

# Early Childhood Measurement and Evaluation Tool Review

Early Childhood Measurement and Evaluation (ECME), a portfolio within CUP, produces *Early Childhood Measurement Tool Reviews* as a resource for those who conduct screening, assessment, and evaluation. To learn more about ECME and CUP, provide feedback, or to access additional reviews, visit our website at [www.cup.ualberta.ca](http://www.cup.ualberta.ca) or email us at [cup@ualberta.ca](mailto:cup@ualberta.ca)

## Universal Nonverbal Intelligence Test (UNIT)

### Measurement Areas:

The Universal Nonverbal Intelligence Test (UNIT) is designed to test **intelligence** (cognitive ability) of children and adolescents ages **5 years 0 months through 17 years 11 months** who may be disadvantaged by verbal and language loaded measures. The UNIT was developed with the purpose of assessing general intelligence, two primary constructs (memory and reasoning), and two secondary processes (symbolic and nonsymbolic internal mediation.)

### Purpose:

The UNIT is a standardized, norm-referenced assessment tool that can be used:

- to measure general cognitive functioning;
- to identify intellectual giftedness, cognitive disabilities, and cognitive strengths and weaknesses;
- for persons belonging to groups where traditional intelligence testing would be inappropriate or unfair such as:
  - Individuals who are deaf or hearing impaired
  - Individuals from different cultural backgrounds
  - Individuals who are intellectually gifted
  - Individuals with learning disabilities
  - Individuals with limited English proficiency
  - Individuals with cognitive disabilities
  - Individuals with serious emotional disturbance or psychiatric disorders
  - Individuals with speech and language impairments
- as a guide for treatment program development;
- for placement decisions in clinical and educational settings; and
- to provide clinical information for neuropsychological evaluation and research.

### Length and Structure:

The UNIT is designed to be administered individually to children and adolescents. The standard test battery contains four subtests: Symbolic Memory, Cube Design, Spatial Memory, and Analogic Reasoning. The Extended battery contains two additional subtests: Object Memory and Mazes. The 4 core subtests take approximately 30 minutes to administer, and an additional 10-15 minutes are

required to administer the two additional subtests for the extended battery. All subtests are administered nonverbally and require a nonverbal response.

### **Materials:**

The UNIT is classified as a “Level C” qualification that is targeted to institutions with personnel possessing masters and doctorates of psychology or education, and/or has licensure in a relevant area of assessment with provincial or national organizations.

The UNIT complete kit is sold by the publisher for USD \$778.00, and includes: the examiner’s manual, two stimulus books, 16 response chips, response grid, 9 cubes, response mat, 10 symbolic memory cards, 25 record forms, 25 mazes response booklets, black pencil and carrying case. Without the carrying case the kit costs USD \$657.00. Scoring and interpretation software (UNIT Compuscore Software Program v1.1) are available for USD \$228.00. Additional record forms are available from the publisher in packages of 25 for USD \$66.00.

### **Administration, Scoring, and Interpretation:**

The UNIT manual suggests that the test examiners and scorers have (at minimum) technical training in test administration. Subtests are administered to examinees using 8 non-verbal gestures. Raw scores can be converted to Standard Scores and Percentile Ranks; Confidence Intervals can be constructed for all quotients as well as scaled scores and test age equivalents for all subtests. Individuals with formal graduate-level or professional training in psychological assessment should interpret test results using the 3 stage interpretation method described in the manual. The manual contains a section on interpretation, including exemplar case studies.

### **Subscales:**

The UNIT consists of 4 subscales: the Memory Quotient (MQ), the Reasoning Quotients (RQ), the Symbolic Quotient (SQ), and the Nonsymbolic Quotient (NSQ). The scaled scores on each of the subscales are combined to form the Full Scale Intelligence Quotient (FSIQ) score.

### **Documentation:**

The UNIT Administration and Examiner’s Manual provides specific procedures for administration and scoring. The manual also contains comprehensive chapters on interpretation, test standardization, norm development, validity, and reliability. Examples of case studies and interpretation are included in the manual.

### **Norming Sample:**

The UNIT was normed and standardized using an American sample of 2,100 children ages 5 years 0 months to 17 years 11 months 30 days. The sample was stratified according to demographic variables such as sex, race/ethnicity, Hispanic origin, parent education level, geographic region, community setting, classroom placement (regular or special education), and special education services (learning disability, speech and language impairments, serious emotional disturbance, cognitively delayed, giftedness, English as a second language and bilingual education). The demographic stratification on many of the variables are close to that of the 1995 US Census. According to the manual, persons classified as being Native American or American Indian were included in the standardization sample in numbers that would be similar to the 1995 US population Census demographic. The manual contains a detailed report of the sample’s demographic information as well as performance on the UNIT.

## Reliability:

**Internal Consistency:** A split-half reliability test determined the reliability of scores for two halves of the test using the standardization sample data. According to the manual, FSIQ reliabilities fell between .89 and .95 for the standard battery, .84 to .94 for the abbreviated battery, and .91 to .94 for the extended battery suggesting that there is a reasonable amount of internal item score consistency. Individual subtest scores were found to have reliability coefficients in the range of .64 to .91. Consistency scores were also compiled for a clinical sample and reliabilities were .96 for the abbreviated, .98 for the standard, and .98 for the extended batteries.

**Test-retest Reliability:** The manual indicates a sample of 197 children was tested with the UNIT twice in a 3-week interval. The test-retest reliabilities were calculated for each subtest, composite score and FSIQ. Reliability coefficients for FSIQ were found to be in the range of .78 and .91 for the standard and extended batteries. The abbreviated battery had ranges across ages of .74 to .89 for the FSIQ, both subtest scores, and composite scores, suggesting reasonably good test-retest reliability.

**Standard Error of Measurement and Confidence Intervals:** The manual also discusses the degree to which measurement error should contribute to an individual's observed score. According to the authors, if a child has a FSIQ score of 110 (standard battery), their "true score" should fall between 102 and 116 points at a 95% confidence interval. The confidence interval accounts for potential measurement error.

**Subtest Floors, Ceilings, and Difficulty Gradients:** The standard and extended batteries are deemed to have excellent floors for age 5, beyond that age subtests floors are deemed to be either good or excellent. In addition the floors of the subtests are deemed to be adequate to assess those at a lower functioning level. The UNIT has been deemed to have excellent ceilings for the oldest age groups as well as those at a higher level of functioning. According to the authors a "difficulty gradient refers to how rapidly a standard score increases as a function of an examinee's success or failure on a single test item" (p. 111). The results indicated that on average UNIT subtest gradients were adequate in detecting minor differences in examinee abilities.

## Validity:

An extensive discussion of validity is contained in the UNIT examiner's manual. Validity will be discussed in terms of internal and external evidence of validity.

### Internal Evidence of Validity:

**Content Validity:** The UNIT subtest tasks were designed to have relevance to central facets of intelligence (memory and reasoning) and to ensure a non-verbal response mode.

**Structural Evidence of Validity:** According to the authors, factor analyses indicated that the subtests do indeed follow a two factor structure: memory and reasoning. In addition, a confirmatory factor analysis was used to test the factor structure and again a two factor, memory and reasoning, model proved to be the best fit for the data.

### External Evidence of Validity:

**Concurrent validity:** is determined by demonstrating the scale is related to other scales that measure the same construct. According to the authors, the UNIT Abbreviated, Standard and Extended FSIQ were found to correlated strongly with FSIQ from the Wechsler Intelligence Scales

for Children-Third Edition (WISC-III; .78, .84, and .83 respectively). The correlation between UNIT abbreviated, standard and extended battery FSIQ and the Woodcock-Johnson-III Tests of Cognitive Abilities FSIQ was found to be .80, .83, and .82 respectively. Correlations between the UNIT and the Kaufman Brief Intelligence Test (K-BIT) were found to be .71 for the abbreviated battery, .82 for the standard battery, and .79 for the extended battery. The UNIT FSIQ was also correlated with scores on three measure of progressive matrices (Raven's Standard Progressive Matrices, Matrix Analogies Test, and the Test of Nonverbal Intelligence) correlations were found to range between .50 to .82.

**Predictive Validity:** The importance of predictive validity is that it demonstrates how well an instrument can predict academic achievement and generally intelligence is said to be strongly related to achievement. The UNIT is equally able to predict achievement as its competitors (e.g. WISC-III, SB5). Correlations between the UNIT FSIQ and the Woodcock-Johnson-Revised (WJ-R) achievement scores are in the range of .37 to .70 for samples of students who speak two languages, are intellectually gifted, and cognitively delayed or learning disabled. A correlation between the UNIT FSIQ and the Wechsler Individual Achievement Test (WIAT) basic skills cluster was found to be .62. According to Fives and Flanagan (2002), the "correlation of the Standard Battery with measures of achievement indicates that the UNIT fares comparably with its competitors." (p.442)

**Special Populations:** Another important facet of cognitive ability tests is the presence of distinctive score profiles for those in special populations (i.e. giftedness, cognitively delayed, developmental delay, limited English competency, speech/language disorders, learning disabilities, deafness and hard of hearing, severe emotional disturbance). The technical manual describes some possible patterns in mean scores that the various groups may demonstrate on the UNIT. Also it is important to note that these special groups were included in the standardization sample as well as in reliability and validity studies. The authors devote a full chapter in the examiner's manual to test fairness among special groups.

### **Additional Reliability and Validity Studies**

Pendley, Myers, and Brown (2004), investigated the UNIT with 29 students diagnosed with ADHD. Consistent with previous findings, children with ADHD scored lower on intelligence tests than those without ADHD. It was found that students with ADHD scored lower on the Memory Quotient portion of the UNIT. In terms of test-retest reliability, general findings were that the correlations were similar to those reported in the manual for the standard battery; however, for the extended battery FSIQ the test retest reliability was found to be .75. This is below what would be considered acceptable standards. Although the authors note that this finding may be indicative that the UNIT may not be adequately reliable for ADHD students, they warn that due to sample size constraints that the conclusion is premature.

### **Publication Information:**

This review is based on the Universal Nonverbal Intelligence Test (UNIT) published in 1998 by Riverside Publishing Company

### **Materials Used for Tool Review:**

- UNIT Examiner's Manual
- The Mental Measurements Yearbook
- Academic Journals

## References:

- Bandalos, D. L. (1998). Review of the Universal Nonverbal Intelligence Test. *Mental Measurements Yearbook*, 14.
- Bracken, B.A., & McCallum, R.S. (1998). *Universal Nonverbal Intelligence Test: Examiner's manual*. Itasca, IL: Riverside Publishing.
- Fives, C.J., & Flanagan, R. (2002). A review of the Universal Nonverbal Intelligence Test (UNIT). *School Psychology International*, 23, 425-448.
- Pendley, J. D., Myers, C. L., & Brown, R. D. (2004). The Universal Nonverbal Intelligence Test with children with attention-deficit hyperactivity disorder. *Journal of Psychoeducational Assessment*, 22, 124-135.
- 

**How to cite this document:** This document was created for CUP. However to cite this document use the following:

Community-University Partnership for the Study of Children, Youth, and Families (2011). *Review of the Universal Nonverbal Intelligence Test (UNIT)*. Edmonton, Alberta, Canada.