

1-2013

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

Anna H. Hall

Clemson University, ah2@clemson.edu

Rory P. Tannebaum

Clemson University

Follow this and additional works at: https://tigerprints.clemson.edu/eugene_pubs



Part of the [Education Commons](#)

Recommended Citation

Please use publisher's recommended citation: <http://jpa.sagepub.com/content/31/5/516.refs>

This Article is brought to you for free and open access by the Eugene T. Moore School of Education at TigerPrints. It has been accepted for inclusion in Publications by an authorized administrator of TigerPrints. For more information, please contact kokeefe@clemson.edu.

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

Anna H. Hall*, Ph.D., Rory P. Tannebaum, M.Ed.

Clemson University, Department of Teacher Education

*Corresponding author at: 100 Russell Ave, Greenville, SC 29609

Phone: 864-243-8413

Email: ah2@clermson.edu

Word Count: 2,376

Wiederholt, J.L., & Bryant, B.R. (2012). *Gray oral reading tests – fifth edition*. Austin, TX: Pro-Ed.

The first edition of the *Gray Oral Reading Tests* (GORT; 1963) was written by Dr. William S. Gray, a founding member and the first president of the International Reading Association. The GORT was designed to measure oral reading abilities (i.e., Rate, Accuracy, Fluency, and Comprehension) of students in Grades 2 through 12 due to the noteworthy advantages this type of assessment has over silent reading assessments (e.g., ability for the examiner to analyze miscues and identify an individual's ability to make letter-sound correspondences). Through the years, revisions of the GORT were published in 1986 (GORT-R), 1992 (GORT-3), and 2001 (GORT-4). The latest edition published in 2012 (GORT-5) includes updated norms extending from 6 years 0 months upward to 23 years 11 months, streamlined basal and ceiling rules, revised items that are passage-dependent, and additional studies showing evidence of sound psychometric properties (Wiederholt & Bryant, 2012a).

The purposes of the GORT-5 are to identify students with oral reading difficulties, determine strengths and weaknesses, evaluate student progress, and provide a standardized norm-referenced test that is appropriate for conducting reading research with school-age children (Wiederholt & Bryant, 2012a). The GORT-5 should be administered individually by examiners trained in formal assessment such as teachers, school psychologists, and diagnosticians.

Test Description

The GORT-5 kit includes an examiner's manual, two parallel forms of a student book (A and B), and one examiner record booklet for each form. Both student books contain 16 reading passages that increase with difficulty as the student moves from one passage to the next. The examiner record booklets contain five comprehension questions for each passage as well as

places to record student information, time spent reading, deviations from print, other reading behaviors (e.g., substitutions, omissions, additions), and prosody (i.e., expression, volume, phrasing, smoothness, and pacing).

Administration. The GORT-5 was designed to be administered to individual students in a quiet setting. Administration time typically ranges from 15-45 minutes with the test ideally being administered in one session. Two sessions are sometimes required if the student becomes fatigued or uninterested.

The examiner determines the first passage for each student based on knowledge of their individual reading abilities or by using the entry point by grade level table provided in the examiner record booklet (Wiederholt & Bryant, 2012b). The students are provided with the passage and directed to read the passage orally as “carefully and quickly as you can” (Wiederholt & Bryant, 2012b, p. 3). The examiner times the student reading and records deviations from print as the student reads the passage. Following the reading, the examiner removes the passage, reads the comprehension questions to the student, and records his/her answers. Testing continues until a basal and ceiling have been reached (determined by Fluency score).

Scoring. Raw scores are recorded for four subtests (Rate, Accuracy, Fluency, and Comprehension) as well as a sum of scaled scores which is used to determine the student’s Oral Reading Index (ORI). The Rate score is determined by the amount of seconds it takes the student to read aloud the passage while the Accuracy score is derived from the amount of words the student correctly pronounces during the reading. The Rate and Accuracy scores are combined to determine the Fluency score. The Comprehension score is the number of correct responses the student has to the questions about the passage. Finally, the ORI is a composite score determined by combining the student’s Fluency and Comprehension scaled scores.

Interpretation. The GORT-5 yields five types of normative scores: grade and age equivalents, percentile ranks, scaled scores, and the ORI (Mullis, 2012). The scores provide helpful information for identifying students who are significantly behind their peers in oral reading ability, selecting intervention strategies, evaluating student progress, and documenting student results to share with families and administrators. The two forms provided allow teachers and other appropriate professionals to test students after interventions for re-evaluation. As with any test, authors caution against using the GORT-5 as the sole basis for diagnoses or instructional decisions. It is best used in conjunction with other assessment strategies and by trained professionals (Wiederholt & Bryant, 2012a).

Technical Adequacy

Development and Standardization. The GORT-5 was normed using a sample of 2,556 students across 33 states. Standardization sites were selected based on representation of the four major U.S. geographic regions (i.e., the South, West, Northeast, and Midwest). Site coordinators were selected for each region to recruit schools with students that had demographics matching closely with the region as a whole. The amount of students tested per age level (i.e., 6-23 years) ranged from 58-237 with 142 being the average group size. Normative information is comprehensive and descriptive at each age level and the stratified variables (e.g., geographic region, gender, Hispanic status) reported conform to national expectations for each age group reported (Wiederholt & Bryant, 2012a).

Reliability. Five types of correlation coefficients were calculated to measure reliability of scores obtained with the GORT-5 including coefficient alpha, alternate forms (immediate administration), test-retest, alternate forms (delayed administration), and interscorer reliability. Averaged coefficient alphas for both forms of the test at all age intervals exceeded .90 which is

considered a highly desirable level. Consistently large alphas for 16 subgroups within the normative sample demonstrate that GORT-5 scores are equally reliable for all subgroups and contain little or no bias relative to any subgroup investigated (Wiederholt & Bryant, 2012a).

The alternate forms (A and B) were correlated and the averaged correlation coefficients for Rate, Accuracy, Fluency, and Comprehension were found to exceed .85. The average corrected ORI coefficient was .93. These results are strong enough to support that the two forms produce equivalent scores and do not suggest content-sampling error (Wiederholt & Bryant, 2012a).

The test-retest method was used with the GORT-5 with a sample of 248 students ranging in age from 6-23 years. The test was administered using both forms twice to all students 1 to 2 weeks apart. The coefficients for the combined sample were .82-.90 supporting the idea that scores on both forms have acceptable test-retest reliability. The same 248 students were used to conduct the alternate form (delayed administration) procedure for the GORT-5. The corrected coefficients for the combined sample ranged from .77-.88 providing further support for reliability (Wiederholt & Bryant, 2012a).

Four separate studies were conducted to test interscorer reliability of the GORT-5. The first three studies used multiple examiners to determine raw scores for Rate, Accuracy, Fluency, and Comprehension. Correlations of results were all .99 or greater. The final study examined interscorer reliability of the miscue analysis system resulting in agreement that exceeded 86% in all instances for each of the miscue categories (i.e., Meaning Similarity, Function Similarity, Graphic/Phonemic Similarity, Multiple Sources, and Self-Correction). The high level of interscorer agreement in these studies provides convincing support for different reporters finding consistent results when scoring the GORT-5. The reliability of the GORT-5 is reported as

consistently high across all five types of reliability tested suggesting that test users can have high levels of confidence in test results (Wiederholt & Bryant, 2012a).

Validity. Validity is the most fundamental consideration in evaluating assessments because it refers to the extent to which one can trust a test to measure what it is intended to measure (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999). Validity requires multiple sources of evidence in order to support the interpretation of scores (Downing, 2003; Messick, 1989). The GORT-5 manual provides strong evidence for content, construct, and criterion-related validity but lacks evidence related to response processes and consequences from testing.

Content validity for the GORT-5 was established by linking the format and scoring procedures to other leading reading tests that include passages as part of their battery (e.g., Test of Reading Comprehension – Fourth Edition [Brown, Hammill, & Wiederholt, 2009] and the Gilmore Oral Reading Test [Gilmore & Gilmore, 1968]). Passages included in the test were written to ensure student interest and avoid bias by focusing on topics of general interest and that were timeless in nature. Two higher level stories were added to the GORT-5 to extend the test for older students. Subtests were determined based on prevalent themes in current oral reading research.

Item difficulty, discrimination, and bias were also examined to provide content-related validity. Difficulty of items were within the acceptable range (i.e., between 15-85%) for all age groups. The discrimination power of items exceeded the .30 criterion for all ages. Finally a differential item functioning analysis was performed to test for item bias. All comprehension items were compared between three dichotomous groups (i.e., male vs. female; African

American vs. non-African American; and Hispanic vs. non-Hispanic). Statistically significant comparisons revealed negligible effect sizes lending support that the comprehension questions possess little or no bias in regard to gender, race, and ethnicity.

Construct-related validity for the GORT-5 was developed using a three-step procedure. This included the identification of several constructs expected to account for test performance, a generated set of hypotheses based on identified constructs, and using verified and empirical methods to verify these hypotheses (Wiederholt & Bryant, 2012a, p. 67). The relationship of performance on the GORT-5 with age and grade levels and intelligence were reported, as well as the intercorrelations of subtests, the differences among groups, and the performance of good and poor readers on the miscue analysis system. Expected patterns were found in all areas with the exception of academic abilities correlating less strongly with GORT-5 performance for secondary students (.33-.77) in comparison to elementary students (.63-.70).

Criterion-related validity evidence was established for the GORT-5 by comparing the test to five previously developed reading tests (e.g., The Nelson-Denny Reading Test [NDRT; Brown, Fishco, & Hanna, 1993]). It was found that GORT-5 scores had a strong correlation (i.e., average correlation coefficients of .68-.77 were large to very large in magnitude) with these other criterion measures developed by authorities in reading ability, therefore providing conclusive evidence of its validity in relation to similar measures.

A binary classification analysis was performed to examine the ability of the GORT-5 to discern which examinees actually have reading problems (Wiederholt & Bryant, 2012a). A sample of 632 students were administered at least two of the criterion validity measures to ensure poor readers were identified. Values reported for the GORT-5 using a cut score of 90 (i.e., 25th

percentile) met or exceeded recommended standards for binary classification indexes and ROC/AUCs including low numbers of false positives.

Although the GORT-5 provides extensive validity evidence from three different sources (content, construct, and criterion), the validity of this test could be strengthened by conducting studies on the response processes of students and examiners. This could be done through think alouds during test administration or through interviews immediately following administration. In addition, evidence is lacking on the consequences of testing of the GORT-5. Further evidence and justification for evaluating the intended and unintended results of score interpretation and use would provide further support for the use of this tool in proper settings.

Commentary and Recommendations

The GORT-5 is a comprehensive assessment of children's oral reading skills measuring four key areas including Rate, Accuracy, Fluency, and Comprehension. Scores on each subtest and the summed ORI score provide teachers with valuable information to help identify specific reading strengths and weaknesses of their students. This is especially important when we consider that there are multiple possible reasons for reading difficulties among students age 6-23 years. Percentile and scaled scores allow teachers to easily share information with families and provide documentation for administrators and policy makers. Reliability and validity evidence are strong and provide evidence across varying samples in 33 states, 2 different forms, and over time. The test is also well-suited for researchers conducting studies on oral reading with school-age children. (Wiederholt & Bryant, 2012a).

Because the GORT-5 uses a detailed method of miscue analysis and measures multiple skills during one assessment, ample training time (i.e., 3-6 hours) is needed, even for trained professionals who are new to using this tool. Before conducting formal assessments, it is

beneficial for examiners to practice using the tool with multiple children (with both forms of the test) to become familiar with the miscue recording methods, timing, and comprehension questions.

In terms of weaknesses, the following concerns about the GORT-5 should be noted. One major change from the GORT-4 to the GORT-5 is the use of open-ended comprehension questions instead of multiple-choice. The open-ended questions, though passage dependent, may be scored inconsistently based on the variability in answers that students provide. There is also a concern about the effect that timed reading may have on a student's overall reading performance. Some children may feel anxious when told to read "quickly" and may have more errors and miscues than they would in a natural reading setting. Finally, the GORT-5 lacks validity evidence related to response processes and consequences of testing.

Overall, the GORT-5 is a thorough oral reading assessment with sound psychometric properties. When used by properly trained professionals, this tool (in conjunction with parent and teacher observations) can help diagnose students with reading difficulties, monitor student progress, inform administrators and policymakers on the effectiveness of interventions, and provide valuable information for researchers.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Brown, J.A., Fishco, V.V., & Hanna, G. (1993). *Nelson-Denny reading test: Manual for scoring and interpretation, forms G & H*. Rolling Meadows, IL: Riverside Publishing.
- Brown, V.L., Hammill, D.D., & Wiederholt, J.L. (2009). *Test of reading comprehension – fourth edition*, Austin, TX: Pro-Ed.
- Downing, S.M. (2003). Validity: On the meaningful interpretation of assessment data. *Medical Education, 37*, 830-837.
- Gilmore, J. V., & Gilmore, E. C. (1968). *Gilmore oral reading test*. New York: Harcourt Brace Jovanovich.
- PAR, Inc. (2012). *Gray oral reading tests-fifth edition (GORT-5)*. Retrieved from <http://www4.parinc.com/Products/Product.aspx?ProductID=GORT-5>
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: American Council on Education.
- Mullis, R. (2012). *Test intro: Gray oral reading test- fifth edition*. Retrieved from <http://lcpdiags.com/lcpdiags/debsays/79-debs-tests/88-gray>
- Wiederholt, J.L., & Bryant, B.R. (2012a). *Gray oral reading test- fifth edition: Examiner's manual*. Austin, TX: Pro-Ed.
- Wiederholt, J.L., & Bryant, B.R. (2012b). *Gray oral reading test- fifth edition: Examiner's record booklet; form-A*. Austin, TX: Pro-Ed.