wayman, M. M., Wallace, T., Wiley, H. I., Ticha, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *Journal of Special Education*, 41, 85–120.

Wiederholt, J. L., & Bryant, B. R. (2012). “Gray Oral Reading Tests – Fifth Edition” (GORT-5). Austin, TX: Pro-Ed, 2012.

APPENDIX A

Approval Letters

MARSHALL UNIVERSITY.
www.marshall.edu

Office of Research Integrity

April 18, 2017

Emily Nestor
School Psychology Program
Marshall University

Dear Ms. Nestor:

This letter is in response to the submitted thesis abstract entitled *"A Comparison Study of the WJIV ACH and the GORT-5."* 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

WE ARE... MARSHALL.
One John Marshall Drive • Huntington, West Virginia 25755 • Tel 304/696-4303
[illegible]

MARSHALL UNIVERSITY
www.marshall.edu

Office of Research Integrity
Institutional Review Board
One John Marshall Drive
Huntington, WV 25755

FWA 00002704

IRB1 #00002205
IRB2 #00003206

January 3, 2018

Rebecca Jennings-Knotts, PhD
School Psychology, MUGC

RE: IRBNet ID# 1139012-1
At: Marshall University Institutional Review Board #2 (Social/Behavioral)

Dear Dr. Jennings-Knotts:

Protocol Title:	[1139012-1] GORT-5 and WJ IV ACH Tests of Reading: A Comparison Study
Expiration Date:	January 1, 2019
Site Location:	MUGC
Submission Type:	New Project APPROVED
Review Type:	Exempt Review

In accordance with 45CFR46.101(b)(1)&(2), the above study and informed consent were granted Exempted approval today by the Marshall University Institutional Review Board #2 (Social/Behavioral) Designee for the period of 12 months. The approval will expire January 3, 2019. A continuing review request for this study must be submitted no later than 30 days prior to the expiration date.

If you have any questions, please contact the Marshall University Institutional Review Board #2 (Social/Behavioral) Coordinator Bruce Day, ThD, CIP at 304-696-4303 or day50@marshall.edu. Please include your study title and reference number in all correspondence with this office.