

Test Review

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive Test of Phonological Processing (CTOPP)*. Austin, TX: PRO-ED.

Purpose and Available Scores

The Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, & Rashotte, 1999) was developed to assess phonological processing ability in individuals 5 through 24 years of age. Phonological processing has been considered by many researchers to be clearly linked to reading deficits in students with learning disabilities. The theoretical model upon which this test is based consists of three correlated, yet separate, processing abilities: phonological awareness, phonological memory, and rapid naming. The CTOPP includes one test version for 5- to 6-year-olds (kindergarten and first graders) and another version for 7- to 24-year-olds (second grade through college). The 5- to 6-year-old version contains seven subtests and one supplemental test, while the 7- to 24-year-old version contains six subtests and six supplemental tests. Both versions are administered individually, requiring approximately 30 minutes for the core subtests. Core subtests administered to both the 5- and 6-year-olds (seven subtests) and the 7- through 24-year-olds (six subtests) include the following: Elision, Blending Words, Memory for Digits, Nonword Repetition, and Rapid Naming (colors, digits, objects and/or letters). Supplemental subtests allow the examiner to include additional nonword or real-word tasks, as examiners often have certain professional preferences regarding whether real words or nonwords provide the most accurate measure of phonological processing skills.

The purpose of the CTOPP is to identify students from kindergarten through college who need instructional activities to increase

their phonological skills. It is designed to identify individuals who perform significantly below their peers, according to national norms. By measuring important phonological abilities for individuals age 5-0 to 24-11 years and months, the CTOPP determines strengths and weaknesses among developed phonological processes, documents individual progress in phonological processing as a consequence of special interventions, and serves as a measurement device in research studies investigating phonological processing. According to the examiner's manual of the CTOPP, the test scores may assist in determining long-term educational goals. However, the test was not designed for planning daily instruction.

The CTOPP yields six types of scores: raw scores, age and grade equivalents, percentile ranks, standard scores, and composite scores. The subtest standard scores have a mean of 10 and a standard deviation of 3. The composite standard scores have a mean of 100 and a standard deviation of 15. While age and grade equivalents are available, the test authors recommended that examiners use percentiles or standard scores. This recommendation is in accord with the *Standards for Educational and Psychological Assessment* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999), as well as the views of numerous researchers who consider developmental scores to be less accurate than other scores, and often misinterpreted (McLoughlin & Lewis, 2005; Rotter, 2005). In addition, explicit descriptions for determining statisti-

cally significant discrepancies at a .05 level of confidence between subtest and composite scores are available.

Technical Adequacy

The sample used to norm the CTOPP included 1,656 participants in 30 states. According to the manual, the normative sample is representative of the nation as a whole based on data from the U.S. Bureau of the Census, 1997. The disability status of the sample was 90% with no disabilities, in addition to the following percent of disability groups: learning disabilities 4%, speech and language disabilities 4%, mental retardation 1% and other disabilities 1%. The demographic characteristics of the normative sample included: geographic region (northeast, midwest, south, west); gender (male, female); race (white, black, and other); residence (urban, rural); ethnicity (Native American, Hispanic, Asian, African American, other); family income (6 levels); and educational attainment of parents (three levels). Each of the demographic characteristics was reported according to six age levels. Samples sizes with far fewer than 100 participants included: Hispanic, Asian, and Native American groups, individuals with mental retardation and other disabilities.

Reliability data are provided for content sampling (homogeneity of items), time sampling (test-retest), and interscorer differences. For content sampling, all the coefficient alphas met or exceeded .70, 76% reached .80, and 19% reached .90. The composite score average coefficient alpha ranged from .83 to .95. Correlations by subtest with a sample size of 91 ranged from .67 (nonword repetition) to .97 (rapid letter naming). The majority of r values are between .70 and .97. Inter-rater reliability is .95 to .99 for all age ranges.

The examiner's manual reports a variety of validity measures: content validity, criterion prediction validity, construct identification validity, as well as conventional item analysis,

item response, and differential item functioning analysis. Evidence of content validity was presented by stating that each subtest was based on experimental tasks used to investigate phonological processing in the research literature. Since this is the first standardized norm-referenced test developed to assess phonological awareness in individuals at the K-16 grade levels, no accepted measures are available for collecting correlation data to verify its validity.

Item analysis, item response theory, and differential item functioning analysis were used to eliminate items that might show bias. To determine criterion prediction validity, a variety of experimental procedures and results were described in the manual. It was concluded that these studies established moderate to strong correlations between CTOPP subtests and other criterion measures, including the Gray Oral Reading Test (GORT-3; Wiederholt & Bryant, PRO-ED., 1992), Wide-Range Achievement Test (WRAT-3; Wilkinson, Jastak Associates, 1995), Woodcock Reading Mastery (WRMT-R; Woodcock, American Guidance Service, 1987), and the Test of Word Reading Efficiency (TOWRE; Torgeson, Wagner, & Rossette, PRO-ED., 1999). Several studies showed that the CTOPP differentiated groups with reading problems from the control groups of adequate readers. When measuring construct identification validity, factor loadings were high for the three factors in phonological awareness. The lengthy discussion of validity measures led to the conclusion that the CTOPP provides valid data to measure phonological processing.

Assessment for Effective Intervention

The CTOPP manual contains easily understood guidelines for the examiner regarding scripts, practice items, prompts, feedback statements, entry points, ceilings, scoring, and raw score conversion tables. One of the strengths of this instrument is that it is the

only individually administered, standardized norm-referenced test that focuses entirely on measuring the major components of the construct of phonological processing as a means of identifying individuals for intervention programs across the wide age span of kindergarten through college. Another strength is the test's use of an audio tape to deliver the subtests, in an effort to standardize the verbal presentation of word, syllable, nonword, and sound units. This eliminates errors that may arise from various speech accents in varying regions of the United States or individual examiner variations in pronunciation, rate of sound delivery, and articulation anomalies. Each subtest begins with practice items to familiarize the examinee with the task before beginning the subtest. In most cases, the subtest is not given if the examinee misses all the practice items. Both of these features, the audiocassette tape and the practice items, provide consistency for measuring phonological processing skills as verified by interscorer reliability coefficients of .95 to .99.

While the reliability of subtests varies considerably, interscorer reliability is consistently high. Thus, while some of the subtests are more reliably acceptable than others, the scoring and administration of the test are clear and consistent. When administering the subtests, examiners may want to be cautious with regard to the subtests that fall below .90 for eligibility decisions and below .80 for program decisions. It is generally accepted that individually administered tests should have minimal reliability coefficients close to or higher than .80, and coefficients of .90 or higher (Salvia & Ysseldyke, 2004).

When there is a strong relationship between instruction and assessment, students are more likely to improve their skills and documentation of their gains is made easier and more consistent (Peck & Scarpati, 2005). More research data showing this relationship will provide more substantial evidence and statistical support for the CTOPP. At this time the CTOPP is not designed to be used to

monitor student performance progress or make informed instructional decisions. Its purpose is to provide a standardized norm-referenced test to determine which students are more likely to be below the norm in phonological process. This is in contrast to assessment measures that provide instructional guidance, such as in curriculum-based measurement, which is the focus of much research at the present time (Safer & Fleischman, 2005).

The CTOPP was developed as the first step in a comprehensive evaluation of phonological processing skills. As stated in the examiner's manual, clinical teaching, criterion-referenced testing, and additional assessment procedures are needed in order to design instructional interventions for individual students.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- McLoughlin, J. A., & Lewis, R. B. (2005). *Assessing students with special needs* (6th ed.). Upper Saddle River, NJ: Merrill/Prentice Hall.
- Peck, A., & Scarpati, S. (2005). Instruction and assessment. *Teaching Exceptional Children*, 37(4), 7.
- Rotter, K. M. (2005). Problems with the use of developmental scores. *Communiqué*, 32(3), 5-6.
- Safer, N., & Fleischman, S. (2005). How student progress monitoring improves instruction. *Educational Leadership*, 62(5), 81-83.
- Salvia, J., & Ysseldyke, J. E. (2004). *Assessment* (9th ed.). Boston: Houghton Mifflin.
- Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of Word Reading Efficiency*. Austin, TX: PRO-ED.
- Wiederholt, J. L., & Bryant, B. R. (1992). *Gray Oral Reading Test—Third Edition*. Austin, TX: PRO-ED.
- Wilkinson, G.S. (1995). *The Wide Range Achievement Test—Third Edition*. Wilmington, DE: Jastak Associates.
- Woodcock, R. W. (1987). *Woodcock Reading Mastery Tests—Revised*. Circle Pines, MN: American Guidance Service.

Sherrel Lee Haight
Central Michigan University