

## GORT-5 and WJ-IV ACH Tests of Reading: A Comparison Study

### **Purpose of Grant Funded Project**

The Gray Oral Reading Tests - Fifth Edition (GORT-5) is a reputable and widely used test of oral reading. It was first published in 1963 and remains a valuable tool in the identification of struggling readers and their specific oral reading strengths and deficits. The GORT-5, published nearly 50 years later in 2012, is ideal for use within a school's multi-tiered system of support (MTSS) when administered in combination with ongoing progress monitoring measures. The instrument has two equivalent forms; normative data for students ages six through 23; and a relatively brief administration time (i.e., approximately 20 minutes) (Wiederholt & Bryant, 2012).

As educators increasingly shift to commercially developed computer-based progress monitoring systems (Greiff, Martin, & Spinath), the use of instruments such as the GORT-5 may become more integral in MTSS to overcome the limitations of computer-based assessment. Computer-based assessments, 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. For these reasons, computer-based assessments can yield a series of highly variable data points for certain student subgroups. The inconsistent data can be difficult for educators to 1) interpret in a response-to-intervention framework and 2) subsequently recommend appropriate interventions and special education referrals.

The GORT-5, conversely, requires direct interaction between the examiner and examinee during testing administration. The direct observation of the examinee throughout the reading process informs intervention and examiner understanding of the extent to which attention and motivation impact reading results. The GORT-5 is consequently ideal for verifying the need for additional intervention and/or a referral for special education for students whose progress monitoring data are unpredictable or suspect. However, in order to validate the use of the GORT-5 as a robust diagnostic tool in MTSS, the instrument must engender reading scores that are comparable with results from other comprehensive reading assessments, such as the Woodcock Johnson Tests of Achievement – Fourth Edition WJ-IV ACH (Schrank, Mather, & McGrew, 2014). To date, an exhaustive review of the literature has produced no peer-reviewed studies describing the correlation between the GORT-5 and WJ-IV ACH reading tests.

### **Research Questions**

The following research questions guided the grant project:

1. What is the magnitude of the correlations between the GORT-5 and WJ-IV ACH tests?
2. Is the GORT-5 ORI standard score for participants equivalent to the WJ-IV Broad Reading, Reading Fluency, and Reading Comprehension standard scores?
3. Does GORT-5 ORI reading performance at or below the tenth percentile, predict WJ-IV performance at or below the tenth percentile for the WJ-IV Broad Reading cluster score and other clusters commonly used in eligibility determinations: Reading Comprehension Ext., Basic Reading, and Reading Fluency?
4. How well does the GORT-5 ORI predict performance on the WJ-IV ACH tests when using a ROC curve analysis?

## Method

### Participants

A total of 118 school age children and adolescents participated in the study. All participants resided in North Central Appalachia in either West Virginia or southeast Ohio. Sixty-five percent of participants (n=77) were male and 34.7% were female (n=41), which is proportionate to the overall gender composition of referred students in this region. The median age was 11-years-old. All participants were in grades two through 12. Sixty (50.8%), 34 (28.8%), and 24 (20.3%) participants were in elementary, middle school, and high school, respectively. The majority of the participants reported a (108; 90.8%) white, non-Hispanic race/ethnicity. The remaining students identified as multiple race/ethnicities (6; 5.0%); Asian (1; 0.8%), African American (1; 0.8%), or Hispanic (2; 1.7%).

This convenience sample consisted of students referred for academic assistance or special education services. It is, however, important to note that participants were referred for a variety of learning problems, social-emotional difficulties, and/or suspected exceptionalities including giftedness since gifted services are included under the state special education criteria in West Virginia. Thus, a participant was not required to have reading difficulties to be included in the study, although a large proportion of the participants had referral reasons linked to reading concerns. All participants were English language proficient.

### Measures

**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. Examiners instruct participants to read each passage aloud as quickly and carefully as possible. During the oral reading, the examiners mark miscues or deviations from print. Immediately after the participant finishes each passage, the examiner records the time in seconds and the total number of deviations. The examiner then removes the participant's response booklet from examinee's preview and asks the five scripted open-ended comprehension 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 (Wiederholt & Bryant, 2012a).

**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. "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's 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 seven of the eight reading measures were used. Letter-Word Identification, Passage Comprehension, Word Attack, Oral Reading, Sentence Reading Fluency, Reading Recall, and Reading Vocabulary (Mather & Wendling, 2015). The Letter-Word Identification measure requires participants to read a list of letters and/or words in isolation, whereas the Word Attack test required participants to read a list of nonsense words to evaluate phonics skills. The Passage Comprehension measure has a missing word designated by a blank space within a sentence or brief passage; the participants identifies the key word after reading the sentence or passage. The Oral Reading task requires participants to read groups of sentences aloud while the examiner marks errors. The Sentence Reading Fluency test is a three-minute timed task wherein the participants read simple sentences and decide if the answer is *Yes* or *No*. For the Reading Recall measure, participants silently read increasingly difficult passages and then retell as much of the story as they can remember to the examiner. Finally, Reading Vocabulary necessitates that participants read a key word aloud and then provide a synonym in the first subtest and an antonym in the latter subtest (Mather & Wendling, 2014).

The WJ-IV measures can be reported independently as single tests or aggregated as cluster scores in the areas of Reading, Broad Reading, Basic Reading Skills, Reading Comprehension, Reading Fluency and Reading Rate, and Phoneme-Grapheme Knowledge (LaForte, et al., 2014; Schrank, et al., 2014). Broad Reading, for example, encompasses the areas of comprehension, rate of reading, and accuracy and provides a comprehensive measure of reading achievement. The WJ-IV ACH tests can be administered to individuals 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).

**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 domains. Pearson correlations between Basic Reading Skills and the two Reading Comprehension clusters rendered coefficients of .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).

## **Procedures**

Certified school psychologists, diagnosticians, and school psychology graduate students administered the GORT-5 and the WJ-IV reading tests. All examiners received formal training on both instruments prior to participation. Testing was conducted in school settings under

standard conditions consistent with instructions, scripts, and time limitations outlined in the respective manuals.

In this study, only form A of the WJ-IV ACH test was administered although both forms A and B of the GORT-5 were counterbalanced since the criterion validity of the GORT-5 was under investigation. To the greatest extent possible, the GORT-5 and WJ-IV ACH reading tests were administered in one test session by the same examiner. However, given constraints of testing in a natural setting, as well as participants' attention spans and school schedules, the administration of both assessments was not always possible in one setting.

### **Data Analysis**

The investigators employed a mixed method and triangulation approach to assessing the criterion and classification accuracy of GORT-5 consistent with Cohen, Manion, and Morrison's (2011) recommendation. To answer Research Question 1, investigators generated Pearson product-moment correlation coefficients in SPSS to examine the strength of association between the GORT-5 ORI and the WJ-IV ACH Broad Reading, cluster and individual reading tests scores. Secondly, although paired samples t-tests are typically used when the same measure is repeated, dependent samples t-tests were incorporated to determine if the GORT-5 ORI standard scores for participants were equivalent to the WJ-IV Broad Reading, Reading Fluency, and Reading Comprehension standard scores.

To answer the third research question, investigators utilized contingency tables and a Fisher's Exact test, a Chi square-like statistic, appropriate for 2x2 binary classifications to determine if the odds of scoring at or below the tenth percentile rank on the WJ-IV clusters was significantly increased when participants scored at or below the tenth percentile rank on the GORT-5 ORI. The percentile ranks of the GORT-5 and WJ-IV ACH were converted into

dichotomous variables. All percentile ranks above the tenth percentile were assigned a value of one and all percentile ranks at or below the tenth percentile were assigned a value of zero. The cross tabs function was applied in SPSS to produce contingency tables, which illustrated dichotomous performance between each participant (Edwards, 2017).

A Receiver Operating Characteristic (ROC) curve analysis was used to answer the fourth and final research question. ROC curves are frequently used in the medical field and are easily adapted for use with psychological and educational assessments. The curve itself visualizes the relationship between sensitivity and specificity of a test and additionally lends to the optimal cut off to ensure the highest rate of true positives and the lowest rate of false positives (Swets, Dawes, & Mohahan, 2000). The Area Under the Curve (AUC) demonstrates how well a measure can discriminate students with a given condition of interest, like a reading disability. An AUC value of 1.0 signals perfect classification accuracy, whereas the minimal AUC of 0.5 indicates accuracy levels commensurate with chance for a binary classification. Moreover, a AUC value of .90 or higher is considered excellent in terms of assessment (Swets et al., 2000)

## **Results**

**Research Question 1:** What is the magnitude of the correlations between the GORT-5 and WJ-IV ACH tests?

The Pearson  $r$  correlation coefficient for the GORT-5 ORI and WJ-IV Broad Reading cluster was very large,  $r = .893$ ,  $p = .001$ . Coefficients for all other GORT-5 and WJ-IV score comparisons outlined in Table 1 ranged from  $r = .725 - .894$ . All correlations were significant at the .001 level. Although the magnitude of all coefficients was deemed very large (.70-.89) by

Hopkins' standards (2002), the GORT-5 Comprehension pairings with the WJ-IV Broad Reading, Reading Fluency, and Basic Reading Skills tests yielded slightly weaker correlations.

Table 1

WJ-IV TESTS	GORT-5 ORI	GORT-5 Fluency	GORT-5 Comprehension
Broad Reading	.893*	.894*	.791*
Reading Fluency	.852*	.872*	.725*
Basic Reading Skills	.869*	.865*	.784*
Reading Comprehension Ext.	.874*	.808*	.855*

\*p. <.001

**Research Question 2:** Is the GORT-5 ORI standard score for participants equivalent to the WJ-IV Broad Reading, Reading Fluency, and Reading Comprehension standard scores?

On average, the GORT-5 ORI standard score was found to be 3.6, 2.8, and 2.4 points higher than the WJ-IV Broad Reading, WJ-IV Reading Comprehension Extended, and WJ-IV Reading Fluency standard scores. All mean score differences were found to be significantly different at the .001 or .01 levels as shown in Table 2. Consequently, GORT-5 scores were not found to have equivalency with the WJ-IV ACH test scores. They were 2.4-3.6 points higher in general.

Table 2

		Mean	SD	SEM	t (df), p=
Pair 1	GORT-5 ORI	87.2	14.6	1.3	
n=118	WJ-IV Broad Reading	83.6	17.4	1.6	-5.03 (117), p = .000
Pair 2	GORT-5 ORI	86.7	14.2	1.3	
n=113	WJ-IV Reading Comp Ext	83.9	15.9	1.5	-3.93 (112), p = .000
Pair 3	GORT-5 ORI	87.4	14.6	1.3	
n=117	WJ-IV Reading Fluency	85.0	17.8	1.6	-2.78 (116), p = .006

**Research Question 3:** 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?

The contingency tables displayed in Table 3 illustrate that 38 participants (32.2%) scored at or below the tenth percentile on both assessments, whereas 64 (54.2%) participants scored above the tenth percentile on both assessments. Eight participants (6.8%) scored at or below the tenth percentile on the GORT-5, but scored above the tenth percentile on the WJ-IV ACH Broad Reading. Another eight participants (6.8%) scored above the tenth percentile on the GORT-5, while performing at or below the tenth percentile on the WJ-IV ACH. Crosstabulations for the GORT-5 and WJ-IV Reading Comprehension, Basic Reading, and Reading Fluency tests along with the Fisher's Exact test values will be reported in final paper for dissemination via a peer-reviewed journal.

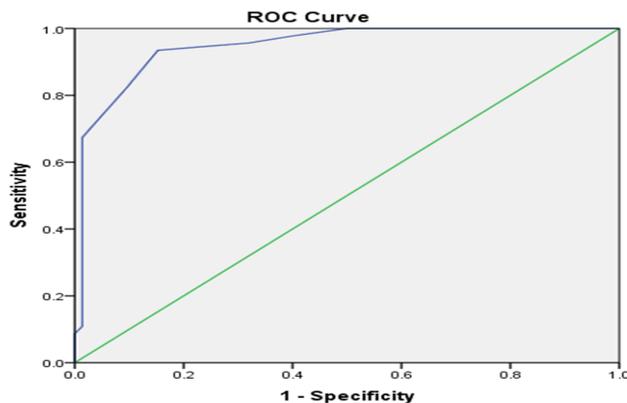
Table 3

		<b>WJ-IV Broad Reading</b>	
		At or below tenth percentile	Greater than tenth percentile
<b>GORT-5 ORI</b>	At or below tenth percentile	32.2% (38 students)	6.8% (8 students)
	Greater than tenth percentile	6.8% (8 students)	54.2% (64 students)

**Research Question 4:** How well does the GORT-5 ORI predict performance on the WJ-IV ACH test when using a ROC curve analysis?

ROC curve analysis yielded excellent results by Swets et al.'s (2000) standards: AUC=0.947, 95% confidence interval: 0.908-0.986,  $p < 0.001$ . The ROC curve is closely aligned to the upper left hand corner of the graph in Figure 1, markedly above the diagonal line which delineates the diagnostic odds associated with chance. The optimal cut point was found to be a standard score of 85 on the GORT-5 wherein the sensitivity value is .935 and 1- specificity value is .153.

Figure 1



## References

- Clemens, N. H., Hagan-Burke, S., Lou, 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, 44*(1), 76-97.
- Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education*. London: Routledge.
- Greiff, 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. Retrieved from <http://muezproxy.marshall.edu:2135/docview/1551095324?accountid=12281>
- Mather, N., & Wendling, B.J. (2014). Examiner's Manual. *Woodcock-Johnson IV Tests of Achievement*. Rolling Meadows, IL: Riverside.
- McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). Technical Manual. Woodcock Johnson IV. Rolling Meadows, IL: Riverside
- Schrank, F. A., & Dailey, D. (2014). Woodcock-Johnson Online Scoring and Reporting [Online format]. Rolling Meadows, IL: Riverside.
- Schrank, F. A., Mather, N., & McGrew, K. S. (2014). Woodcock-Johnson IV Tests of Achievement. Rolling Meadows, IL: Riverside.
- Swets, J.A., Dawes, R.M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest, 1*, 1–26.
- Wiederholt, J.L., & Bryant, B.R. (2012). Gray Oral Reading Tests-Fifth Edition (GORT-5). Austin, TX: Pro-Ed.