

Wilkinson, G. S., & Robertson, G. J. (2006). *Wide Range Achievement Test—Fourth Edition*. Lutz, FL: Psychological Assessment Resources. WRAT4 Introductory Kit (includes manual, 25 test/response forms [blue and green], and accompanying test materials): \$243.00.

Purpose and Nature of Test

The Wide Range Achievement Test—Fourth Edition (WRAT4) is designed to provide “a quick, simple, psychometrically sound assessment of academic skills” (Wilkinson & Robertson, 2006, p. 3). The test was first published in 1946 by Joseph F. Jastak, with the purpose of augmenting the cognitive performance measures of the Wechsler–Bellevue Scales, developed by David Wechsler. Jastak believed that academic performance should also be considered during a cognitive assessment battery. The WRAT has since been widely used as a norm-referenced measure of basic academic skills—specifically in reading, spelling, and mathematical calculations. It has subsequently undergone several revisions, with the WRAT–Revised (WRAT-R) in 1978, the WRAT–Third Edition (WRAT3) in 1993, and the WRAT4 in 2006. Expanding on the earlier versions, new features of the WRAT4 include a Sentence Comprehension subtest and a reading composite score (Wilkinson & Robertson, 2006).

Designed for use with individuals aged 5 through 94, the WRAT4 consists of four subtests—namely, Word Reading, Sentence Comprehension, Spelling, and Math Computation. As with the previous version, the WRAT4 is a single-level test with alternate forms (blue and green) that can be used interchangeably with comparable results. The blue and green forms can also be administered together for a more comprehensive evaluation.

The Word Reading subtest includes the letter recognition (15 items) and word reading (55 words). The letter recognition section is administered to children 7 years old or younger or to individuals who do not meet the basal scoring guidelines for the Word Reading subtest. This subtest is designed to measure letter and word recognition rather than speech or dictation.

The Sentence Comprehension subtest contains 50 items of one to two sentences each, where the participants fill in the blank with one or two words to indicate their comprehension of the sentence. The Sentence Comprehension subtest is new to the WRAT4, and it was added to the battery to address criticisms regarding the lack of construct validity for reading skills (Marby, n.d; Ward, n.d.).

Also new to the WRAT4 is the reading composite score, which combines the standard scores of the Word Reading and Sentence Comprehension subtests to provide a more comprehensive measure of reading achievement.

The Spelling subtest is made up of two parts. The first part, administered to children aged 7 years or younger, includes letter writing (i.e., 13 letters must be written); the second part consists of 42 words that must be spelled correctly. The

Spelling subtest may be administered individually or in small groups.

The Math Computation subtest is also made up of two parts: oral math (15 items) and math computation (40 problems to solve). The oral math section (for those aged 7 years or younger) must be administered individually; the math computation section may be administered individually or in small groups. Calculators may not be used for the Math Computation subtest, and there is a time limit of 15 min. The subtest is designed to measure an individual’s “ability to perform basic mathematics computations through counting, identifying numbers, solving simple oral problems and calculating written mathematics problems” (Wilkinson & Robertson, 2006, p. 2).

According to the manual, the “WRAT4 is designed to be administered and interpreted by individuals who have the necessary training and experience in administering and interpreting individually administered tests” (Wilkinson & Robertson, 2006, p. 5). The publisher lists the WRAT4 as a B-level test, which means that to purchase and use the test, an individual must have a

degree from an accredited 4-year college or university in psychology, counseling, speech-language pathology, or a closely related field PLUS satisfactory completion of coursework in test interpretation, psychometrics and measurement theory, educational statistics, or a closely related area; OR license or certification from an agency/organization that requires appropriate training and experience in the ethical and competent use of psychological tests. (Psychological Assessment Resources, 2008, n.p.)

Practical Applications

Recommendations for administration of the WRAT4 include providing a quiet room that is free from distractions, well lit, and so on, to ensure the most accurate reflection of the abilities of the individual who is taking the test. The manual also provides suggestions for developing rapport with the individual.

Although the test’s order of administration is somewhat flexible, the suggested order for administering the four subtests is as follows: Word Reading, Sentence Comprehension, Spelling, Math Computation. Furthermore, the authors indicate that the four subtests may be given separately or in combination of two or more at one sitting; however, they emphasize that the Word Reading subtest should be given before the Sentence Completion subtest. In addition, whereas the Spelling and Math

Computation subtests may be administered in a group setting, the Word Reading and Sentence Comprehension subtests must be given individually. According to the manual, administration time for the WRAT4 requires 15 to 45 min, depending on “the age, skill level and response style” (Wilkinson & Robertson, 2006, p. 7) of the individual being tested.

The WRAT4 has two forms, blue and green, considered to be equivalent versions and designed so that retesting may be conducted soon after the initial examination without the potential for practice effects.

The booklets and scoring sheets have administration instructions for ease of use, and the manual provides detailed instructions for test administration, using basal–ceiling rules. These rules are explained in detail as well as noted on the test forms and cards. For the Word Reading, Spelling, and Math Computation subtests, the *5 rule* refers to the minimum (basal) number of items that must be answered correctly on the second part before the administrator can exclude the administration of the preliminary items of each subtest (first part). The *7 rule* is a point in the testing when the examiner ends the Sentence Comprehension subtest after the test taker answers seven consecutive items incorrectly (ceiling value). The *10 rule* refers to the point of discontinuing the Word Reading and Spelling tests after 10 consecutive items are answered incorrectly (ceiling value).

Testing materials for administration of the WRAT4 include blue and green response forms for the Spelling and Math Computation subtests; test forms for recording the scores of the Word Reading, Spelling and Math Computation subtests; test forms for the Sentence Comprehension subtest; word reading list–spelling list cards; sentence comprehension cards and sentence comprehension sample cards; and optional place marker cards. In addition, the examiner will need pens and pencils, as well as a stopwatch or timepiece with a second hand. The examiner must be thoroughly familiar and comfortable with the testing materials and procedures.

Raw scores are obtained on each subtest and then converted to standard scores, based on a mean of 100 and a standard deviation of 15, according to age- and grade-normative groups, as provided in manual appendices. Optional scores—including percentile ranks, normal curve equivalents, stanines, and grade equivalents corresponding to standard scores—can also be obtained from the manual.

The publisher also offers two optional scoring and interpretation resources: the WRAT4 Scoring Program and the WRAT4 Interpretive Report. Both are capable of generating reports after raw scores are obtained, and both include a score summary table, a score profile for interpretation purposes, and a standard score comparison table so that comparisons may be made according to age or grade level. Confidence levels and percentile ranks are included in the reports. The WRAT4 Scoring Program allows for retention of individual records for future comparisons on achievement.

Technical Aspects

Information regarding psychometric properties of the WRAT4 is included in the professional manual (Wilkinson &

Robertson, 2006). As mentioned, raw scores are recorded on the scoring sheet, which administrators convert into standard scores and percentile ranks based on charts provided. Scoring may include confidence levels at 85%, 90%, or 95%. Standard scores may also be converted to normal curve equivalents, grade equivalents, stanines, or to one of seven verbal descriptions (ranging from lower extreme to upper extreme).

Norms

A stratified quota-based sampling procedure was used to standardize the WRAT4 on the basis of the 2001 U.S. Census. The sample size was 3,007, with 100 to 150 participants matched across 19 age groups. Samples were also matched to census data by gender, race/ethnicity, educational attainment, and geographic region. Race/ethnicity samples included African American, Caucasian, Hispanic (including Latino and Spanish equivalents), and others (consisting of Asian, American Indian, and Hawaiian / Pacific Islander). Educational attainment was used to indicate socioeconomic status, and it was based on years of school completion for those aged 18 and older. The four census categories were employed: college graduate, some college (no degree), high school graduate only, and less than a high school diploma. For those younger than age 18, parent educational attainment was used. In addition, race/ethnicity and educational attainment variables were matched to geographic regions, including the Northeast, the South, the North Central, and the West.

For grade-based samples, individuals with educational disability classifications were included in the norming process—namely, those with specific learning disabilities, speech and language impairments, mental retardation, emotional disturbance, physical impairments (including hearing, orthopedic, visual, and other health impairments), and attention-deficit/hyperactivity disorder. However, the representation of people with various types of disabilities was not adequately accomplished during the standardization process. Visually, hearing, and physically disabled schoolchildren, although included in the school-based sampling, had a lower percentage of representation when compared to that of the National Center for Education Statistics reports for 2002. For example, the reports indicate that 1.1% of the school-age population had visual, hearing, and physical disabilities. The sample obtained for standardization was only 0.2% ($n = 3$). Whereas the percentage of students with speech and language difficulties was 2.3% of the 2003 U.S. school-age population, this group represented only 0.7% of the standardization sample ($n = 12$). There is no evidence that adults with these types of disabilities were included in the sampling for standardization.

Reliability

Analysis in reliability for the WRAT4 has been conducted, including internal consistency and alternate-form reliability (Wilkinson & Robertson, 2006). For a test to have high levels of reliability, coefficients must be around or above .90 (Gay, 1992). The WRAT4 overall has high levels of internal consistency, ranging from .92 to .98. It also has moderate levels of

internal consistency within its subtests, with reliability coefficients ranging from .87 to .93. The reading composite score coefficients are high, ranging from .95 to .96 on both the blue form and the green.

Alternate-form reliability is also at moderate levels. Immediate retest coefficients range from .82 (for those aged 19–34) to .90 (for those aged 10–11). For all ages, the alternate-form reliability coefficient is .88. The delayed testing coefficient is .84 (for all ages). Test administrators are advised to use caution when retesting using alternate forms for delayed testing.

Validity

Validity studies for WRAT4 are limited, given that the revised version is relatively new. Wilkinson and Robertson (2006) state that the validity evidence presented in the manual is a continuation of the previous work done for earlier revisions of the WRAT and that it must continue as the WRAT4 version is utilized in the field. The authors show evidence for both internal validity and external validity, including content and concurrent validity. The test developers include other important measures of validity, including differential item functioning, subtest intercorrelations, and developmental changes.

Content Validity

Content validity refers to how well a test measures the intended content, and it depends on the validity of the items and that of sampling (Gay, 1992). Much of the questions from WRAT3 have been maintained and are included in WRAT4. Data provided in the manual indicate that content validity is acceptable, and the new version includes input from experts from the field. Intercorrelations among subtests (or measures of overlap) indicate a slight increase in divergent validity over the WRAT3 items, suggesting that subtests are more effectively measuring content.

Concurrent Validity

Concurrent validity refers to how closely scores of a given test are related to those of another established test based on the same criteria (Gay, 1992). Data provided in the manual indicate some degree of concurrent validity. The WRAT4 subtests were compared to several tests of academic achievement, including the WRAT–Expanded, Wechsler Individual Achievement Test–Second Edition, Woodcock Johnson III, Kaufman Test of Educational Achievement–Comprehensive Form, Kaufman Test of Educational Achievement–Brief Form, Wide Range Intelligence Test, Wechsler Intelligence Scale for Children–Fourth Edition, Stanford–Binet Intelligence Scale–Fifth Edition, Wechsler Abbreviated Scale of Intelligence, Kaufman Brief Intelligence Test–Second Edition, and Reynolds Intellectual Assessment Scales. Studies conducted for concurrent validity found low to moderate relationships between full-scale IQ and subtest scores, ranging from .57 (for spelling) and .72 (for reading). The authors note that these correlations are consistent with relationships found between IQ and achievement in other studies.

Concerns with validity for the previous WRAT3 were raised by several authors (e.g., Knoop, 2004; Ward, n.d.).

Marby (n.d.) strongly challenged construct validity for reading comprehension. In her test review, the author stated that reading ability cannot possibly be assessed if no text is read and that pronunciation alone is not sufficient to represent the domain of reading. She also argued that spelling ability is not related to a general cognitive ability, as claimed. Marby further questioned the validity of the mathematics subtest, given that it has a limited computational scope and as such cannot accurately determine computational skills. In terms of construct and concurrent validity, the author argued that a comparison with IQ scores obtained on the Wechsler Intelligence Scale for Children–Third Edition are not appropriate, because the constructs of IQ and academic achievement are not the same, nor are they based on the same normative samples. Ward (n.d.) also disputed the construct validity of WRAT3, indicating that

an obvious attempt has been made to address some of the problems identified in the revise of the previous edition, the WRAT-R. However, the major problem has not been addressed: the failure to identify the construct for the test and to provide evidence to support that construct. (para. 3)

Upon review of the technical data provided in the manual, the criticisms of Marby and Ward have not been fully addressed. Construct and concurrent validity indicators are based on the same tests of intelligence and achievement and so remain questionable.

Use of WRAT4 for Children and Adults With Disabilities or Learning Difficulties

The WRAT4 standardization process included samples of children with various disabilities, and subsequent studies were conducted to determine the capacity of the battery to discriminate among three groups: children with learning disabilities, those with low cognitive skills, and those with high cognitive skills. Results were significant for the learning disabled and low cognitive ability groups ($p < .01$) but not for the high cognitive ability group. Only three measures were statistically significant ($p < .01$)—namely, the Word Reading and Sentence Completion subtests and the Reading Composite score. The Spelling subtest ($p < .11$) and the Mathematics subtest ($p < .08$) did not discriminate between average- and high-ability children.

Validity has been in question since the development of the original WRAT. Marby (n.d.) and Ward (n.d.) maintained that the WRAT3 could not be used to determine reading skills, reading ability, or identification of children in need of special education services. In addition, Pikulski and Shanahan (1982) questioned the use of the original WRAT to place students at accurate reading levels. Their research indicated that when WRAT reading scores (for those aged 7–16) are compared to the reading scores of two other indicators, the former places students at levels two to three grades higher than that of the latter 30% to 37% of the time. Snart, Dennis, and Brailsford (1983) questioned the content validity of WRAT, claiming that it overestimated achievement, especially in math. They also determined that there was too much

of an emphasis on word recognition over essential reading constructs, such as decoding and reading comprehension. WRAT4 now includes the reading composite score, which combines the scores of the Sentence Comprehension and Word Reading subtests. Decoding is also included in the Spelling and the Reading Comprehension subtests, although to a limited degree. Sentence completion may assess reading comprehension but to a limited degree, given that it is a fill-in-the-blank-type question.

Gay (1992) discusses validity not as a constant construct but as that which can be valid for some and not for others:

It is the "valid for whom" concern that makes the description of the norm group so important. Only to the degree that persons in the norm group are like the persons to whom we wish to administer the test can proper interpretation of results be made. (p. 139)

A limited number of individuals with visual and auditory impairments were included in the norming samples for WRAT4, and no validity or reliability studies have been conducted on those groups. Because the subtests consist of verbal and visual modes, those who have difficulty seeing and/or hearing will not be effectively assessed. A Braille supplement is not available. A problem therefore exists for the valid assessment of children and adults with various types of disabilities—namely, those with visual and/or auditory impairments, as well as those with language difficulties.

Hannan (2007) discusses the problems of standardized tests and the lack of validity for those with blindness or low vision. This effectively excludes many individuals who need alternative testing procedures. Subsequently, this potentially affects school-age children, given that they may not benefit from No Child Left Behind initiatives. Hannan further explains that content validity may be examined for cultural and visual bias, namely, because an alteration of the test threatens validity—especially, construct validity: "The transcription of test items into Braille, for example, may change the construct of the test item" (p. 71). Because WRAT4 includes no normative data on visually or auditorily impaired groups or those with language difficulties, comparisons and valid assessments cannot be assured. Hannan recommends that new tests be developed with normative data on children with these types of disabilities.

Conclusion

The WRAT4 is an easy-to-use assessment of reading, spelling, and mathematical achievement for individuals aged 5 to adulthood. Adjustments have been made to improve the assessment of reading comprehension and, to a limited degree, decoding, as in the case of spelling and word pronunciation. However, much of the research and commentary regarding the WRAT, WRAT-R, and WRAT 3 suggest that it should not be used to identify children in need of special education services, nor should it determine reading levels of school-age children. These recommendations remain for the use of WRAT4. Although WRAT4 is compared to many widely used measurements of achievement and intelligence, its construct validity (namely, achievement) is

still in question. Its usefulness has not been established for children and adults with speech difficulties, as well as for those with visual and auditory impairments. Administrators should use caution when making evaluation judgments based on the use of the WRAT4 exclusively, especially when used to place children into special education services or to determine reading ability.

The authors of the test state,

The WRAT4 is intended for use by those professionals who need [an] . . . assessment of important fundamental academic skills. Such measures are valuable in initial evaluation of individuals referred for learning, behavior or vocational difficulties. The results of WRAT4 by themselves are not intended to provide formal identification of learning or cognitive disorders. (Wilkinson & Robertson, 2006, p. 3)

As such, it is important to highlight that the WRAT4 may be used to assist in the determination of basic academic levels of achievement but that it does not provide information on behavioral or vocational difficulties. The test may be one tool to use in the evaluation of vocational potential based on basic academic achievement. However, placement into special education services or determination of academic skill levels should not depend on the WRAT4 alone.

Cindy Ann Dell
Barbara Harrold
Thomas Dell
Montana State University–Billings

References

- Gay, L. R. (1992). *Educational research: Competencies for analysis and application*. Upper Saddle River, NJ: Prentice Hall.
- Hannan, C. K. (2007). Exploring assessment processes in specialized schools for students who are visually impaired. *Journal for Visual Impairment and Blindness*, 101(2), 67-69.
- Knoop, A. (2004). Test review: Wide Range Achievement Test 3. *Rehabilitation Counseling Bulletin*, 47(3), 184-185.
- Marby, L. (n.d.). [Review of the Wide Range Achievement Test 3]. In *The 12th mental measurements yearbook*. Retrieved January 23, 2008, from EBSCOHost Mental Measurements Yearbook Database.
- Pikulski, J. J., & Shanahan, T. (1982). *Informal reading inventory: A critical analysis. Approaches in the informal evaluation of reading*. Newark, DE: International Reading Association.
- Psychological Assessment Resources. (2008). *Qualification level: B*. Retrieved May 12, 2008, from <http://www3.parinc.com/products/product.aspx?Productid=WRAT4#>
- Snart, F., Dennis, S., & Brailsford, A. (1983). Concerns regarding the Wide Range Achievement Test (WRAT). *Canadian Psychology*, 24(2), 99-103.
- Ward, A. W. (n.d.). [Review of the Wide Range Achievement Test 3]. In *The 12th mental measurements yearbook*. Retrieved January 23, 2008, from EBSCOHost Mental Measurements Yearbook Database.
- Wilkinson, G. S., & Robertson, G. J. (2006). *Wide Range Achievement Test: Fourth Edition*. Lutz, FL: Psychological Assessment Resources.

Copyright of Rehabilitation Counseling Bulletin is the property of PRO-ED and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.