PSYCHOLOGICAL TEST LIST
PSYCHOLOGICAL TEST LIST
INDEX

AAMD Adaptive Behavior Scale
Adult Neuropsychological Questionnaire
Alcohol Use Inventory
Balthazar Scales of Adapt. Behavior
Beck Anxiety Inventory
Beck Depression Inventory
Beck Depression Inventory - 2nd Ed
Beck Hopelessness Scale
Bender Gestalt Monograph
Bender Gestalt Test
Bennett Mechanical Comprehension Test
Benton Visual Retention Test - 4th Ed
Boehm Test of Basic Concepts - Revised
Booklet Category Test
Boston Diagnostic Aphasia Examination
Boston Naming Test
BRL Sorting Test
California Psychological Inventory
Carlson Psychological Survey
Cattell Infant Intelligence Scale
Child Behavior Checklist
Child Neuropsychological Questionnaire
Childhood Trauma Questionnaire
Children’s Apperception Test
Children’s Depression Inventory
Children’s Manifest Anxiety Scale - Revised
Clinical Analysis Questionnaire
College and University Environment Scales
Columbia Mental Maturity Scale
Conners’ Rating Scales - Revised
Differential Ability Scales
Dementia Rating Scale
Developmental Test of Visual-Motor Integration
Devereux Elem Schl Beh Rating Scale
Dissociative Experiences Scale
Eating Disorder Inventory
Eating Disorder Inventory - 2nd Ed
Edwards Personal Preference Schedule
Family Environment Scale
Family Environment Scale: Children’s Ver
Forer Structured Sentence Completion Test
Frostig Developmental Test of Visual Percep
General Clerical Test
Geriatric Sentence Completion Form
Halstead-Reitan Neuropsych Test Battery
Harris Test of Lateral Dominance
Holtzman Inkblot Technique
House-Tree-Person Interrogation Form
Illinois Test of Psycholinguistic Abilities, Rev
Jenkins Activity Survey
Kaufman Adolescent & Adult Intel Test
Kaufman Assessment Battery for Children
Kaufman Test of Educational Achievement
Lawrence Psychological-Forensic Exam
Luria-Nebraska Neuropsychological Battery
McCarthy Scales of Children’s Abilities
Millon Adolescent Personality Inventory
Millon Clinical Multiaxial Inventory
Millon Clinical Multiaxial Inventory- 2nd Ed
Millon Clinical Multiaxial Inventory- 3rd Ed
Minnesota Multiphasic Personality Inventory
MMPI- 2 | Short Form
MMPI- Adolescent
Minnesota Percepto-Diagnostic Test
Myers-Briggs Type Indicator
NEO Personality Inventory - Revised
Paulhus Deception Scales
Peabody Individual Achievement Test
Peabody Individual Achievement Test - Revised
Peabody Picture Vocabulary Test - Revised
Peabody Picture Vocabulary Test - 3rd Ed
Personal Assessment Inventory-IPAT Dep Scale
Personality Assessment Inventory
Personality Research Form
Porteus Maze Test
Psychological Screening Inventory
Psychosocial Pain Inventory
Questionnaire on Resources and Stress
Quick Neurological Screening Test - Revised
Raven Standard Progressive Matrices
Rogers Criminal Responsibility Assessment Scales
Rorschach Inkblot Test
Rotter Incomplete Sentences Blank - 2nd Ed
Short Employment Tests
Sixteen Personality Factor Questionnaire
Slosson Drawing Coordination Test
Slosson Intelligence Test
Slosson Oral Reading Test
Stanford-Binet Intelligence Scale: Form L-M - 3rd Ed
Stanford-Binet Intelligence Scale - 4th Ed
Stanford-Binet Intelligence Scale - 5th Ed
Stanford Diagnostic Mathematics Test
State-Trait Anxiety Inventory
Study of Values
Symptom Checklist 90 - Revised
System of Multicultural Pluralistic Assessment
Szondi Test
Tell Me A Story
Tennessee Self Concept Scale
Test Anxiety Inventory
Test of Language Development
Test of Language Development - 2nd Ed
Thematic Apperception Test
Vineland Adaptive Behavior Scales
Wahler Physical Symptoms Inventory
Wechsler Memory Scale - 3rd Ed.
Wechsler Adult Intelligence Scale
Wechsler Adult Intelligence Scale - Revised
Wechsler Adult Intelligence Scale - 3rd Ed
Wechsler Bellevue Intelligence Scale
Wechsler Individual Achievement Test
Wechsler Intelleligence Scale for Children - Rev
Wechsler Intelleligence Scale for Children - 3rd Ed
Wechsler Intelleligence Scale for Children - 4th Ed
Wechsler Preschool & Primary Scale of Intelligence
Wechsler Preschool & Primary Scale of Intel - Rev
Wechsler Preschool & Primary Scale of Intel - 3rd Ed
Whiticker Index of Schizophrenic Thinking
Wide Range Achievement Test
Wide Range Achievement Test - Revised
Wide Range Achievement Test - 3rd Ed
Wide Range Assessment of Memory and Learning
Wonderlic Personnel Test
Woodcock-Johnson III Tests of Achievement
Woodcock-Johnson III Tests of Cognitive Abilities
Word Association Test
AAMD Adaptive Behavior Scale

**Purpose:** Designed to measure children’s personal independence and social skills.

**Population:** Mentally retarded and emotionally maladjusted ages 3- adult, grades 2-6.

**Score:** Domain, Factor, and Comparison Scores.

**Time:** (30-120) minutes.

**Authors:** Kazuo Nihira, Ray Foster, Max Shellhaas, and Henry Leland.

**Publishers:** Publishers Test Service.

**Description:** The AAMD Adaptive Behavior Scale is designed to measure children’s personal independence and social skills. Adaptive behavior is a critical component in the diagnostic classification of the mentally retarded and is defined as “the effectiveness or degree with which the individual meets the standards of personal independence and social responsibility expected for his or her age and cultural group.”

**Scoring:** Three main types of scores are derived from the item response scores (Domain, Factor, and Comparison scores.) The Domain score is simply the summation of all items within a given domain, the Factor score is the summation of the Domain scores with a given factor, and the Comparison score is a weighted summation of three Factor scores. Thus, there is a progression in score derivation that results in a score that compares a child to same-aged peers in either Regular, EMR, or TMR reference groups.

**Reliability:** The only form of reliability data documented in the technical manual is the internal consistency of each factor via the coefficient alpha technique. With the exception of the Personal Adjustment Factor, the coefficient alphas are high (range .71 to .97). The omission of test-retest and interrater reliabilities is a major concern for a test that can be administered to several parties (i.e., teachers, parents, guardians) and, in the case of a handicapped child, possibly several times over the course of a few years.

**Validity:** Two types of validity data are presented in the manual: (a) data on the relationship between adaptive behavior ratings and intelligence test performance and (b) data on the predictive power of the ABS for accurately classifying normal and mentally retarded children. With respect to intelligence, most of the 21 ABS domains have low to moderate correlations with IQ test performances. Noticeable exceptions were observed between IQ and the Language Development domain (r ranging from .39 to .63 depending on age group) and IQ and the Numbers and Time domain (r ranging from .33 to .62).

**Norms:** The ABS was standardized on a sample of 6,523 individuals in California and Florida ranging in ages from 3 through 17 years. Individuals sampled were classified into one of three groups: Regular, EMR, or TMR. Individuals from various racial/ethnic groups and locales were included. Although the sample represents only two states, the size and diversity of the sample are very good compared to other adaptive behavior scales.

**Suggested Uses:** Recommended as part of a classification/diagnostic battery in screening and placement decisions regarded the mentally handicapped.
Adult Neuropsychological Questionnaire

**Purpose:** Designed to screen for brain dysfunction.

**Population:** Adults.

**Time:** (5-10) minutes.

**Author:** Fernando Melendez.

**Publisher:** Psychological Assessment Resources, Inc.

**Description:** The Adult Neuropsychological Questionnaire consists of 54 items designed to be administered as a semi-structured interview. It can also be self-administered. The purpose of the questionnaire is to “inquire about complaints, symptoms, and signs that may suggest underlying brain dysfunction or other organic conditions” (manual, p.1). It was developed initially as an aid for students who were not well versed in signs and symptoms associated with neurologic conditions.

**Scoring:** The test manual is essentially an interpretive guide that contains a rationale for the items that have been included. The authors suggest that formal assessment will often confirm or clarify the symptoms endorsed, although no data are presented to support this assertion.

**Reliability and Validity:** The manual contains no information on reliability or validity.

**Norms:** The manual contains no normative data.

**Suggested Uses:** Although designed to help students and non-neuropsychologist inquire about symptoms, history, or other complaints that may be associated with neuropsychological disorders, reviewers discourage its use claiming poor organization and over-interpretations about the significance and possible attribution of specific symptoms.
**Alcohol Use Inventory**

**Purpose:** Designed to assess the nature of and problems associated with alcohol use pattern.

**Population:** Adults and adolescents 16 years of age and over.

**Scales:** Primary Scales measuring Benefits, (Social Improvement, Mental Improvement, Manage Moods, Marital Coping, Gregarious, Compulsive, Sustained), Consequences (Loss of Control, Role Maladaptation, Delirium, Hangover, Marital Problems), Concerns and Acknowledgements (Quantity, Guilt and Worry, Help Before, Receptivity, Awareness), Second Level, Enhanced, Obsessed, Disruption 1, Disruption 2, Anxious Concern, Receptive Awareness, and Alcohol Involvement.

**Time:** (35-60) minutes.

**Authors:** John L. Horn; Kenneth W. Wanberg, F. Mark Foster.

**Publisher:** NCS Assessments

**Description:** The Alcohol Use Inventory (AUI) is a self-report inventory comprised of 24 scales designed to assess patterns of behavior, attitudes, and symptoms pertaining to the use of alcohol of individuals 16 years of age or older who drink to some extent. It was developed for use with individuals admitted to an alcoholism treatment program. The AUI reflects the multiple condition theory about drinking problems and has evolved from studies of the features of a single aspect of alcoholism. The AUI is not appropriate for use with individuals who do not drink. The AUI is designed to provide operational indicators to describe patterns of alcohol use, allowing mental health workers to identify and understand different kinds of alcoholics.

**Scoring:** The Primary Scales focus on the “benefits,” “styles,” “consequences,” and “concerns and acknowledgments” of drinking. There are six second-level scales derived from factor analysis of the relationships among the primary scales. The third level score is a measure of broad involvement with alcohol. Total raw scores for both the hand scoring and the computerized forms, are plotted on a graph so the decile rank are read.

**Reliability & Validity:** The AUI presents internal consistency reliabilities and test-retest information on the scales from a number of different years and groups. The third-level scale and second-level scales have higher coefficients than the primary scales as expected. Of the Primary Scales, most coefficients ranged from .65-.80. The scales appear to be reasonably associated with problem drinking.

**Norms:** The AUI was standardized on a sample of 1290 people applying for first time inpatient alcohol treatment. The mean age was 39.16, (SD= 11.89) and the mean for years of education were 11.50 (SD= 2.75). There were about5 15% women, 8% Black, and 21% Hispanic. About 45% were married or living in a “marriage-type” relationship. Other than that, there is not a good demographic description of the makeup of the norming group.

**Suggested Use:** A useful assessment tool when working with individuals with alcohol problems. The scales can be interpreted based on the benefits, styles, consequences and concerns associated with alcohol use, and is helpful in treatment. Combinations of scores have been developed into typologies which indicate ways to relate to the client, and can help with treatment planning.
Balthazar Scales of Adaptive Behavior I: Scales of Functional Independence

Purpose: Designed to measure self-care differences in mentally retarded children and adults.

Population: Developmentally disabled individuals ages 5 and over.

Score: Three scale scores.

Time: N/A.

Author: Earl. E. Balthazar.

Publisher: Consulting Psychologists Press, Inc.

Description: The Balthazar Scales of Adaptive Behavior I: Scales of Functional Independence (BSAB-I) were developed to measure very small differences in the self-care experiences of profoundly and severely mentally retarded children and adults. One use of the scales is to assess individual behavior in self-care skills and compare these scores with scores in normative groups. These scale measurements are also used to identify weaknesses in eating, dressing, and toileting skills in order to give appropriate training or treatment, after which they can be applied again to determine the amount of improvement in the skills where inadequacies were assessed.

Scoring: The scoring produces scores for three scales: the Eating Scale, the Dressing Scale, and the Toilet Scale. Detailed and explicit scoring instructions are included in the manual. Reviewers suggest that two to three hours are required to become thoroughly familiar with the scoring procedures.

Reliability: The reliability coefficient was .873 for the Eating Scale, .965 for the Dressing Scale, and .939 for the Toileting Scale, resulting in respectable interrater reliability.

Validity: The manual reports no external validity studies.

Norms: The BSAB-I was standardized on profoundly and severely mentally individuals aged from 5 to 57 years, with IQs ranging from 20 to 35 on global intelligence tests.

Suggested Uses: Recommended uses include assessment of the developmentally disabled in clinical and research settings.
Beck Anxiety Inventory

Purpose: Designed to discriminate anxiety from depression in individuals.

Population: Adults.

Score: Yields a total score

Time: (5-10) minutes.

Author: Aaron T. Beck.

Publisher: The Psychological Corporation.

Description: The Beck Anxiety Inventory (BAI) was developed to address the need for an instrument that would reliably discriminate anxiety from depression while displaying convergent validity. Such an instrument would offer advantages for clinical and research purposes over existing self-report measures, which have not been shown to differentiate anxiety from depression adequately.

Scoring: The scale consists of 21 items, each describing a common symptom of anxiety. The respondent is asked to rate how much he or she has been bothered by each symptom over the past week on a 4-point scale ranging from 0 to 3. The items are summed to obtain a total score that can range from 0 to 63.

Reliability: The scale obtained high internal consistency and item-total correlations ranging from .30 to .71 (median=.60). A subsample of patients (n=83) completed the BAI after 1 week, and the correlation between intake and 1-week BAI scores was .75.

Validity: The correlations of the BAI with a set of self-report and clinician-rated scales were all significant. The correlation of the BAI with the HARS-R and HRSD-R were .51 and .25, respectively. The correlation of the BAI with the BDI was .48. Convergent and discriminant validity to discriminate homogeneous and heterogeneous diagnostic groups were ascertained from three studies. The results confirm the presence of these validities.

Norms: The three normative samples of psychiatric outpatients were drawn from consecutive routine evaluations at the Center for Cognitive Therapy in Philadelphia, Pennsylvania. The total sample size was 1,086. There were 456 men and 630 women.

Suggested Uses: Recommended for use in assessing anxiety in clinical and research settings.
Beck Depression Inventory (BDI)

**Purpose:** Designed to measure presence of depression.

**Population:** Adolescents and adults.

**Score:** Produces single score indicating intensity of the depressive episode.

**Time:** Not reported.

**Author:** Aaron T. Beck.

**Publisher:** Center for Cognitive Therapy.

**Description:** The Beck Depression Inventory (BDI) is a 21-item test presented in multiple choice format which purports to measure presence and degree of depression in adolescents and adults. Each of the 21-items of the BDI attempts to assess a specific symptom or attitude “which appear(s) to be specific to depressed patients, and which are consistent with descriptions of the depression contained in the psychiatric literature.” Although the author, Aaron T. Beck, is associated with the development of the cognitive theory of depression, the Beck Depression Inventory was designed to assess depression independent of any particular theoretical bias.

**Scoring:** Each of the inventory items corresponds to a specific category of depressive symptom and/or attitude. Each category purports to describe a specific behavioral manifestation of depression and consists of a graded series of four self-evaluative statements. The statements are rank ordered and weighted to reflect the range of severity of the symptom from neutral to maximum severity. Numerical Values of zero, one, two, or three are assigned each statement to indicate degree of severity. Beck admits that there is no arbitrary cutoff score and the specific cutoff depends on the characteristics of the patients used and the purpose for which the inventory is given.

**Reliability:** Test-retest reliability has been studied in the case of 38 patients who were given the BDI on two occasions. It was discovered that the changes in BDI scores tended to parallel changes in the clinical reading of the depth of depression, indicating a consistent relationship between BDI scores and the patient’s clinical state. The reliability figures here were above .90. Internal consistency studies demonstrated a correlation coefficient of .86 for the test items, and the Spearman-Brown correlation for the reliability of the BDI yielded a coefficient of .93.

**Validity:** In assessing the validity of the BDI, the readily apparent face validity of the BDI must be addressed. The BDI looks as though it is assessing depression. While this may be quite advantageous, it may make it easy for a subject to distort the results of the test. Content validity would seem to be quite high since the BDI appears to evaluate well a wide variety of symptoms and attitudes associated with depression. One study addressing concurrent validity demonstrated a correlation of .77 between the inventory and psychiatric rating using university students as subjects. Beck reports similar studies in which coefficients of .65 and .67 were obtained in comparing results of the BDI with psychiatric ratings of patients.

**Norms:** Normative sample included 226 psychiatric in- and out-patients.

**Suggested Uses:** The BDI is recommended in research and clinical settings.
Beck Depression Inventory - 2nd Edition

**Purpose:** Designed to determine presence and severity of symptoms of depression.

**Population:** Adolescents and adults.

**Score:** Produces single score indicating intensity of the depressive symptoms.

**Time:** 5-10 minutes, longer for patients with severe depression or obsessional disorders.

**Author:** Aaron T. Beck, Robert A. Steer, and Gregory K. Brown.

**Publisher:** the Psychological Corporation.

**Description:** The Beck Depression Inventory Second Edition (BDI-II) is a 21-item self-report instrument intended to assess the existence and severity of symptoms of depression as listed in the American Psychiatric Association’s Diagnostic and Statistical Manual of Mental Disorders Fourth Edition (DSM-IV; 1994). This new revised edition replaces the BDI and the BDI-1A, and includes items intending to index symptoms of severe depression, which would require hospitalization. Items have been changed to indicate increases or decreases in sleep and appetite, items labeled body image, work difficulty, weight loss, and somatic preoccupation were replaced with items labeled agitation, concentration difficulty and loss of energy, and many statements were reworded resulting in a substantial revision of the original BDI and BDI-1A. When presented with the BDI-II, a patient is asked to consider each statement as it relates to the way they have felt for the past two weeks, to more accurately correspond to the DSM-IV criteria.

**Scoring:** Each of the 21 items corresponding to a symptom of depression is summed to give a single score for the BDI-II. There is a four-point scale for each item ranging from 0 to 3. On two items (16 and 18) there are seven options to indicate either an increase or decrease of appetite and sleep. Cut score guidelines for the BDI-II are given with the recommendation that thresholds be adjusted based on the characteristics of the sample, and the purpose for use of the BDI-II. Total score of 0-13 is considered minimal range, 14-19 is mild, 20-28 is moderate, and 29-63 is severe.

**Reliability:** BDI has been used for 35 years to identify and assess depressive symptoms, and has been reported to be highly reliable regardless of the population. It has a high coefficient alpha (.80) its construct validity has been established, and it is able to differentiate depressed from non-depressed patients. For the BDI-II the coefficient alphas (.92 for outpatients and .93 for the college students) were higher than those for the BDI-1A (.86). The correlations for the corrected item-total were significant at .05 level (with a Bonferroni adjustment), for both the outpatient and the college student samples. Test-retest reliability was studied using the responses of 26 outpatients who were tested at first and second therapy sessions one week apart. There was a correlation of .93, which was significant at p < .001. The mean scores of the first and second total scores were comparable with a paired t (25)=1.08, which was not significant.

**Validity:** One of the main objectives of this new version of the BDI was to have it conform more closely to the diagnostic criteria for depression, and items were added, eliminated and reworded to specifically assess the symptoms of depression listed in the DSM-IV and thus increase the content validity of the measure. With regard to construct validity, the convergent validity of the BDI-II was assessed by administration of the BDI-1A and the BDI-II to two sub-samples of outpatients (N=191). The order of presentation was counterbalanced and at least one other measure was administered between these two versions of the BDI, yielding a correlation of .93 (p<.001) and means of 18.92 (SD = 11.32) and 21.888 (SD = 12.69) the mean BDI-II score being 2.96 points higher than the BDI-1A. A calibration study of the two scales was also conducted, and these results are available in the BDI-II manual. Consistent with the comparison of mean differences, the BDI-II scores are 3 points higher than the BDI-1A scores in the middle of the scale. Factorial Validity has been established by the inter-correlations of the 21 items calculated from the sample responses.

**Norms:** The normative sample included 500 outpatients from rural and suburban locations. All patients were diagnosed according to DSM-III-R or DSM-IV criteria were used to investigate the psychometric characteristics of BDI-II. The group was comprised of 63% women, and 37% men, the mean age was 37.20 years, range of 13-86 years. The racial/ethnic makeup was 91% White, 4% African American, 4% Asian American, and 1% Hispanic. A student sample of 120 college students in Canada served as a comparative normal group.

**Suggested Use:** The BDI-II is intended to assess the severity of depression in psychiatrically diagnosed adults and adolescents 13 years of age and older. It is not meant to serve as an instrument of diagnosis, but rather to identify the presence and severity of symptoms consistent with the criteria of the DSM-IV. The authors warn against the use of this instrument as a sole diagnostic measure, as depressive symptoms may be part of other primary diagnostic disorders.
Beck Hopelessness Scale

**Purpose:** Designed to measure negative attitudes about the future.

**Population:** Ages 17 and over.

**Score:** A total scale score.

**Time:** (5-10) minutes.

**Author:** Aaron T. Beck.

**Publisher:** The Psychological Corporation.

**Description:** The Beck Hopelessness Scale (BHS) is a 20-item scale for measuring negative attitudes about the future. Beck originally developed this scale in order to predict who would commit suicide and who would not. The conceptual basis for the scale derives from the writings of the social psychologist Ezra Stotland.

**Scoring:** Scoring is straightforward; one simply adds up each of the items marked in the direction keyed for “hopelessness.” Using the scoring template, one counts the number of blackened circles that show up under the circles on the template. The user bases his or her interpretation on the total scale score. The manual contains general cutoff guidelines, although the author recommends that cut-off scores should be based upon clinical decisions.

**Reliability:** The manual reports KR-20 coefficients (measures of the scale’s internal consistency) ranging from .82 to .93. The manual states that when 21 patients with mixed diagnoses were tested at the Center for Cognitive Therapy both during an intake evaluation and 1 week later, before beginning their therapy, the correlation between their scores on the two occasions was .69. In another sample of patients from the Center for Cognitive Therapy (N = 99), the test-retest reliability over a 6-week span was .66. Both of these test-retest coefficients are statistically significant.

**Validity:** Concurrent validity. Beck examined the relationship between clinical ratings of hopelessness and BHS scores in two samples: a) 23 outpatients in a general medical practice and b) 62 hospitalized patients who had recently attempted suicide. In the general practice sample, the correlation between the BHS and the ratings of hopelessness was .74; in the suicide-attempt sample, it was .62.

**Norms:** The normative sample consisted of 294 psychiatric inpatients who had made recent suicide attempts. This sample included 125 men and 169 women. Of the total sample, 150 were white, 139 were black, and 5 belonged to other racial groups. The average age was 29.9 years. On average, the participants had finished 10.85 years of school. As to marital status, 41.5% were single, 17.3% were married, 31.6% were separated or divorced, 2.7% were widowed. and the rest were unmarried but living together or of unknown status.

**Suggested Uses:** The BHS is recommended for measuring extent of negative attitudes in clinical and research settings.
Bender Gestalt Monograph

**Purpose:** Published to disseminate unusual techniques and research of clinical interest in the Gestalt tradition.

**Population:** N/A.

**Score:** N/A.

**Time:** N/A.

**Author:** Lauretta Bender.

**Publisher:** The American Orthopsychiatric Association, Inc.

**Description:** This book is authored by the creator of the Bender-Gestalt test. The author’s motivation for this published work is the “mass of data that has been accumulated which is offered as a contribution to the psychology of gestalt, to the psychology of the functioning of the personality, and to the science of clinical medicine. Furthermore, it will be the effort of the author to indicate in what way this new data adds new understanding to the functions of the normal mind and to various forms of pathological mental states.” Bear in mind that this “new” data the author was referring to was published in before 1938. This book illuminates the processes inherent in Gestalt theory and has tremendous clinical applicability for clinicians using the Bender-Gestalt test.

**Scoring, Reliability, Validity, Norms, Suggested Uses:** N/A.
**Bender-Gestalt Test**

**Purpose:** A rapid, efficient measure of perceptual-motor and cognitive development in children.

**Population:** Ages 4 and over.

**Time:** (10) minutes.

**Author:** Lauretta Bender.

**Publisher:** The American Orthopsychiatric Association, Inc.

**Description:** The Bender Visual Motor Gestalt Test (Bender-Gestalt) is the most frequently administered and thoroughly researched of all the drawing (copying) tests. It consists of nine geometric designs (numbered A and 1-8) originally developed by Wertheimer to demonstrate the perceptual tendencies to organize visual stimuli into configural wholes (Gestalten). Each design is presented sequentially to the subject whose task is to reproduce them on a blank sheet of paper.

**Scoring:** Scoring is usually relatively easy and rapid, rarely requiring more than three or four minutes, regardless of whether a formalized or intuitive scoring system is employed.

**Reliability:** The results involving the Bender with young children reveal interscorer reliability to be very high with correlations of .90 and above. Test-retest reliability coefficients with children range from a low of about .50 with kindergarten children measured 8 months apart to .90 with the same age group measured two weeks apart. The majority of more than 20 different reliability studies reported by Koppitz reveal correlation coefficients in the .80+ range and suggest that normal elementary school children show relatively stable patterns of Bender-Gestalt scores from one administration to the next.

**Validity:** With respect to the validity of the Bender with children, Koppitz reported correlation coefficients from about .50 to as high as .80 between the Bender-Gestalt and intelligence as measured by the Stanford-Binet or Wechsler Intelligence Scale for Children up to about the age of 10. Beyond this age the correlations drop to essentially zero as most older children obtain nearly perfect scores. She also reported relatively high correlations between Bender scores and subsequent educational achievement of first-grade children. Koppitz also reported a relatively high correlation between the Bender and intellectual and academic performance for retarded children as well. With children diagnosed as having minimal brain damage, she reported that the Bender is a valuable diagnostic tool but cautioned that it should not be used alone but in combination with other psychological tests and any background information available.

**Norms:** Norms for a wide variety of clinical groups, including mentally retarded, organically brain-damaged, psychotic, and normal adults are included in Bender’s classic research monograph.

**Suggested Uses:** Designed for use in educational, research, and clinical settings.
**Bennett Mechanical Comprehension Test**

**Purpose:** Designed to measure abilities related to mechanical comprehension.

**Population:** Individual applicants for mechanically-related jobs or schooling.

**Score:** Percentile score based on the appropriate normative population.

**Time:** (30) minutes.

**Author:** George K. Bennett.

**Publisher:** The Psychological Corporation.

**Description:** The Bennett Mechanical Comprehension Test (BMCT) measures a complex set of abilities composed of three primary facets or constructs: 1) mechanical information, 2) spatial visualization, and 3) mechanical reasoning or understanding. The knowledge measured by this test is based on common experiences, and special training seems to have little effect on test scores.

**Scoring:** The BMCT is scored by hand with a scoring template. The raw score is simply the number correct. Raw scores are converted to percentile scores based on the appropriate normative population. Scoring instructions for the BMCT are clear and easy to master. Scoring the 68 items should take an experienced scorer no longer than 2 minutes. Interpretation of the BMCT is based on an objective raw score (number correct) that is converted to a percentile score. The percentile score represents the percent of a similar population that would have scored at or below the raw score attained by the candidate.

**Reliability:** The manual for the BMCT presents split-half reliability coefficients (corrected by the Spearman-Brown formula) of .81 to .93, with a median of .86. The standard errors of measurement range from 3.0 to 3.8. The manual indicates that the reliability of the BMCT is low because of the lower mean item intercorrelations, due to the variety of topics covered in the 68 items. Test-retest reliability data are not presented in the 1969 manual.

**Validity:** Content validity is not discussed in the manual because there is no specific job or curriculum against which to compare the test items. Construct validity is also not specifically addressed in the 1969 manual. A table of correlations of the BMCT with other standardized tests from an industrial data base is presented. For applicants for skilled trade jobs in an automobile company, the BMCT correlates .46 with the Personnel Test (mixed forms) and .55, .48, and .59 with the Wesman Personnel Classification Test (Forms A and B) Verbal, Numerical, and Verbal and Numerical scores, respectively. For applicants for electrical inspector trainee jobs in an aviation company, the BMCT correlated .57 and .68 with the Verbal and Numerical scales of the Personnel Tests for Industry (Forms A and B) and .15 and .04 with the Numbers and Names subtests of the Minnesota Clerical Tests.

**Norms:** To develop Forms S and T, 180 experimental items (95 existing items, 43 original items that had been modified, and 42 new items) were administered to 706 male students in Grades 11 and 12 in Philadelphia, Pennsylvania, and Kansas City, Missouri.

**Suggested Uses:** Recommended uses include the prediction of job performance.
**Benton Visual Retention Test**

**Purpose:** Designed to assess visual perception, visual memory, and visuoconstructive abilities.

**Population:** Ages 8 and over.

**Score:** Difference between error score and correct score.

**Time:** Not reported.

**Author:** Arthur Benton.

**Publisher:** The Psychological Corporation.

**Description:** The Benton Revised Visual Retention Test is a widely used instrument that assesses visual perception, visual memory, and visuoconstructive abilities. Because it measures perception of spatial relations and memory for newly learned material, it is used in clinical diagnosis of brain damage and dysfunction in children and adults, as well as in research. The Benton, as it is usually called, has three alterate forms, each of which consists of ten designs. In addition, there are four possible modes of administration.

**Scoring:** Test interpretation is based on an assessment of the number and types of errors made and involves several levels of analysis for diagnostic purposes. The examiner compares the examinee’s obtained scores with the expected scores found in the norm tables. When examining the difference between these scores for the number correct, the wider the discrepancy in favor of the expected score, the more probable it is that the examinee has suffered neurological impairment.

**Reliability:** The interscorer agreement for total error score is high ($r = .95$) and for major categories of errors reliability is moderate to high ($r = .66$ to $.97$). The category of errors having the lowest interscorer reliability is substitutions (.66), whereas distortion and omission/addition categories produce a correlation of .75. All other categories have correlations over .85. Alternate form reliability for Forms C, D, and E is high ($r = .85$) for Administration A. The multiple-choice forms [F and G] are reported to have a moderate internal consistency, split-half reliability is .76. To determine test-retest reliability, Administration C was given to 194 Kindergarten children twice in a four-month interval. The correlation of .75 between the two sets of scores is very promising.

**Validity:** A correlation of .42 was found between the Benton and the Digit Span WAIS subtest. This low correlation indicates discriminate validity since the Benton was created to supplement the Digit Span test. Various studies have examined the ability of the Benton to assist in the diagnosis. Using a cutoff score of -3, the test identified as brain-injured 22% of the suspected brain injured, 24% of the true brain injured, and 6% of normals.

**Norms:** 600 people were used in the norming of this test, but the manual does not indicate the demographic distribution of its sample.

**Suggested Uses:** The Benton is recommended for use as part of a neuropsychological battery to assess specific dysfunction.
Boehm Test of Basic Concepts--Revised

**Purpose:** Designed to measure the basic positional concepts of young children.

**Population:** Kindergarten through second grade.

**Score:** Percentile or Total Group scores.

**Time:** (45) minutes.

**Author:** Ann E. Boehm.

**Publisher:** The Psychological Corporation.

**Description:** The Boehm Test of Basic Concepts-Revised (BTBCR) was developed to measure the understanding of basic positional concepts of young children, that is, whether individuals in kindergarten, Grade 1, and Grade 2 can correctly identify a picture from among a choice of three when presented with verbal cues incorporating such terms as over, least, left, and so on. Based on these results, children can be identified as deficient in conceptual development, with resultant curricular implications.

**Scoring:** Once scored, the results can be translated into percentiles, which may be done by socioeconomic level, or by the total group.

**Reliability:** A number of measures of the reliability of the instrument are presented. Correlations reported between Forms C and D (N=625) are .82 for Kindergarten, .77 for Grade 1, and .65 for Grade 2. Correlations of Forms C and D with Form A of the BTBC (N=173) are .65 and .62, respectively. Split-half reliability coefficients and standard errors of measurement are presented by form, grade, and socioeconomic level. These relationships vary between .55 and .87, with the coefficients being lower at Grade 2 than at Kindergarten and Grade 1. Stability was measured through test-retest with a 1-week interval (N = 548) and varies between .55 and .88.

**Validity:** Validity is measured through the relationship of scores on the Boehm-R and various measures of academic achievement (the total batteries of the Comprehensive Test of Basic Skills, the California Achievement Test, and the Iowa Test of Basic Skills), as well as through reading level as measured on the Harcourt Brace Jovanovich Bookmark Reading Program. The coefficients of correlation varied from .24 (Grade 2 applications with ITBS) through .64 (Kindergarten Form D with CTBS), with a median of .44.

**Norms:** The norming of the Boehm-R occurred in 1983, involving approximately 10,000 children in public school districts selected as nationally representative by size of the district and geographic location. Socioeconomic status was estimated based on each building’s percentage of students who qualified for subsidized lunches, not on the characteristics of the individual children.

**Suggested Uses:** Recommended for assessment of receptive conceptual development in educational and research settings.
The Booklet Category Test

**Purpose:** Predictor of general neurological dysfunction.

**Population:** Adolescents and adults.

**Score:** Not reported.

**Time:** Not reported.

**Authors:** Nick DeFilippis and Elizabeth McCampbell.

**Publisher:** Psychological Assessment Resources.

**Description:** This test is a booklet version of the Category Test originally introduced by Halstead and revised by Reitan as a subtest within the Halstead-Reitan Neuropsychological Battery. The Category Test is often seen as a rather complex concept formation test that is not especially difficult for normal subjects, but requires abstraction ability that is often diminished in subjects with cortical damage. According to Reitan and Davison, the purpose of the test is to “determine the ability of the subject to profit from both negative and positive experience as a basis for altering his [sic] performance.”

**Scoring:** Patients must attempt to abstract the organizing principle or concept involved in seven groups of stimuli presented on sheets of paper in a binder. The number of errors made by the subjects is recorded as a measure of the individual’s ability to profit from experience. Difficulty exist in interpretation because it is unlikely that any two subjects would exhibit the same pattern of positive and negative feedback.

**Reliability, Validity, and Norms:** The manual relies primarily upon the body of research developed with the Category Test. The manual reports the correlations between the Booklet Category Test and the Category Test to range from .91, for 30 normal students, to .76, for a heterogeneous group of 38 psychiatric patients. While promising, these data are less than sufficient in establishing the equivalence of the two tests. Although the relationship between the Category Test and the booklet version may be viewed as an estimate of its stability-equivalence, no formal reliability data are reported. The manual also lacks any standardization or normative data.

**Suggested Uses:** The Booklet Category Test offers a promising adaption of the Halstead-Reitan version which reduces equipment cost and increases the portability of the test.
Boston Diagnostic Aphasia Examination

Purpose: Designed as a comprehensive measure of aphasia.

Population: Adults.

Score: Percentiles or standard scores for the subtests.

Time: (180) minutes.

Authors: Harold Goodglass and Edith Kaplan.

Publisher: The Psychological Corporation.

Description: The Boston Diagnostic Aphasia Examination is a comprehensive, multifactorial battery designed to evaluate a broad range of language impairments that often arise as a consequence of organic brain dysfunction. The Examination is designed to go beyond simple functional definitions of aphasia into the components of language dysfunctions (symptoms) that have been shown to underlie the various aphasic syndromes. Thus, this test evaluates various perceptual modalities (e.g., auditory, visual, and gestural), processing functions (e.g., comprehension, analysis, problem-solving), and response modalities (e.g., writing, articulation, and manipulation). This approach allows for the neuropsychological analysis and measurement of language-related skills and abilities from both ideographic and nomothetic bases, as well as a comprehensive approach to the symptom configurations that relate to neuropathologic conditions.

Scoring: The manual provides clear statements and rules for scoring protocols.

Reliability: Reliability of the subtests was studied by selecting protocols of 34 patients with a degree of severity of aphasia ranging from slight to severe. Kuder–Richardson reliability coefficients for subtests ranged from .68 to .98, with about two-thirds of the coefficients reported ranging from .90 upwards. Since test-retest reliability is difficult if not impossible to attain with patients suffering from aphasic symptoms, the current reliability coefficients demonstrate very good internal consistency in terms of what the items within the subtests are measuring.

Validity: A discriminant analysis comparing “unambiguous exemplars of a single syndrome” was carried out. Thus, unambiguous cases of Broca’s aphasia, Wernicke’s aphasia, conduction aphasia, and anemic aphasia were selected. Ten variables were selected on the assumption of providing the most useful data. From these, five variables were selected for the discriminant analysis (body part identification, repetition of high probability sentences, verbal paraphasias, articulatory agility rating, and automated sentence rating). This classification yielded no misclassifications.

Norms: Standardization of the revised Boston Diagnostic Aphasia Examination is based on a normative sample of 242 patients with aphasic symptoms tested at the Boston VA Medical Center between 1976–1982. It is important to consider that this sample includes only male, presumably English-speaking patients and is thus highly selected.

Suggested Uses: Designed for the assessment of aphasia for inpatient or outpatient populations.
Boston Naming Test

Purpose: Designed to measure object naming from line drawings.

Population: Not reported.

Score: N/A.

Time: Not reported.

Author: Sandra Weintraub.

Publisher: The Psychological Corporation.

Description: The Boston Naming Test (BNT) represents a measure of object naming from line drawings that takes into account the finding that patients with dysnomia often have greater difficulties with the naming of low frequency objects. Thus, instead of there being a simple category of anomia, naming difficulties may be rank ordered along a continuum. Items have been rank ordered in terms of their ability to be named, which is thought to be correlated with their frequency. This type of picture-naming vocabulary test is useful in the examination of children with learning disabilities and the evaluation of brain-injured adults.

Scoring: The BNT contains 60 items.

Reliability, Validity, and Norms: This information is not provided in the manual.

Suggested Uses: Recommended as a supplement to the Boston Diagnostic Aphasia Examination.
BRL Sorting Test

**Purpose:** Designed to assess sorting behavior and its psychological implications.

**Population:** Unknown.

**Score:** Unknown.

**Time:** Unknown.

**Author:** Unknown.

**Publisher:** Unknown.

**Description:** This test attempts to measure level of concept formation through sorting behavior. The authors assert that sorting behavior, in response to the instruction to “put together those which belong together,” is an expression of concept formation. It may take both the form of actively putting together those objects that “belong together,” and the form of discovering why the objects put together “belong together.”

**Scoring:** Scoring takes into account both sorting and verbalization. Specifically, the administrator scores: 1) the adequacy of sorting and verbalization— that is, to what degree the sortings or verbalizations approximate or deviate from the norm on that item; 2) the conceptual level of verbalization— that is, whether the definition of the group was on an abstract, functional, or concrete level; 3) and the concept span— that is, the patient’s regard for all the objects in the test, and whether he or she omitted or included too many objects in the group that was sorted.

**Reliability, Validity, and Norms:** The manual contains no information on reliability, validity, and norms.

**Suggested Uses:** Suggested uses include assessment of concept formation in a research setting.
California Psychological Inventory

**Purpose:** Designed to evaluate interpersonal behavior and social interaction within normal individuals.

**Population:** Ages 13 and over.

**Scores:** Inventory produces 18 scale scores.

**Time:** (45-60) minutes.

**Author:** Harrison Gough

**Publisher:** Consulting Psychologists Press, Inc.

**Description:** The purpose of the CPI scales is stated by the author as follows: "Each scale is designed to forecast what a person will say or do under defined conditions, and to identify individuals who will be described in characteristic ways by others who know them well or who observe their behavior in particular contexts. The scales are grouped for convenience into four broad categories, bringing together those having related implications. The underlying logic here is interpretational, not factorial, i.e., these four categories do not necessarily constitute psychometric entities."

**Scoring:** The items on the inventory produce scores for 18 scales, which are divided into four classes: measures of poise, ascendancy, self-assurance and inter-personal adequacy; measures of socialization, responsibility, intra-personal values, and character; measures of achievement potential and intellectual efficiency; and measures of intellectual and interest modes.

**Reliability:** The manual contains no information on reliability.

**Validity:** Correlations between individual CPI scales and relevant external criteria fall in the .2 to .5 range. Such relationships are typical in personality research, and extremely high correlations are unlikely to be found since the scales are developed to assess rather broad behavioral tendencies. The manual contains no information on item intercorrelations and factorial analyses.

**Norms:** The manual contains no information on norms.

**Suggested Uses:** Using the CPI as a research tool and as an assessment device for adolescents and adults have been popular among psychologists.
Carlson Psychological Survey

Purpose: Designed as “a psychometric instrument intended primarily for individuals accused or convicted of crimes, or otherwise referred for socially deviant behavior.”

Population: Criminal offenders.

Score: Five scale scores.

Time: (10-20) minutes.

Author: Kenneth A. Carlson

Publisher: Research Psychologists Press, Inc.

Description: The primary intent of the Carlson Psychological Survey (CPS) is for initial assessment and classification of incarcerated male adults and secondarily for non-incarcerated offenders on probation.

Scoring: The CPS has four scales referred to as Chemical Abuse, Thought Disturbance, Antisocial Tendencies, and Self-Deprecation and also an added Validity Scale. The scoring procedure involves summation of the scale values of the alternatives checked [1 to 5] and then transposing these raw scores to the profile sheet provided. The author has provided 18 profile “Types” or score patterns based on multivariate analysis. Predictions are also provided regarding the intrastititutional security status most appropriate for these subjects, possibility of parole violation, and post-release adjustments.

Reliability: The test-retest reliability values (N=32), with an interval of 2 weeks for the four scales, range from .97 to .92; for the Validity scale, it is .49. Faking studies indicate that the “CPS appears susceptible to faking.”

Validity: Correlations between the CPS and MMPI scales and intergroup comparison of scores between correctional officers and offenders, as well as data on offenders from three different prisons, confirm the solidity of this instruments psychometric base.

Norms: The standardization sample consists of 412 adult male subjects, all inmates in an Ontario correctional center.

Suggested Uses: It is suggested that the CPS is a psychometrically sound instrument for measuring certain aspects of the personality of adult males incarcerated in correctional centers for purposes of assessment and providing possible psychological help.
Cattell Infant Intelligence Scale

Purpose: Designed to assess mental ability in children.

Population: Ages 2-30 months.

Score: Mental Age score.

Time: Not reported.

Author: Psyche Cattell.

Publisher: The Psychological Corporation.

Description: The Cattell Infant Intelligence Scale was developed with the intention of being a standardized assessment of mental ability for children aged 2-30 months. Underlying the test is a view of intelligence as being maturationally and genetically controlled. The Cattell proposes to focus on mental development and not on motor development, to be standardized, be objective in scoring, appeal to young children, and provide numerical rather than simply descriptive assessments of mental ability.

Scoring: The test consists of 95 items: five for each month period from 2-12 months, five for each two-month period during the second year of life, and five for each of the two quartiles of the first half of the third year of life. At each of these age periods, one or two alternate items, which are described as acceptable but somewhat less satisfactory than the standard items, are also included for use as needed. All 95 items are not administered to a given child at a given test administration or age level; only the items necessary to establish basal and ceiling levels are presented in order to obtain a child’s mental age score.

Reliability and Validity: At three months, the reliability estimate obtained by the split-half method and corrected by the Spearman-Brown formula was 0.56. Correlating those scores obtained at three months with scores at 36 months on Form L of the Stanford-Binet resulted in a value of only 0.10. Thus, the scores at age three months show neither statistical reliability or predictive validity. For 6, 9, and 12 months the split-half reliability coefficients were .88, .86, and .89, respectively, and correlations with the Stanford-Binet at 36 months were .34, .18, and .56, respectively. Although reliability estimates reach acceptable levels, the predictive validity scores do not. At ages 18, 24, and 30 months the reliability coefficients were .90, .85, and .71, respectively, with correlations with the Stanford-Binet for 36 months falling at .67, .71, and .83, respectively. Again, acceptable reliability estimates are achieved with the exception of the 30-month score. The predictive validity score at 24, and possibly at 18 months, appears acceptable, but the score at 30 months must be interpreted cautiously, given the low level of reliability at that age.

Norms: The Cattell was developed and tested on a standardization sample of 2,346 examinations made on 274 children enrolled in the Normal Child Series study at the Center for Research in Child Health and Development at the Harvard School of Public Health. Minorities were not included in this sample. This restricted sample, which is neither socioeconomically, ethnically, nor geographically representative of the United States population, severely restricts the confidence with which one can usefully apply the scale and interpret the obtained results.

Suggested Uses: The Cattell is recommended for use in clinical, educational, and research settings.
**Child Behavior Checklist**

**Purpose:** Designed to assess “social competence” and “behavior problems” in children.

**Population:** Ages 4-18.

**Score:** Five scale scores.

**Time:** Not reported.

**Authors:** Thomas M. Achenbach and Craig Edelbrock.

**Publisher:** Thomas M. Achenbach.

**Description:** The Child Behavior Checklist (CBCL) was designed to address the problem of defining child behavior problems empirically. It is based on a careful review of the literature and carefully conducted empirical studies. It is designed to assess in a standardized format the behavioral problems and social competencies of children as reported by parents.

**Scoring:** The CBCL can be self-administered or administered by an interviewer. It consists of 118 items related to behavior problems which are scored on a 3-point scale ranging from not true to often true of the child. There are also 20 social competency items used to obtain parents’ reports of the amount and quality of their child’s participation in sports, hobbies, games, activities, organizations, jobs and chores, friendships, how well the child gets along with others and plays and works by him/herself, and school functioning.

**Reliability:** Individual item intraclass correlations (ICC) of greater than .90 were obtained “between item scores obtained from mothers filling out the CBCL at 1-week intervals, mothers and fathers filling out the CBCL on their clinically-referred children, and three different interviewers obtaining CBCLs from parents of demographically matched triads of children.” Stability of ICCs over a 3-month period were .84 for behavior problems and .97 for social competencies. Test-retest reliability of mothers’ ratings were .89. Some differences were found between mothers’ and fathers’ individual ratings.

**Validity:** Several studies have supported the construct validity of the instrument. Tests of criterion-related validity using clinical status as the criterion (referred/non-referred) also support the validity of the instrument. Importantly, demographic variables such as race and SES accounted for a relatively small proportion of score variance.

**Norms:** Normative data, obtained from parents of 1,300 children, were heterogeneous with respect to race and socioeconomic status and were proportionate to the composition of the general U.S. population.

**Suggested Uses:** It is suggested that the CBCL is a viable tool for assessing a child’s behaviors, via parent report, in a clinical or research environment.
Child Neuropsychological Questionnaire

**Purpose:** Designed to evaluate children suspected of having brain dysfunctions.

**Population:** Children.

**Scores:** Overall evaluation of neuropsychological impairment.

**Time:** Not reported.

**Author:** Fernando Melendez.

**Publisher:** Psychological Assessment Resources, Inc.

**Description:** The Child Neuropsychological Questionnaire consists of 42 items that cover pre-, peri-, and postnatal problems, developmental milestones, medical events, psychosocial development, and educational history. The manual author suggests the aim is to “try to make a three-way differentiation between entities that generally show very similar signs and symptoms; childhood schizophrenia, abnormal electrical functions, and posterior fossa neoplasms.” The questions cover the standard information that should be obtained in a diagnostic interview that is directed toward a child’s psychiatric or neuropsychological problems.

**Scoring:** The Questionnaire consists of 42 items; 7 require narration and 35 require only a yes/no answer. The rationale for each of the questions and suggestions concerning the possible diagnostic significance of each item and the type of follow-up that might be needed are provided in the manual.

**Reliability, Validity, and Norms:** The manual contains no reliability, validity and normative data.

**Suggested Uses:** Suggested uses include differential diagnosis in pediatric neuropsychology or as a training device to train students and to alert them to the possible significance of various behaviors and the reasons for pursuing diverse areas of inquiry.
**Childhood Trauma Questionnaire**

**Purpose:** Designed as a self-report inventory to screen for a history of abuse and neglect.

**Population:** Ages 12 and up.

**Score:** Emotional Abuse, Physical Abuse, Sexual Abuse, Emotional and Physical Neglect, and Minimization–Denial of Abuse.

**Time:** 5 minutes.

**Author:** David P. Bernstein, Laura Fink

**Publisher:** The Psychological Corporation

**Description:** The Childhood Trauma Questionnaire is a 28-item self-report retrospective inventory intending to measure childhood or adolescent abuse and neglect. It is straightforward and easy to use. The CTQ can be administered individually or to a group. The examinee responds to 28 simple questions on a 5-point Likert scale ranging from Never True to Very Often True. The central constructs underlying the questionnaire are emotional, physical neglect and abuse, and sexual abuse. Other traumatic events that may occur during childhood, such as the death of a parent or a major illness, are not assessed. The items are written at a sixth grade reading level and reading level and intellectual functioning should be assessed before administering the scale.

**Scoring:** The CTQ contains five subscales, three assessing abuse (Emotional, Physical, and Sexual) and two assessing neglect (Emotional and Physical). Each subscale has five items and there is a three-item Minimization-Denial subscale to check for extreme response bias, specifically attempts by respondents to minimize their childhood abuse experiences. A 5-point frequency of occurrence scale is utilized: (1) never true, (2) rarely true, (3) sometimes true, (4) often true, and (5) very often true. Each subscale score ranges from 5 (no history of abuse or neglect) to 25 (very extreme history of abuse and neglect).

**Reliability:** Internal consistency coefficients are offered for all of the subsamples with generally favorable patterns reported. The three largest subsamples, the Sexual Abuse (alphas of .93 to .95) and Emotional Neglect (alphas of .88 to .92) are the most reliable. Emotional Abuse (alphas of .84 to .89) and Physical Abuse (alphas of .81 to .86) have acceptable reliabilities. The internal consistency of Physical Neglect (alphas of .63 to .78) is marginal. Test-retest reliabilities with testing over an average 3.6-month period yielded stability coefficients near .80, suggesting good consistency of responses over time.

**Validity:** The results of confirmatory factor analyses, which tested the goodness of fit of the five-factor CTQ subscale model for the adult substance abusers, the adolescent psychiatric inpatients, and a subsample of the HMO members, showed structural invariance across the three samples suggesting that they measured the same constructs across groups. There are good correlations between scores on the CTQ and ratings derived from semistructured interviews administered by clinicians and ratings by therapists. The correspondence of the CTQ scores with ratings of Sexual Abuse and Physical Abuse seem particularly noteworthy (typically correlations from .50 to .75).

**Norms:** Norms were derived from six samples. Only three had more than 300 subjects: (a) 378 mostly Black, male inpatient substance abusers; (b) 398 adolescent psychiatric inpatients; and (c) 1,225 all female, mostly White HMO members. These three subsamples comprise 2,001 of the 2,201 individuals in the CTQ norm group. The first two subsamples are from New York psychiatric facilities and the latter sample was from the Pacific Northwest. Despite using nonrepresentative samples, the authors offer “norms” for each of these subsamples. The adult substance user norms present percentile ranks derived from just 58 female respondents. Percentile ranks are also provided for adolescent psychiatric patients and are derived from just 171 males and 227 females. The distributions of scores, even in these high-risk samples, were quite skewed. The CTQ norms were also used to create the severity classification categories of (1) None or Minimal, (2) Low to Moderate, (3) Moderate to Severe, and (4) Severe to Extreme. How these classifications apply to regional groups in the U.S. is completely unknown. These categories should be used with extreme caution.

**Suggested Use:** The stated uses of the CTQ are (a) for rapid abuse-history taking for treatment planning, (b) to encourage disclosure of childhood abuse as part of clinical assessments, and (c) for use in epidemiological and correctional studies involving childhood abuse. As a screening tool to identify individuals with a childhood history of abuse and neglect, this device may have utility.
Children’s Apperception Test

**Purpose:** Designed as a projective method of describing personality.

**Population:** Ages 3 to 10 years.

**Score:** N/A.

**Time:** (30) minutes.

**Authors:** Leopold Bellak and Sonya Sorel Bellak.

**Publisher:** C.P.S., Inc.

**Description:** The Children’s Apperception Test (CAT-A) is a projective method of describing personality by studying individual differences in the responses made to stimuli presented in the form of pictures of animals in selected settings. The 10 items consist of 10 scenes showing a variety of animal figures, mostly in unmistakably human social settings. The use of animal rather than human figures was based on the assumption that children of these ages would identify more readily with appealing drawings of animals than with drawings of humans. The author discusses interpretation on the basis of psychoanalytic themes, but there is no compelling reason that Children’s Apperception Test protocols could not be interpreted from other theoretical frameworks.

**Scoring:** This projective technique is not “scored” in a quantitative sense. The gist of stores is recorded, and the presence or absence of thematic elements is indicated on the form provided.

**Reliability and Validity:** No statistical information is provided on the technical validity and reliability of the CAT.

**Norms:** Information on norms is not included in the manual.

**Suggested Uses:** Designed for use in clinical and research settings.
Children’s Depression Inventory (CDI)

**Purpose:** Designed for assessment of self-rating of depressive symptoms in children.

**Population:** Children, aged 7 through 17 years.

**Score:** Total Score, Five Scale A: Negative Mood, B: Interpersonal Problems, C: Ineffectiveness, D: Anhedonia, and E: Negative Self Esteem.

**Time:** [15] minutes or less.

**Author:** Maria Kovacs

**Publisher:** Mental Health Systems

**Description:** The CDI has been designed to measure self-rated assessment of depressive symptoms for school aged children and adolescents. There are 27 items quantifying symptoms such as depressed mood, hedonic capacity, vegetative functions self-evaluation and interpersonal behaviors. It covers the consequences of depression as they relate to children and functioning in school and with peers. A short form with 10 items can be used when a quick screening is necessary. While both forms are reported to give comparable results, the longer form provides factor scores and generally gives a more robust description of the child’s symptoms. The reading level of the CDI is at the first grade level, the lowest of any measure of depression for children.

**Scoring:** For each item the child has three possible answers; 0 indicating an absence of symptoms, 1 indicating mild symptoms, and 2, definite symptoms. The total score can range from 0 to 54. The CDI can be administered using a QuickScoreTM Form to assist in scoring and transforming the scores to a profile. Factor scores are calculated by adding the scores for each letter assigned to each item. Total scores and factor scores converted to t scores on the profile form. The CDI can also be administered and scored using a microcomputer, or administered using paper forms and scored using the computer program.

**Reliability:** Internal consistency reliability has been found to be good, with coefficients ranging from .71 to .89 with various samples. Test-retest reliability correlations appear to be acceptable. It is however expected that the symptoms of depression would change over time, and regression to the mean is associated with repeated testing over time.

**Validity:** Numerous research studies have supported the CDI as assessing important constructs both for explanatory and predictive uses for characterizing symptoms of depression in children and adolescents. Studies of discriminant validity found significant differences of Negative Mood factor scores ($p < .05$) but no significant difference for total CDI scores among a sample of 134 children and adolescents with various depressive disorders. Some studies report the CDI to successfully distinguish normals from diagnostic categories, while other studies have been less favorable, and it is agreed that more research on the discriminant validity is needed for the CDI.

**Norms:** The normative sample included 1266 public school students from Florida in grades 2 through 8. There were 592 boys between the ages of 7 and 15, and 674 girls ages 7 to 16. Assuming the sample to be representative of the total demographics of the school, it is estimated that 77% were white, and 23% African American, Native American, or Hispanic. The population was mostly middle class, and about 20% of the students were from single homes. Separate norms were developed for two groups based on ages, (7-12, and 13-17) as developmental trends result in higher scores for the older group.

**Suggested Use:** The CDI measures severity of symptoms of depression in children and adolescents. The inventory should be used with other assessment instruments for diagnosis and monitoring treatment progress.
Revised Children’s Manifest Anxiety Scale

**Purpose:** Designed to assess the level and nature of anxiety in children and adolescents.

**Population:** Ages 6 to 19.

**Score:** Physiological Anxiety, worry-Oversensitivity, Social concerns-concentration, Total Anxiety, Lie

**Time:** 10-15 minutes

**Author:** Cecil R. Reynolds and Bent O. Richmond

**Publisher:** Western Psychological Services

**Description:** The Revised Children’s Manifest Anxiety Scale (RCMAS) is a self-report instrument designed to measure anxiety for children and adolescents aged 6-9 years. For children over 9 and a half years of age, it can be administered in a group situation. For first and second graders the examiner should read the items to the child. There are 37 items each of which requires a yes or no answer. The RCMAS was developed in 1978 to address criticisms of the original Children’s Manifest Anxiety Scale (CMAS). Goals for revision of the scale were to (a) create an objective measure of children’s anxiety suitable for group administration; (b) keep administration time to the minimum required for accurate, valid assessment; (c) make the reading level of items suitable for elementary school students but yet allow for use throughout the school years; (d) cover new areas of anxiety and determine whether anxiety would best be treated as unidimensional or multidimensional; (e) increase norms and information for diverse groups of children; and (f) assure that all items are good test items.

**Scoring:** The Total Anxiety score is based upon 28 items with 9 items comprising the Lie Scale. The Total Anxiety score and the Anxiety subscale scores are determined by the number of ‘yes’ responses to the anxiety items. The Lie Score is determined by ‘yes’ responses to the Lie subscale items and is used to determine if the child was making a valid attempt to respond. The three anxiety subscales should be interpreted cautiously and should be used only as an aid in hypothesis generation due to limited reliability levels. The Total Anxiety Score is expressed as a T score (M=50, sd=10) and the subscales are expressed as scale scores (M=10, sd = 3). Percentile ranks are provided for each of the RCMAS scores. Norms are provided at 1-year intervals and for each ethnic-sex combination for blacks and whites. The Lie Scale is a positive feature of the instrument and is designed to detect acquiescence, social desirability, or faking of responses.

**Reliability:** Reliability estimates for internal consistency for white and black males and females for each of the 12 age levels ranged from .42 (Age 6, black females) to .87 (Age 12, black males and Age 15, black males). Reliability estimates collapsed across the 12 age levels ranged from .79 to .85 (median=.82). Of the 48 coefficient alphas reported across age, race, and sex, 17 alphas fall below .80, but in general the Total Anxiety score is reliable enough across most ages to permit reliable interpretations. It should be noted however, that for black females; at ages 6, 8, 10, and 11, reliability coefficients are significantly lower than for white females at the same ages, and the authors urge that special care be taken when using the RCMAS with black females under 12 years of age.

Coefficient alphas are also reported for each of the three anxiety subscales and the Lie Scale. Most of these alphas fall below .80, some far below (.15) indicating poor domain sampling. Individual anxiety subscales are too unreliable to merit interpretation for most persons in the standardization sample. One study reported in the manual found a stability coefficient of .68 for the Total Anxiety score and .58 for the Lie Scale score for 534 children tested 9 months apart. Stability estimates increase dramatically when the interval between testing is shorter. All stability estimates reported in the manual are for children below grade 7. No stability data are reported for junior and senior high school students.

**Validity:** The manual reports higher correlations between the RCMAS Total Anxiety score and a measure of trait anxiety (r=.67) than between Total Anxiety and a measure of state anxiety (r=.10) despite the questions about its measure of trait anxiety raised by the low test-retest stability coefficients. Other than this, the validity data is sparse and the factors appear theoretically unrelated. Empirically, the correlation between the RCMAS factors range from .49 to .85 (median=.67), suggesting correlated factors and an oblique solution. However, the factor analysis of the instrument appears to be flawed. The standardization data should be refactored before test users can have faith in the dimensions underlying the RCMAS.

**Norms:** The RCMAS was standardized on 4,972 children between the ages of 6 and 19 years. The standardization sample contained 2,208 white males (44%), 2,176 white females (44%), 289 black males (5.8%), and 299 black females (6%). The sample was obtained from children in 13 states and more than 80 school districts from all major geographic regions of the U.S. The sample appears to be representative of the U.S. population, and the authors believe that “an excellent cross-section of school children was obtained” even though the manual provides no data comparing it to census data. Approximately 600 children in the sample were classified as handicapped (i.e., educably mentally retarded, learning disabled, or gifted). Users of the instrument can, however, be reasonably confident the RCMAS has a well-standardized and representative sample.

**Suggested Use:** With future work with the scale, particularly the factor analysis undertaken the RCMAS may develop into one of the better measures of anxiety in children and adolescents.
Clinical Analysis Questionnaire

**Purpose:** Designed as a shortened version of the Sixteen Personality Factor Questionnaire.

**Population:** Ages 16 and over.

**Score:** Twenty-eight primary factor scores and nine second-order factor scores.

**Time:** (120) minutes.

**Author:** Samuel E. Krug.

**Publisher:** Institute for Personality and Ability Testing.

**Description and Scoring:** The Clinical Analysis Questionnaire (CAQ) has been developed in the tradition of the 16 PF and by the same organization. Part I of CAQ is an abbreviated, simplified version of the 16 PF, measuring the same 16 factors with half of its items the same as those in Form A of the 16 PF. Part II consists of 144 items measuring 12 factorially defined clinical dimensions. This part was designed to measure dimensions of pathology and to perform functions such as those done by the MMPI.

**Reliability:** Test-retest reliabilities range from .51 to .74 in Part I and from .67 to .90 in Part II.

**Validity:** The author of the manual argues that validity is defined as the correlation of a scale with the underlying factor it is designed to measure. Given this definition, validity coefficients are high, often higher than reliability. But this is not the usual meaning of validity.

**Norms:** Information concerning normative data was not included in the manual.

**Suggested Uses:** It is suggested, as described in the manual, that the CAQ should be seen as a research device.
College and University Environment Scales

Purpose: Designed to assess students’ conceptions of the prevailing atmosphere of the campus.

Population: College students.

Score: Yields seven scores.

Time: (30) minutes.

Author: C. Robert Pace.

Publisher: Educational Testing Service.

Description: The purpose of the College and University Environment Scales (CUES) is to help institutions define the cultural, social, and intellectual climate of the campus. The instrument is composed of 160 items, that may be administered to individuals or groups, that are grouped into seven scales.

Scoring: The CUES yields seven scores: practicality, community, awareness, propriety, scholarship, campus morale, quality of teaching and faculty-student relationships. The CUES provides only group scores. Group scores are obtained by adding the number of items answered by 66 percent or more of the students in the keyed direction, subtracting the number of items answered by 33 percent or fewer of the students in the keyed direction, then adding 20 points to avoid the possibility of a negative score.

Reliability: The reliability of CUES scores as measures of institutional differences was determined by means of Cronbach’s coefficient alpha. These range from .89 to .94. The manual reports test-retest comparisons made from comparable samples of reporters over a one- or two-year period or comparisons of scores for different groups judged to be qualified reporters have been summarized for 25 different colleges and universities. The finding is that of different groups within a single institution 80 percent differed by 3 points or less and 90 percent differed by 4 points or less.

Validity: The manual provides an extensive set of tables showing intercorrelations between CUES scale scores, college aptitude measured by mean SSAT scores of entering freshmen, and a wide variety of other factors. Although these various relationships are reasonably congruent with expectations, reviewers have noted that this information still leaves unsettled the question of just what the college environment is and what the CUES is measuring.

Norms: Acquiring normative data required a compromise between two approaches involving a straightforward averaging of the numbers or proportions for a sample of 100 institutions categorized in terms of region, level or program, form of control (public or private), and type of institution (liberal arts, teachers’ college, etc.).

Suggested Uses: The CUES is recommended as a useful resource for systematically gathering data on student perceptions of the college environment and its impact on them.
Columbia Mental Maturity Scale

Purpose: Designed to assess the reasoning ability of children.

Population: Ages 3 to 9.

Score: Age Deviation Score.

Time: [15-20] minutes.

Authors: Bessie Burgemeister, Lucille H. Blum, and Irving Lorge.

Publisher: The Psychological Corporation.

Description: The Columbia Mental Maturity Scale (CMMS) is an individually administered instrument designed to assess the general reasoning ability of children between the ages of 3 years, 6 months to 9 years, 11 months. The CMMS consists of 92 pictorial and figural, classification items arranged in a series of eight overlapping levels. Each of the eight levels contains between 51 and 65 items that are appropriate for a specific chronological age.

Scoring: Administration of the CMMS takes approximately 15 to 20 minutes and yields several scores: raw score, Age Deviation Score, percentile rank, stanine, and Maturity Index. The Age Deviation Score is a standard score with a mean of 100 and a standard deviation of 16. The maturity indexes are comparable to mental ages, although they are more global, employing the use of ranges of age rather than specific mental ages.

Reliability: The CMMS manual reports both split-half and test retest reliabilities. The split-half reliabilities are reported for each of the 13 age levels, with the items for each age level divided in half forming two half-tests for each age level. The manual reports internal consistency coefficients ranging from a low of .85 to a high of .91 with a median split-half coefficient of .90 for the standardization group, indicating excellent internal consistency. Test-retest reliability coefficients for three different age groups are reported for an interval of 7-10 days. A median test-retest reliability of .85 was obtained.

Validity: The CMMS manual reports correlational data between the CMMS and the subtests of the SAT. These data indicate that the interlevel standard scores of the CMMS correlate substantially with the various subtest scores of the SAT with a median value of .57 (.31 to .61) for all Primary I Battery subtests and a median value of .47 (.43 to .61) for all Primary II Battery subtests. The correlation of CMMS scores and two measures of intellectual ability, the Otis-Lennon Mental Ability Test (OL) and the Stanford-Binet Intelligence Scale (SB), are reported as concurrent validity. The ADS of the CMMS and the Deviation IQ of the Otis-Lennon correlated from .62 to .69. The Deviation IQ score of the SB and the ADS of the CMMS correlated .67 for 52 preschool and first-grade children from a large southern city.

Norms: The CMMS was standardized on 2,600 children stratified on the basis of parental occupation, race, geographic location, and size of residence community. A sample of 200 children were selected for each age level closely reflecting the population indicated by the 1960 U.S.

Suggested Uses: The CMMS is recommended as an instrument for screening the general reasoning ability in young children.
Conners’ Rating Scales - Revised

Purpose: Designed to assess attention-deficit/hyperactivity (ADHD) and related problems in children.

Population: Children, aged 3 years through 17 years.

Score: Short and long versions of parent, teacher, and self-report forms with subscales.

Time: 20 minutes or less, longer for reading disabilities, ESL, or psychiatric problems.

Author: C. Keith Conners

Publisher: MHS

Description: The Connors Rating Scale - Revised (CRS-R) updates the original Conners’ Rating Scale, provides three scales rather than two, with long and short versions of each scale. The “Hyperactivity Index” had been renamed the Conners’ Global Index, and is included on the forms for teachers and parents. While the long versions require more time to complete, they correspond more closely to the criteria of the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV, APA, 1994). The short version is useful when time is limited or when repeated administrations are needed.

Scoring: Computer programs are available for scoring, calculating standardized T-scores from raw scores, and providing graphic display and a report of the results.

Reliability: The coefficient alphas for internal reliability were highly satisfactory for the normative groups. For the long form there was a range from .728 to .942 and .857 to .938 for the short form of the CRS-R, indicating that the CRS-R subscales are accurate in measuring the constructs they were intended to measure.

Validity: One The CRS-R has been compared to the CDI, The CRS, and the CPT overall index. Correlations were also done between the teacher, parent, and adolescent ratings. The results indicate that the CRS-R does in fact identify childhood and adolescent ADHD behavioral problems and psychopathology. Validity studies are continuing.

Norms: The normative sample consisted of over 8000 cases. Data are from over 200 schools in over 45 states and 10 provinces throughout the US and Canada for parents, teacher and self-reports was included.

Suggested Use: The main use of the CRS-R is for the assessment of ADHD. The subscales however provide information useful for assessment of conduct problems, cognitive problems, family problems, emotional, anger control and anxiety problems. The CRS-R can be used for screening, for treatment monitoring, as a research instrument, and as a clinical diagnostic aid.
Differential Ability Scales

**Purpose:** Designed to assess the cognitive ability and achievement of children.

**Population:** Children, aged 2 years 6 months through 17 years 11 months.

**Scores:** General Conceptual Ability, cluster scores and core subtest scores.

**Time:** (25-65) minutes for Preschool level, (40 –65) minutes for School-Age Level on the Cognitive Battery, (15-25) minutes for the School Achievement tests.

**Author:** Colin D. Elliott

**Publisher:** The Psychological Corporation

**Description:** The Differential Ability Scales is an individually administered test battery intending to measure cognitive and achievement levels for children for classification and diagnostic purposes. Its diverse nature makes it possible to profile a child’s strengths and weaknesses. This instrument intends to provide a wider range of measurement possibilities than found in other similar batteries. The tasks and scores of the DAS are said to reflect a wide range of theories to accommodate a variety of theoretical views. It consists of 20 subtests, 17 cognitive and 3 achievement subtests yielding an overall cognitive ability score and achievement scores. Differences between cognitive abilities and between cognitive ability and achievement can be explored. Harder or easier sets of items can be administered if high or low ability is expected in out-of-level testing.

**Scoring:** Scores are obtained on three levels, the General Conceptual Ability (GCA) made up the cluster scores, with a foundation provided by the individual subtests. The GCA score is based on a definition of psychometric g as the general ability of an individual to perform complex mental processing that involves conceptualization and the transformation of information (conceptual and reasoning ability), the cluster scores represent verbal, spatial, and nonverbal reasoning abilities, and the subtest represent specific ability or processes. Special ability scores for Verbal Ability, Nonverbal Reasoning Ability, and Spatial Ability are reported as percentiles and standard scores. Such measures as perceptual and memory skills on the diagnostic subtest are reported by age as both percentile and T scores.

**Reliability:** The reliability coefficient (IRT in most cases,) was high for The GCA in all ages, with an average of .90 at the lowest preschool lever, and .94 for the upper pre school level. And .95 for the School Age level. Test-retest reliability scores were very stable for the GCA and cluster scores. Ranging from .79 to .94. Tests with high internal reliability were also found to have high test-retest reliability.

**Validity:** Inter-correlation of subtests and composites by age ranges were 100.3 and 99.7 for GCA, and SNV for ages 2:6 –3:5, 99.6 and 99.8 for ages 3:6-5:11. There were high correlations between the DAS and the WPPSI-R on the composite scores for 4 and 5 year olds. The correlations between the verbal composites of the and the DAS and the SB-IV composites were reasonably high, at .74 to .77 for 4 and 5 year olds. Correlations with the McCarthy Scales of Children’s Ability (MSCA) showed highest correlations with the Verbal, Perceptual-Performance, or Qualitative scales. For school age level all of the DAS composites correlated highly with the WISC-R Full Scale IQ, and the DAS Verbal ability cluster correlated very highly with the WISC-R Verbal IQ for 8 to 10 year olds and for 14 to 15 year olds. The verbal composites of the DAS and the SB-IV for 9 to 10 year olds also correlated very highly.

**Norms:** The normative sample included 3,475 children and adolescents representative of the US population census for race/ethnicity, gender, community size, and parent education.

**Suggested Use:** According to the author, the GCA of the DAS is an excellent predictor of academic achievement. This instrument is able to address a wide variety of referral questions for a broad age range of children in school and clinical settings, as well as in research.
**Dementia Rating Scale**

**Purpose:** Designed to measure and track mental status in adults with cognitive impairment.

**Population:** ages 56 to 105

**Score:** Attention, Initiation-Perseveration, Construction, Conceptualization, Memory, Total.

**Time:** 15 – 30 minutes

**Author:** L. Steven Mattis, Paul, J. Jurica, and Christopher L. Leitten

**Publisher:** Psychological Assessment Resources

**Description:** The Dementia Rating Scale (DRS-2), is a 36–task and 32-stimulus card individually administered instrument designed to assess level of cognitive functioning for individuals with brain dysfunction. The DRS-2 is sensitive at the lower ends of functioning and differentiating levels of deficits. Conversely, the instrument generally will not discriminate individual functioning in the average or higher range of intelligence due to the design to minimize floor effects of clinically impaired individuals. The DRS-2 consists of a professional manual, scoring booklets, and 32-stimulus cards. (DRS-2) is an enhanced version of the original DRS designed to provide standardized, quantitative cognitive functioning assessment in neurologically impaired populations. The initial pool of items was revised for comprehensive and brief administration, however, allowing for a low floor so that even severely impaired individuals could be evaluated. The DRS-2 has a wider age range than the original DSR, and the age corrected scaled and percentile ranks are more sensitive to change in cognitive status. The task and stimulus card have not been changed from the original. In the hands of an experienced neuropsychologist-clinical psychologist, this is an excellent instrument but it is highly dependent upon qualifications and skill of each individual test user.

**Scoring:** The DRS-2 assesses cognitive functioning on five subscales: Attention (ATT, 8 items); Initiation-Perseveration (I-P, 11 items); Construction (CONST, 6 items); Conceptualization (CONCEPT, 6 items); and Memory (MEM, 5 items).

**Reliability:** The reliability and validity properties of the DRS-2 are excellent. The DRS-2 uses the previously established reliability and validity scores. A test-retest reliability correlation coefficient was .97 with subscale correlation coefficients ranging from .61 to .94. The DRS was administered twice with a 1-week interval between administrations to a group of 30 patients diagnosed with dementia of the Alzheimer’s type. A split-half reliability coefficient was .90, utilizing a sample of 25 patients ages 65 to 94 years who received diagnoses of either organic brain syndrome or senile dementia. A t test indicated no significant differences between scores on the two halves. The alpha coefficients were calculated for four DRS subscales using a combined dementia sample. The alpha coefficients were Attention (.95), Initiation-Perseveration (.87), Conceptualization (.95), and Memory (.75). Five factors were found; however, there is a confound of scoring dependence as patients who score positive on the initial items are given credit for the remaining items on the subtest, which results in an artificial correlation.

**Validity:** The DRS-2 was compared with the Mini-Mental State Examination (MMSE), which displayed a significant correlation (r = .82) with the DRS-2 showing a greater sensitivity to change than the MMSE in patients with severe dementia. In addition, correlations with the Wechsler Adult Intelligence Scale indicated a correlation of .75 between the WAIS full scale and the DRS-2 total score.

**Norms:** The normative data for the DRS-2 is provided by the Mayo Older American’s Normative Study (MOANS) taken from 623 healthy adults living in community settings from ages 56–105 (199 men and 424 women). The normative data-base are not representative of U.S. population, as they consisted of predominately Caucasian adults with higher than average educational levels.

**Suggested Use:** The DRS-2 is very useful in the assessment and progression of dementia of Alzheimer’s type, vascular dementia, Parkinson’s disease, Huntington’s disease, and age-related dementia in mental retardation and Down’s syndrome.
Developmental Test of Visual-Motor Integration

**Purpose:** Designed to assess visual-motor integration in children.

**Population:** Ages 2-8 (short form), 2-15 (long form).

**Score:** Pass/fail criteria.

**Time:** (10) minutes.

**Authors:** Keith E. Beery and Norman A. Buktenica.

**Publisher:** Follett Educational Corporation.

**Description:** The long form of the VMI consists of 24 geometric forms that are to be copied in a test booklet. These geometric forms are arranged in order of increasing difficulty, and an individual’s score is calculated as the number of forms that have been copied successfully prior three consecutive failures. A separate age equivalent norm for each sex is also provided for each of the 24 items on the test. Educational assessment is the stated purpose for which the test has been designed, with particular emphasis place on the preschool group.

**Scoring:** Scoring procedures contain a high amount of subjectivity, but the authors have made an effort to provide scoring criteria included in the manual.

**Reliability:** No report of reliability is made in the administration and scoring manual. The technical report presents a collection of reliability coefficients, most of which were obtained from small samples and some samples of mentally retarded children. An internal consistency reliability (K-R 20) of .93 was computed for an unknown portion of the standardization population consisting of 594 children from suburban schools. Test-retest reliability over a two-week period, obtained for another portion of the standardization population consisting of 171 children from rural schools, was .83 for boys and .87 for girls. These coefficients were obtained from the entire age range and should be considered high.

**Validity:** A correlation of .89 between scores on the VMI and chronological age is the only validity evidence reported in the administration and scoring manual. This procedure provides a verification of only the developmental sequence of the items. The technical report describes an amount of empirical evidence that tangentially points to test validity, but no extensive prediction studies using an outside criterion have been conducted. One study that includes 342 subjects indicates that the VMI correlates .50 with reading achievement for first grade children, while results of another study using approximately 60 subjects at three grade levels show that the correlation with mental age decreases from .59 to .38 from the first to the seventh grade.

**Norms:** Standardization of the test was accomplished by using a group of 1,039 children from Illinois. Over one-half (57%) of the group selected were from suburban schools, while the remainder were selected from schools in urban (26%) and rural (17%) areas. All the groups of children were identified by school officials as average so as to create an average standardization population. The age levels for the suburban group were fairly well represented.

**Suggested Uses:** Recommended uses include assessment of visual-motor integration in educational, clinical, and research settings.
Devereux Elementary School Behavior Rating Scale

**Purpose:** Designed to rate problem behaviors of students by teachers.

**Population:** Grades K-6.

**Scores:** Ten factor scores and four cluster scores.

**Time:** (5-10) minutes.

**Author:** Marshall Swift.

**Publisher:** Devereux Foundation Press.

**Description:** The Devereux Elementary School Behavior Rating Scale (DESB) is used primarily for screening children in regular and special education programs, or for recording progress of these children with regard to school behavior and adjustment. The scale is also used in research on incidence of specific classroom behavior problems and in research on development of social competence.

**Scoring:** The DESB consists of 10 factors (work organization, creative initiative/involvement, positive toward teacher, need for direction in work, socially withdrawn, failure anxiety, impatience, irrelevant thinking/talking, blaming, negative-aggressive) that result in their respective scores. There are four cluster scores: perseverance, peer cooperation, confusion, and inattention.

**Reliability:** The manual reports test-retest reliabilities over a 1 week ranging from .89 to .91, the median being .87.

**Validity:** The manual does not include any information on validity.

**Norms:** Normative DESB data were obtained from 13 elementary schools in a consolidated small city public school system. Thirty-two kindergarten through sixth grade teaches made ratings of the behaviors of 809 children. The sample was representative of the United States population.

**Suggested Uses:** Recommended uses include screening in an educational setting.
**Dissociative Experiences Scale**

**Purpose:** Designed to help identify patients with dissociative psychopathology and as a research tool to quantify dissociative experiences.

**Population:** Late adolescence to adult

**Score:** Total score only

**Time:** 10 minutes

**Author:** Eve Bernstein Carlson, Frank W. Putnam

**Publisher:** Eve Bernstein Carlson Department of Psychology Beloit College

**Description:** The Dissociative Experiences Scale (DES) is a brief self-report questionnaire designed to identify patients with dissociative psychopathology and to provide a means of quantifying dissociative experiences. Content for the scale’s 28 items was garnered from extensive interviews with dissociative patients and from consultations with clinical experts. The scale taps a broad range of dissociative experiences including disturbances in memory, identity, and cognition, and feelings of derealization, depersonalization, absorption, and imaginative involvement. The DES comes in two forms: the original DES and the DES II. Both versions use the same items, but they differ in their item formats.

**Scoring:** The DES in its original format uses a visual analogue scale that requires examinees to mark their responses along a numerically anchored 100-millimeter line. Item responses range from 0%, “This never happens to you,” to 100%, “This always happens to you.” The DES II uses a more convenient 11-point Likert scale. Total scores are obtained by averaging the 28 item scores.

**Reliability:** Reliability findings from six studies are reported in the manual. The weighted means of the test-retest and internal consistency reliabilities from these studies are .85 and .93, respectively. These figures indicate that the DES yields reliable scores in many assessment contexts.

**Validity:** The primary validation studies with the DES feature Spearman rank-order correlations of the instrument with age and socioeconomic status. Item-total correlations also use this statistic. The manual is a compendium of psychometrically informative text and tables that should be consulted by potential users of the scale. It contains useful information on norms, reliability, validity, structure (via factor analysis), and use of cutoff scores for classification.

**Norms:** Norms are presented for both clinical and nonclinical samples including subgroups with anxiety disorder, affective disorder, eating disorders, schizophrenia, borderline personality disorder, PTSD, dissociative disorders, and multiple personality. Scale means or medians are reported for each diagnostic group.

**Suggested Use:** The DES is useful as a screening test for major dissociative psychopathology, and will serve as a useful research tool.
Eating Disorder Inventory

Purpose: Designed as a self-report measure of "psychological and behavioral traits common in anorexia nervosa and bulimia."

Population: Ages 12 and over.

Scores: Eight subscale scores.

Time: (15-25) minutes.

Authors: David M. Garner, Marion P. Olmsted, and Janet Polivy.

Publisher: Psychological Assessment Resources, Inc.

Description: The Eating Disorder Inventory (EDI) is a 64-item, 6-point forced-choice inventory assessing several behavioral and psychological traits common in two eating disorders, bulimia and anorexia nervosa. The EDI, a self-report measure, may be utilized as a screening device, outcome measure, or part of typological research. It is not purported to be a diagnostic test for anorexia nervosa or bulimia.

Scoring: The EDI consists of 8 subscale scores: drive for thinness, bulimia, body dissatisfaction, ineffectiveness, perfectionism, interpersonal distrust, interoceptive awareness, and maturity fears.

Reliability: The average item total correlation of the eight subscales of the Eating Disorder Inventory was .63 (SD = .13). Reliability information was based on 271 college women on whom completed information on all subscales was obtained. Reliability coefficients (standardized Chronbach’s alphas) for the anorexia nervosa group ranged from .83 (Interoceptive Awareness) to .93 (Ineffectiveness). Reliability coefficients for the female college students ranged from .72 (Maturity Fears) to .92 (Body Dissatisfaction).

Validity: Criterion-related validity studies were performed by comparing the EDI patient profiles with the judgments of clinicians familiar with the patient’s psychological presentation. A subgroup of 49 of the anorexia nervosa patients who had completed the EDI was assigned two raters: a psychologist and psychiatrist who were familiar with the patients, being their primary therapist or consultant. The raters were provided with the description of the subscale content and with the patients’ total score percentile rank within the entire anorectic sample. All interrater correlations were significant at the p < .001 level and ranged from .43 (Maturity Fears) to .68 (Ineffectiveness).

Norms: Two groups of respondents participated in the validation of the EDI. The criterion group (n = 129) was composed of three subsamples of women, mostly anoretics. These women averaged 20% below the expected weight for their height and age. In this sample, 56 were classified as “restricters” and 73 were diagnosed as “bulimic.” The comparison group (n = 770) consisted of three samples of female university students who were enrolled in introductory and upper-level psychology classes. These volunteers were administered the EDI in their classes.

Suggested Uses: The EDI is recommended to delineate subtypes of anorexia nervosa in clinical or research settings.
Eating Disorder Inventory-2

**Purpose:** Designed as a self-report measure of psychological features commonly associated with anorexia nervosa and bulimia nervosa.

**Population:** Ages 12 and over.

**Scores:** 11 scores: Drive for Thinness, Bulimia, Body Dissatisfaction, Ineffectiveness, Perfectionism, Interpersonal Distrust, Interoceptive Awareness, Maturity Fears, Asceticism (provisional), Impulse Regulation (provisional), Social Insecurity (provisional).

**Time:** (20) minutes.

**Author:** David Garner.

**Publisher:** Psychological Assessment Resources, Inc.

**Description:** The Eating Disorder Inventory (EDI-2) is a self-report measure of symptoms frequently related to anorexia nervosa or bulimia nervosa. As such, the EDI-2 was designed as an aid to forming a diagnosis and not as the exclusive basis for making a diagnosis. The EDI-2 provides clinical information regarding the psychological and behavioral dimensions of eating disorders. The EDI-2 package also includes the EDI Symptom Checklist (EDI-SC), a structured self-report form soliciting current and historical information about the client’s eating-related and menstrual history.

**Scoring:** The EDI-2 retains the 64 items (grouped into eight scales) of the EDI and adds 27 new items into three provisional scales: Asceticism, Impulse Regulation, and Social Insecurity.

**Reliability:** Internal consistency reliability coefficients for the EDI-2 scales are between .44 and .93. Test-retest reliability for EDI-2 administered one week apart to 70 student and staff nurses revealed coefficients of .79 to .95 for all subscales except Interoceptive Awareness. After 3 weeks, test-retest reliabilities for 70 nonpatient university undergraduates were all above .80, excluding Maturity Fears.

**Validity:** The original subscales show appropriate content, convergent, and discriminant validity. Further, many of the findings from these earlier validation studies have been replicated by new research. Correlations between EDI scales and other personality instruments also yielded many positive correlations (from about -.08 to .76) indicating that the constructs measured by the EDI involve to some degree other personality characteristics not unique to eating disorders. The psychometric properties of the instrument are sound and the constructs measure symptom domains and have clinical utility. The authors of the measure have evidence the EDI-2 is sensitive to clinical change and that it can play a valuable role in clinical evaluations of eating disorder patients.

**Norms:** College-age females.

**Suggested Uses:** Psychiatrists, psychologists, and social workers who work with young women or men who are suspected or known to have eating disorders would benefit from using the EDI-2. It can also help expand clinical knowledge about eating disorders and associated attributes.
Edwards Personal Preference Schedule

**Purpose:** Personality inventory.

**Population:** College and adults.

**Scores:** 15 scores.

**Time:** (40-55) minutes.

**Author:** Allen L. Edwards.

**Publisher:** The Psychological Corporation.

**Description:** The Edwards Personal Preference Schedule (EPPS) is a forced choice, objective, non-projective personality inventory, derived from the theory of H. A. Murray, which measures the rating of individuals in fifteen normal needs or motives. On the EPPS there are nine statements used for each scale. Social Desirability ratings have been done for each item, and the pairing of items attempts to match items of approximately equal social desirability. Fifteen pairs of items are repeated twice for the consistency scale.

**Scoring:** The EPPS consists of 15 scales: achievement, deference, order, exhibition, autonomy, affiliation, interception, succorance, dominance, abasement, nurturance, change, endurance, heterosexuality, and aggression.

**Reliability:** Split-half reliability coefficients, or coefficients of internal consistency for 1,509 students in the college normative group range from .60 to .87 with a median of .78. The author also presents test-retest stability coefficients with a one-week interval. These are based on a sample of 89 students and range from .55 to .87 with a median of .73. Other researchers have reported similar results over a three-week period, showing correlations of .55 to .87 with a median of .73.

**Validity:** The manual reports studies comparing the EPPS with the Guilford Martin Personality Inventory and the Taylor Manifest Anxiety Scale. Other researchers have correlated the California Psychological Inventory, the Adjective Check List, the Thematic Apperception Test, the Strong Vocational Interest Blank, and the MMPI with the EPPS. In these studies there are often statistically significant correlations among the scales of these tests and the EPPS, but the relationships are usually low-to-moderate and often are difficult for the researcher to explain.

**Norms:** 1,509 students in college.

**Suggested Uses:** Recommended primarily for instructional value and research settings.
**Family Environment Scale**

**Purpose:** Designed to “measure the social-environmental characteristics of all types of families.”

**Population:** Family members.

**Score:** Ten scores are derived from the subscales to create an overall profile of family environment.

**Time:** (15-20) minutes.

**Authors:** Rudolf H. Moos and Bernice S. Moos.

**Publisher:** Consulting Psychologists Press, Inc.

**Description:** The Family Environment Scale (FES) was developed to measure social and environmental characteristics of families. The scale is based on a three-dimensional conceptualization of families. Additionally, three separate forms of the FES are available that correspondingly measure different aspects of these dimensions. The Real Form (Form R) measures people’s perceptions of their actual family environments, the Ideal Form (Form I) rewords items to assess individuals’ perceptions of their ideal family environment, and the Expectations Form (Form E) instructs respondents to indicate what they expect a family environment will be like under, for example, anticipated family changes.

**Scoring:** The Relationship dimension includes measurements of Cohesion, Expressiveness, and Conflict. The Personal Growth dimension involves assessments of Independence, Achievement Orientation, Intellectual-Cultural Orientation, Active-Recreational Orientation, and Moral-Religious Emphasis. The System Maintenance dimension includes Organization and Control measures. Scores for each of these 10 subscales are derived to create an overall profile of family environment. Based on these scores, families are then grouped into one of three family environment typologies based on their most salient characteristics.

**Reliability:** Internal consistency reliability estimates for the Form R subscales range from .61 to .78. Intercorrelations among these 10 subscales range from -.53 to .45. These data suggest that the scales are measuring relatively distinct characteristics of family environment and with reasonable consistency. Test-retest reliabilities for the Form R subscales for 2-month, 3-month, and 12-month intervals range from .52 to .91. These estimates suggest that the scale is reasonably stable across these time intervals.

**Validity:** The face and content validity of the instrument are supported by clear statements about family situations that relate to subscale domains. Evidence of construct validity is presented in the manual through comparative descriptions of distressed and normal family samples; comparisons of parent responses with those of their adolescent children; descriptions of responses by families with two to six or more members; and descriptions of families with a single parent, of minority families, and of older families. Additional validity evidence is provided in the manual through summaries or references to approximately 150 additional research studies.

**Norms:** Information regarding normative data is not included in the manual.

**Suggested Uses:** The FES recommended as a viable instrument in the study of family systems.
Family Environment Scale: Children’s Version

**Purpose:** Designed to measure a young child’s subjective appraisal of their family environment.

**Population:** Ages 5-12.

**Score:** 10 scores.

**Time:** Not reported.

**Authors:** Christopher J. Pino, Nancy Simons, and Mary J. Slawinowski.

**Publisher:** Slosson Educational Publications, Inc.

**Description:** The Children’s Version of the Family Environment Scale (CVFES) is a downward extension of the Family Environment Scale (FES). As such, its purpose is to enable children, ages 5 to 12 to provide self-reports of family relationships.

**Scoring:** Children’s perceptions of 10 dimensions in three general areas of family functioning are assessed: Relationship Dimensions (Cohesion, Expressiveness, and Conflict); Personal Growth Dimensions (Independence, Achievement Orientation, Intellectual-Cultural Orientation, Active-Recreational Orientation, and Moral-Religious Emphasis); and System Maintenance Dimensions (Organization and Control).

**Reliability:** The authors report that the 4-week test-retest reliability was .80. No range of reliabilities is given, so one presumes that this is the reliability for the total score, even though no indication is given that a total score should be computed. Information about differences in reliabilities for different ages should also be given.

**Validity:** The validity of the CVFES rests on the validity of the FES and the pictures developed to tap its content. While the FES has adequate reliability, the evidence for its validity is weak. In developing the CVFES, the authors grouped the items of the FES so that a smaller number of items would tap the same content as the adult version. They then chose those items which “best” cut across the nine FES scales, but they do not indicate what their criteria were.

**Norms:** As noted by the authors, the standardization sample for the CVFES is very restricted: 158 children from grades 1 to 6 of the Buffalo parochial schools. No specific breakdown of the sample is provided.

**Suggested Uses:** Reviewers have recommended that the CVFES should be considered an experimental instrument and as such, should be used as part of an assessment battery in clinical and research settings.
Forer Structured Sentence Completion Test

**Purpose:** Designed to identify individuals’ attitudes and views about themselves, others, and the world.

**Population:** Adolescents and adults.

**Score:** Four categories.

**Time:** (15-20) minutes.

**Author:** Bertram R. Forer.

**Publisher:** Western Psychological Services.

**Description:** The Forer Structured Sentence Completion Test (FSSCT) is comprised of 100 sentence stems developed to identify individuals’ attitudes and views about themselves, others, and the world. Sentence stems reflect an approximately equal balance among third-person and first-person singular and plural stimulus presentations. The structure of the FSSCT involves both the specificity of the sentence stems and the evaluation system used to assess the quality of responses.

**Scoring:** Forer has developed a Checklist and Clinical Evaluation Form for both the adolescent and adult forms. These forms provide a structured evaluation scheme that the examiner can use to group individual items into one of four categories: Interpersonal Figures, Wishes, Causes of Own (feelings and behaviors), and Reactions (to other people). Each item is rated according to “attitudes toward” and “characteristics of” in the first two categories and “attitudes toward” in the last two categories. All of the terms used have specific clinical operational definitions and clear implications about the structure and organization of one’s personality. Forer has attempted to define these terms operationally in the manual.

**Reliability and Validity:** The manual contains no information on reliability and validity.

**Norms:** Norms are not available.

**Suggested Uses:** The FSSCT is recommended as a supplement to a battery of psychological tests in clinical or research settings.
Frostig Developmental Test of Visual Perception

**Purpose:** Designed to assess visual perception skills in children.

**Population:** Ages 4-8.

**Score:** Perceptual Quotient.

**Time:** (30-45) minutes.

**Author:** Marianne Frostig.

**Publisher:** Consulting Psychologists Press, Inc.

**Description:** The Frostig Developmental Test of Visual Perception (DTVP) is a test designed to assess visual perceptual skills in children. It provides information that includes 1) an estimation of the overall visual perception ability of the child and 2) a delimitation of the distinct visual perception difficulties in need of training.

**Scoring:** The five DTVP subtests are presumed to measure five distinct subareas of visual perception: Eye Motor Coordination, Figure Ground, Constancy of Shape, Position in Space, and Spatial Relationships. Raw scores are obtained for each of the subtests and then converted to Age Equivalents or Perceptual Ages (PAs) and Scale Scores (SSs). The total test results are expressed in Perceptual Quotient (PQ) and percentile rank scores.

**Reliability:** The authors report three test-retest reliability studies. The reliability coefficients based on two- to three-week intervals for the total test (PQ) ranged from .69 to .98. Test-retest reliability coefficients for the subtests (SS) were lower, ranging from .29 (Subtest 1) to .80 (Subtest 3), with most coefficients falling in the .50-.60 range. Split-half reliability coefficients for the total test ranged from .78 (8-9 years) to .89 (5-6 years). Coefficients for the DTVP subtests ranged from .35 to .96, with The Figure Ground subtest (2) showing the highest reliability (.91-.96) and The Position in Space subtest (4) showing the lowest reliability (.35-.70).

**Validity:** Correlations between scores on the DTVP and teacher ratings of classroom adjustment, motor coordination, and intellectual functioning were .44, .50 and .50, respectively. Correlations between scores on the DTVP and the Goodenough Draw-A-Man Test ranged from .32 to .46 for Kindergartners and first- and second-graders. Factor analytic studies reported by other researchers have indicated that a single perceptual factor exists rather than five independent factors represented by each subtest. This suggests that the PQ is a valid overall measure of visual perception but the DTVP’s ability to differentially assess specific areas of visual perception is questionable.

**Norms:** The DTVP was standardized based on 2,116 normal school children between the ages of 3-9 years. The sample was drawn from a restricted area and was predominantly middle class (93%) in nature. Children from low socioeconomic groups and minority groups were poorly represented with only a small number of Hispanics and Asians included. No black children were represented.

**Suggested Uses:** Recommended as part of a larger assessment battery of visual-perceptual skill development.
General Clerical Test

**Purpose:** Developed to assess clerical speed and accuracy, numerical skills, and language-related skills.

**Population:** Clerical applicants and workers.

**Score:** Four: Clerical, Numerical, Verbal, and Total.

**Time:** (46-51) minutes.

**Author:** The Psychological Corporation.

**Publisher:** The Psychological Corporation.

**Description:** The General Clerical Test is a nine-part right/wrong test assessing skills related to general clerical and office management. According to the manual, the test may be used to predict success in jobs requiring applications of the three specific areas, the suitability of a given job for a particular applicant, or in assigning an inexperienced person to appropriate work.

**Scoring:** The main test consists of nine parts, summing to three subscores labeled Clerical, Numerical, and Verbal. The first section, Clerical, consists of two parts and reflects speed and accuracy on perceptual tasks. The Numerical section contains three parts reflecting computation, reasoning, and error location. The Verbal component consist of four sections addressing spelling, grammar, vocabulary, and reading comprehension. All items have one correct answer and responses are entered directly onto the test booklet.

**Reliability:** Test-retest correlations are moderate given continued evidence from reliability studies that considerable improvements in scores is noted on retesting.

**Validity:** Correlations included in the manual provide reasonable evidence for content, criterion-related, concurrent, contrasted groups, and correlations with other tests. A note of caution: users must be aware of the fact that not all types of validity evidence are available for all populations.

**Norms:** Norms are included for paralegal students, community college trainees, customer service applicants, clerical/service positions in public utilities, blue collar and management trainee positions, and multiple levels of clerical tasks in industry, government, and university settings. It should be noted that in most cases the norms reflect scores of white females. Scores reflective of male respondents are included in blue collar, utility, and management trainee positions. No separate gender-based norms are available.

**Suggested Uses:** It is recommended that the General Clerical Test may be most appropriately used in educational rather than actual employment settings, although using it in such settings will garner some helpful information.
Geriatric Sentence Completion Form

**Purpose:** Designed to assess personality characteristics of older adults.

**Population:** Ages 60 and over.

Scores: Personal responses in four domains.

**Time:** (20-30) minutes.

**Author:** Peter LeBray.

**Publisher:** Psychological Assessment Resources, Inc.

**Description:** The Geriatric Sentence Completion Form was developed to assess personality characteristics of older adults in a projective format.

**Scoring:** The GSCF elicits personal responses in four domains: physical, psychological, socioenvironmental, and temporal orientation. The manual lacks any suggested scoring system.

**Reliability:** No information on reliability was included in the manual.

**Validity:** No information on validity was included in the manual. The manual does provide five clinical case illustrations. Four of these concern individuals who are at least 79 years old, so the question of applicable range is a pertinent one.

**Norms:** No information on norms was included in the manual.

**Suggested Uses:** The GSCF offers clinicians an assessment tool that may be useful in working with the elderly.
Halstead-Reitan Neuropsychological Test Battery

Purpose: Developed as a comprehensive neuropsychological test battery.

Population: Ages 5-8, 9-14, 15 and over.

Scores: One combined score: Halstead impairment index.

Time: Not reported.

Author: Ralph M. Reitan

Publisher: Reitan Neuropsychology Laboratories, University of Arizona.

Description: The Halstead-Reitan Neuropsychological Test Battery (HRNB) was developed to predict the presence of brain damage while offering a comprehensive view of a patient’s individual functions. The purpose of this battery is to provide the clinician with a database for inferring the nature, location, and extent of the structural changes in the brain that may underlie and explain the pattern of intact and impaired functions derived from the measures and qualitative information yielded by the battery. The present battery consists of 10 tests which have been shown empirically to best discriminate between normals and patients with documented cortical damage. The authors offer convincing data favoring clearer brain damage localization with the HRNB for acute lesions rather than more chronic neuropathology.

Scoring: The battery includes tests purported to measure elements of memory, abstract thought, language, sensory-motor integration, imperception, and motor dexterity.

Reliability: Information on reliability is not included in the manual.

Validity: The battery has been shown to discriminate normal controls from patients with brain damage with considerable accuracy (84-98%). The normative base for establishing a profile of individual standard score patterns can be found in a large collection of studies designed to demonstrate validity of the battery to differentiate (1) organic neurological from normal populations, (2) organic neurological from functional psychiatric populations, (3) focal from diffuse neurological disease, (4) regional focal cerebral dysfunction by major zones, and (5) the etiological conditions associated with individual differences in outcome pattern.

Norms: The manual reports no standardization or normative data, but relies upon the body of research which has evolved for the past 30 years with the battery.

Suggested Uses: It is suggested that the HRNB is a more than adequate neuropsychological instrument for use in clinical and research settings.
Harris Test of Lateral Dominance

**Purpose:** Designed to measure directional confusion.

**Population:** Children and adults.

**Score:** Total hand dominance rating.

**Time:** Not reported.

**Author:** Albert J. Harris.

**Publisher:** The Psychological Corporation.

**Description:** The author defines lateral dominance as the preferred use and better performance of one side of the body as compared to the other side. Dominance is said to be crossed when the dominant hand and dominant eye are on opposite sides. Hand dominance is called mixed or incomplete when the individual does not show a consistent preference for, or superiority of, one hand. The author claims that this battery, which consists of ten tests, are sensitive indicators of directional confusion.

**Scoring:** Test 1 measures subject’s knowledge of right and left; test 2 assesses hand preferences; test 3 assesses simultaneous writing; test 4 assesses handwriting; test 5 measures tapping; test 6 assesses card dealing ability; test 7 measures strength of grip; test 8 consists of monocular tests; test 9 consists of binocular tests; and test 10 involves stereoscopic testing. After the separate hand dominance tests have been rated, a total hand dominance rating is ascertained. The author states that this rating is a matter of qualitative judgment.

**Reliability:** The author states that a rough approximation of reliability coefficients can be obtained by the use of the coefficient of contingency. On this basis, the author states that the maximum value for the four hand dominance tests is .894. Split-half techniques yield Spearman-Brown reliability coefficients of .85 and .88. Test-retest reliabilities based on group administration in college classes yield coefficients ranging from .75 to .83.

**Validity:** The author claims that the nature of the tasks are appropriate for the purposes for which the test are intended. On this basis, he claims the tests evidence content validity. The author also asserts that the hand dominance tests are reliable and discriminative measures which are more sensitive to mixed dominance and directional confusion than many previous sets of hand dominance tests.

**Norms:** The manual provides no information on norms.

**Suggested Uses:** Recommended uses include assessment of hand dominance or directional confusion in research and clinical settings.
The Holtzman Inkblot Technique

Purpose: Designed as a multi-variable projective personality test.

Population: Ages 5 and over.

Score: 20-22 scores.

Time: (75) minutes.

Author: Wayne H. Holtzman.

Publisher: The Psychological Corporation.

Description: The Holtzman Inkblot Technique (HIT) is a multi-variable projective personality test that was designed to overcome psychometric limitations of the Rorschach.

Scoring: The HIT is an inkblot test that consists of two alternate forms of 45 inkblots. It is scored on 22 variables. These are: reaction time, rejection. Location, space, form definiteness, form appropriateness, color, shading, movement, pathognomic verbalization, integration, content (human, animal, anatomy, sex, abstract), anxiety, hostility, barrier, penetration, balance, and popular. Administration and scoring are rather time consuming, except in computerized versions.

Reliability: The author reported intrascorer consistency on three examiners after a period of about 3 months. These reliabilities average about .95 on 9 scores. Interscorer consistency is reported for a number of different scorers for various scores. The lowest reported score reliabilities are .57, the highest .99. Intrasubject reliabilities are reported for 15 different groups on all 22 scores. These split-half reliabilities vary widely, but generally range between .50 and .90 with an approximate average in the .70s or .80s. No test-retest reliabilities were found which might furnish information about whether the scores are measuring persisting traits.


Norms: College students, average adults, 7th graders, elementary school children, 5-year olds, chronic schizophrenics, mental retardates, and depressed patients.

Suggested Uses: Recommended uses of the Holtzman Inkblot test include experimental and clinical applications.
House-Tree-Person Interrogation Form

**Purpose:** Designed as a supplemental scoring form for the house-tree-person technique.

**Population:** Children and adults.

**Score:** IQ score.

**Time:** Not reported.

**Author:** John N. Buck.

**Publisher:** Western Psychological Services.

**Description:** The House-Tree-Person (H-T-P) projective technique developed by John Buck was originally an outgrowth of the Goodenough scale utilized to assess intellectual functioning. Buck felt artistic creativity represented a stream of personality characteristics that flowed onto graphic art. He believed that through drawings, subjects objectified unconscious difficulties by sketching the inner image of primary process. Since it was assumed that the content and quality of the H-T-P was not attributable to the stimulus itself, he believed it had to be rooted in the individual’s basic personality. Since the H-T-P was an outcropping of an intelligence test, Buck developed a quantitative scoring system to appraise gross classification levels of intelligence along with at qualitative interpretive analysis to appraise global personality characteristics.

**Scoring:** The Post-Drawing Interrogation form (PDI) consists of 60 questions varying from direct and concrete to indirect and abstract. Once the PDI has been administered and the interview has been completed, the examiner records items of detail, proportion, and perspective in the Scoring Folder. After completing the scoring tables, the examiner derives an IQ figure for the percentage of raw G, a net weighted score, a weighted “good” score, and a weighted “flaw” score, which then comprise the items for the profile configuration.

**Reliability and Validity:** The manual contains no information on validity and reliability.

**Norms:** The standardization sample included 140 adults. No attempt was made to randomly select a stratified sample of subjects from the general population. Twenty adults were selected for each of seven intellectual levels (imbecile, moron, borderline, dull average, average, above average, and superior).

**Suggested Uses:** This instrument is recommended for projective assessment in research and clinical settings.
Illinois Test of Psycholinguistic Abilities, Revised Edition

**Purpose:** Designed to measure acquisition and use of language.

**Population:** Ages 4-8 years.

**Score:** Scaled Scores and a Composite Score.

**Time:** (60) minutes.

**Authors:** Samuel A. Kirk, James J. McCarthy, and Winifred D. Kirk.

**Publishers:** University of Illinois Press.

**Description:** The Revised Illinois Test of Psycholinguistic Abilities (ITPA) is an individually administered test for children aged 4-8 years, measuring 12 functions employed in the acquisition and use of language.

**Scoring:** The test consists of 10 main subtests (Auditory Reception, Visual Reception, Auditory Association, Visual Association, Verbal Expression, Manual Expression, Grammatic Closure, Visual Closure, Auditory Sequential Memory, and Visual Sequential Memory) and two supplementary subtests (Auditory Closure and Sound Blending). Raw scores for each subtest are used to derive Scaled Scores for each subtest, a Composite Score, Psycholinguistic Age Scores for subtests and Composite, and Psycholinguistic Quotients for subtests and Composite. The book entitled Psycholinguistic Learning Disabilities: Diagnosis and Remediation, by Samuel A. Kirk and Winifred D. Kirk is offered in the testing library as a supplement to those using this test.

**Reliability:** Internal consistency coefficients range from .60-.96, with a median of .88, except for one coefficient of .45 for Sound Blending for the age level 8.7-9.1. High internal consistency was confirmed (range .63-.93) for a group of educable mentally retarded children. Internal consistency coefficients for difference scores among the ITPA subtests range from .67-.91 with a median of .81. Five-month test-retest coefficients for 4-, 6-, and 8-year-old children ranged from .28-.90 with a median of .71 for subtests, .87-.93 for Composite scores, and .86-.91 for Psycholinguistic Quotients.

**Validity:** Several studies have obtained correlations between ITPA scores and intelligence measures. Paraskevopoulos and Kirk report significant correlations of ITPA Composite scores and Psycholinguistic Quotients with Stanford-Binet Mental Age, IQ, and Vocabulary scores across several age levels (standardization data). Huizinga tested 100 six-year-old children and found ITPA Total Scaled Scores correlated significantly with Stanford-Binet IQ (.88), as well as WISC Verbal, Performance, and Full Scale IQS (.75, .58 and .80). Guest obtained significant correlations for ITPA Total Score with Wechsler Preschool and Primary Scale of Intelligence Verbal, Performance, and Full Scale IQS (.69, .47, and .67) for 47 kindergarten children. Wechsler Full Scale IQ was significantly correlated with ITPA Total scores (.61) for 73 third-grade children (Bartin, 1971) and with ITPA Psycholinguistic Quotients (.87) for 136 children aged 4-10 years.

**Norms:** 962 children aged 2 to 10 who had average: intelligence, achievement, and spoken English.

**Suggested Uses:** Recommended for assessing children’s language in educational, research, and clinical settings.
**Jenkins Activity Survey**

**Purpose:** Designed as a measure of Type A behavior and coronary prone behavior.

**Population:** Employed adults ages 25-65.

**Scores:** Type A, speed and impatience, job involvement, hard-driving and competitive.

**Time:** (20-30) minutes.

**Authors:** C. David Jenkins, Stephen Zyzanski, and Ray Rosenmen.

**Publisher:** The Psychological Corporation.

**Description:** The Jenkins Activity Survey (JAS) was developed in an attempt to duplicate the clinical assessment of the Type A behavior pattern by employing an objective psychometric procedure. Individuals displaying a Type A behavior pattern are characterized by extremes of competitiveness, striving for achievement and personal recognition, aggressiveness, haste, impatience, explosiveness and loudness in speech, characteristics which the JAS attempts to measure.

**Scoring:** The JAS is a self-administered, multiple-choice questionnaire that yields a composite Type A scale score and three factor analytically-derived subscales: Speed and impatience, Job involvement, and Hard-driving and competitive. The Type A scale consists of 21 items, while Speed and impatience (Factor S), Job involvement (Factor J) and Hard-driving and competitive (Factor H) contain 21, 24, and 20 items respectively.

**Reliability:** Reliability estimates for the JAS Type A scale appear to be adequate. Estimates of item reliabilities derived from squared multiple correlation coefficients range from .27 to .75, with the coefficient for the Type A scale reported to be .85. Test-retest reliability estimates generally range between .60 and .70 for retest intervals of from six months to four years. However, most of the correlations are based upon significant modifications in the items in later versions of the JAS.

**Validity:** Concurrent Validity has been established by comparing JAS scores to Type A ratings based upon the structured interview. Despite a statistically significant association between the two measures, as many as 30% of respondents are classified differently by the JAS and interview, and correlations between the JAS and interview in younger populations (e.g., college students) are less than .50. Evidence for the predictive validity of the JAS comes primarily from the prospective findings of the Western Collaborative Group Study. Analysis of JAS Type A scores of 2,750 healthy men showed the Type A scale to distinguish the 120 future clinical cases of coronary heart disease from those men who subsequently remained healthy. Numerous studies have also found patients with coronary heart disease to score higher on the JAS Type A scale than patients without heart disease.

**Norms:** The norms cannot be considered representative of any population of broad general interest. The standardization sample did not include women, young or elderly, or persons with low socioeconomic status.

**Suggested Uses:** Although the JAS is currently the most widely used instrument to assess Type A behavior, evidence indicates that the JAS should be limited to experimental or clinical research.
**Kaufman Adolescent and Adult Intelligence Test**

**Purpose:** Designed to measure general intelligence.

**Population:** ages 11 to 85+.

**Scales:** Crystallized Scale (Gc), Fluid Scale (Gf), Measures of Delayed Recall, Mental Status, Total

**Time:** [58-73] minutes for Core Battery; [83-102] minutes for Expanded Battery.

**Authors:** Alan S. Kaufman and Nadeen L. Kaufman

**Publisher:** American Guidance Service Inc.

**Description:** The Kaufman Adolescent and Adult Intelligence Test (KAIT) is an individually administered intelligence test battery composed of three intelligence scales: Fluid (Gf), Crystallized (Gc), and Composite Intelligence. The tests for fluid scale include paired-associative learning, deductive reasoning, and a test of both inductive and deductive reasoning. The tests for crystallized scores contain measures of lexical knowledge and listening ability. They require synthesis and integration, and memory for meaningful material. In addition to the six core battery subtests, the Expanded Battery contains an alternate test for the fluid score that involves visual memory, analysis, and synthesis, and an alternate test for crystallized score assessing general factual knowledge, long term retrieval and integration of facts, and comparing immediate and delayed memories. All KAIT subtests were constructed following Piaget's stage of formal operations and Luria and Golden's conception of planning ability. The 10 items on the Mental Status subtest are designed to assess attention and orientation in time and space, and performance is classified as Average, Below Average, or Lower Extreme. The category of Lower Extreme is subdivided into Mild, Moderate, and Severe Deficit, and for individuals in this range, the Kaufmans suggest administration of the Famous Faces subtest only.

**Scoring:** These six subtests comprise the Core Battery of the KAIT and, together, yield the Composite IQ, an index of general intelligence. The standard IQ scores have M=100 and SD=15. The standard scores of the ten subtests have M=10 and SD=3 and are called subtest scaled scores.

**Reliability:** Test-retest reliability coefficients ranged from .87 (Fluid IQ Scale) to .97 (Crystallized IQ Scale). In general, the KAIT IQ scales are sufficiently reliable regarding general intellectual ability and Fluid and Crystallized cognitive functioning. The subtests test-retest reliability coefficients were below .80 and should be used in conjunction with other data.

**Validity:** The information presented suggests that KAIT scores show a developmental progression similar to what might be expected if the tests measured crystallized and fluid intelligence. Exploratory and confirmatory factor analyses suggest that the KAIT does indeed measure two factors similar in structure to the organization of the test. The crystallized factor is similar to the verbal factor found for the Wechsler tests, whereas the fluid scale appears relatively distinct from a Wechsler perceptual organization factor. Multi-sample confirmatory factor analyses suggest that the KAIT does indeed measure the same abilities across its age span. In correlations with other tests, the Crystallized scale appears to correlate well with other measures of crystallized ability, but the Fluid scale does not always behave in the manner that would be predicted if it were indeed a measure of fluid intelligence.

**Norms:** The normative sample included 2,000 individuals ages 11 to 85+ years, spanned 13 age levels with at least 125 individuals at each level. The sample was representative of the US population census for gender, socioeconomic status (examinee or parental education levels), and race or ethnic group.

**Suggested Use:** The KAIT is considered a viable alternative to the Wechsler Scales as a measure of general intellectual functioning, and has shown its potential utility for making differential diagnoses. The KAIT Fluid Scale and Wechsler Scale can be used to supplement each other for additional information about cognitive abilities. The KAIT may be useful in cross-battery approach to intellectual assessment.
Kaufman Assessment Battery for Children

**Purpose:** Designed for assessing cognitive development in children.

**Population:** Children, ages 2.5-12.5.

**Score:** 16 subtests.

**Time:** (40-85) minutes.

**Author(s):** Alan Kaufman and Nadeen Kaufman.

**Publisher:** American Guidance Service.

**Description:** The Kaufman Assessment Battery for Children (K-ABC) is a clinical instrument for assessing cognitive development. Its construction incorporates several recent developments in both psychological theory and statistical methodology. The K-ABC also gives special attention to certain emerging testing needs, such as use with handicapped groups, application to problems of learning disabilities, and appropriateness for cultural and linguistic minorities. The authors rightly caution, however, that success in meeting these special needs must be judged through practical use over time. They also point out that the K-ABC should not be regarded as “the complete test battery”; like any other test, it should be supplemented and corroborated by other instruments to meet individual needs, such as the Stanford-Binet, Wechsler scales, McCarthy scales, or neuropsychological tests.

**Scoring:** The 16 subtests are grouped into a mental processing set and achievement set, which yield separate global scores. The mental processing set is then grouped into those requiring primarily sequential processing of information and those requiring simultaneous processing, with separate global scores for each.

**Validity and Reliability:** Odd-even reliabilities within one-year age groups averaged in the .70s and .80s for subtests; for global scores, the averages were in the high .80s and .90s. Test-retest reliabilities were computed within age groups spanning 3 or 4 years, retested after intervals of 2 to 4 weeks. For subtests, these reliabilities ranged from .59 to .98, clustering in the .70s and .80s; for global scores, they ranged from .77 to .97. In general, reliabilities were higher for the achievement than for the mental processing tests. Concurrent and predictive validity (6- to 12 interval) against standardized achievement tests, were investigated in several small groups of both normal and exceptional children. The correlations vary widely, but most appear promising, and the patterns of correlations with subtests tend to fit theoretical expectations. Analyses by ethnic groups yielded closely similar validities for Blacks, Hispanics, and Whites.

**Norms:** Norms for the battery are based on administration of the tests to representative samples of 100 children at each 6-moth age interval from 2.5 to 12.5, a total of 2000 individuals. A variety of supplementary norms are provided, some requiring the testing of additional subjects. Sociocultural norms are provided based on a cross-tabulation by race (black-white) and by parental education (less than high school education, high school graduate, and one or more years of college or technical school).

**Suggested Uses:** Recommended uses of the K-ABC include integration as a component of a cognitive assessment battery in clinical situations.
Kaufman Test of Educational Achievement

Purpose: Designed to measure school achievement of children.

Population: Grades 1-12.

Score: Age and grade norms.

Time: (60-75) minutes.

Authors: Alan S. Kaufman and Nadeen L. Kaufman.

Publisher: American Guidance Service.

Description: The Kaufman Test of Educational Achievement (K-TEA) was designed to measure school achievement of children enrolled in Grades 1-12. It consists of two overlapping forms: Comprehensive and Brief. The Brief Form globally samples the areas of reading, mathematics, and spelling, whereas the Comprehensive Form measures more specific skills in the areas of reading decoding and comprehension, mathematics applications and computation, and spelling. Norm-referenced measures are included in both forms. The Comprehensive Form also provides criterion-referenced assessment data to analyze the students’ errors in each subtest content area. Additionally, all standard scores in both forms are set at a mean of 100, with a standard deviation of 15, to allow for comparisons between the K-TEA and previously obtained standard IQ scores.

Scoring: The Comprehensive Form subtests include: Mathematics/Applications (60 items), Reading/Decoding (60 items), Spelling (50 items), Reading/Comprehension (50 items), and Mathematics/Computation (60 items). Raw scores can be converted to age and grade norms.

Reliability: The overall reliability coefficients ranged from .87 to .95 for all ages. Internal consistency showed strong reliability in this area. Coefficients ranged from .77 to .85 by grade level and from .82 to .88 by age group. Test-retest intervals ranged from 1 to 35 days. In most cases the results showed a .90 or better test-retest coefficient. Because there are two forms of the K-TEA (Brief and Comprehensive), reliability between the two was examined. The overall results showed interform reliability coefficients to be in the low .90s, with a range from .87 to .96 for the different grade levels and .90 to .97 for the separate age groups. Interform reliability coefficients in the areas of reading were higher (in the .90s) for younger children and fell into the upper .70s for older children.

Validity: Data that correlate performance on both forms of the K-TEA with other achievement tests are presented in the manual. The other tests included the Wide Range Achievement Test, the Peabody Individual Achievement Test (PIAT), the Metropolitan Achievement Test, the Stanford Achievement Test, and the K-ABC. The test authors present data to support relatively strong correlations on most of these measures (e.g., K-ABC ranged from .83 to .88; PIAT ranged from .75 to .86).

Norms: The standardization sample was representative with regard to geographic region, sex, socioeconomic status, and educational level of parents.

Suggested Uses: The K-TEA is recommended for achievement assessment in educational and research settings.
Lawrence Psychological-Forensic Examination

**Purpose:** Designed as a guide for comprehensive psychological assessment with clients in the criminal, juvenile, and civil justice systems.

**Population:** Clients within the criminal, juvenile, and civil justice systems.

**Score:** N/A.

**Time:** N/A.

**Author:** Stephen B. Lawrence.

**Publisher:** Unknown.

**Description:** The Lawrence Psychological-Forensic Examination (LAW-PSI) was gradually developed during ten years of practical experience working with clients, attorneys and judges within the criminal justice system. The LAW-PSI provides mental health professionals with a concise yet detailed handbook as a guide for comprehensive psychological assessments and report writing with clients in the criminal, juvenile, and civil justice systems. The author cautions that this instrument is not a definitive forensic tool, but rather a detailed guideline to aid both the skilled clinician in making a wide variety of clinical and legal judgments with criminal justice clients. These judgments often require a clinician to determine clients’ motivations, specific intentions and other legally precise mental states of mind at the time of the commission of alleged offenses which may have occurred days, weeks, or months in the past.

**Scoring:** N/A.

**Reliability and Validity:** Information on reliability and validity is not contained in the manual.

**Norms:** N/A.

**Suggested Uses:** Recommended uses include assessment of clients within the criminal justice system.
The Luria-Nebraska Neuropsychological Battery

**Purpose:** Designed to assess neurologically impaired patients.

**Population:** Ages 15 and over.

**Scores:** Fourteen scores.

**Time:** (90-150) minutes.

**Authors:** Charles J. Golden, Thomas A. Hammeke, and Arnold D. Purisch.

**Publisher:** Western Psychological Services.

**Description:** The Luria-Nebraska Neuropsychological Battery (LNNB) is based on the theory of higher cortical functioning by Aleksandr R. Luria. In contrast to the more psychometric and quantitative approach of most western clinical neuropsychologists, Luria utilized unstructured qualitative techniques in assessing neurologically impaired patients. Golden and his colleagues have attempted to integrate Luria’s techniques into American clinical neuropsychology.

**Scoring:** The LNNB generates fourteen scores: motor, rhythm, tactile, visual, receptive speech, expressive speech, writing, reading, arithmetic, memory, intellectual processes, pathognomic, left hemisphere, and right hemisphere.

**Reliability and Validity:** Several studies have reflected positively on the LNNB. When compared to the Halstead Neuropsychological Battery, they were found to be roughly equivalent in the hands of experienced clinical neuropsychologists in discriminating between a mixed group of psychiatric and brain-damaged populations. Each battery achieved an overall hit rate of approximately 80%. The LNNB was also found to separate brain-injured patients from pseudoneurological patients with an approximately 80% hit rate. The hit rate for various other groups are as follows: ideopathic seizure group (77.3%), seizures secondary to brain injury group (88.9%), and normals (87.5%).

**Norms:** Neurologically impaired patients.

**Suggested Uses:** Recommended uses include screening and neuropsychological evaluations for neurologically impaired patients.
McCarthy Scales of Children’s Abilities

Purpose: Designed to assess the abilities of preschool children.

Population: Children, ages 2.5 - 8.5.

Score: Six scale scores.

Time: (45-60) minutes.

Author: Dorothea McCarthy.

Publisher: The Psychological Corporation.

Description: The McCarthy Scales of Children’s Abilities (MSCA) is a measurement device used to assess the abilities of preschool children.

Scoring: The results of the MSCA produce six scale scores: verbal, perceptual-performance, quantitative, composite (general cognitive), memory, motor.

Reliability: In the MSCA manual, McCarthy provides information on the internal consistency and stability of test scores, as they were obtained from the standardization sample. The internal consistency coefficients for the General Cognitive Index (GCI) averaged .93 across 10 age groups between 2.5, and 8.5 years. Mean reliability coefficients for the other five Index Scales ranged from .79 to .88. The manual includes information on test-retest reliability over a one month interval on a stratified sample of 125 children grouped into three age levels. The average coefficient for the GCI was .80, with correlations ranging from .69 to .89 for the other scales. Other studies of long and short-term stability resulted in stability coefficients for the GCI of .81 and .84, respectively. Stability coefficients of the cognitive scales ranged from .62 to .76 with the Motor Scale emerging as the only scale that lacked stability (r=.33).

Validity: According to the manual, the content of the MSCA and the organization of the six scales were determined primarily through “intuitive and functional considerations” based on McCarthy’s extensive teaching and clinical experience. Analyses of the standardization data for five age groups and for separate groups of blacks and whites have given generally good support for the construct validity of the battery for normal children, although each factor did not emerge for every age group. The results also provide evidence that a child’s profile of MSCA Index scores reflect real and meaningful performance in domains of cognitive and motor ability. The major practical implication of these results for test users is that a child’s strengths and weaknesses can be determined through the interpretation of differences on Scale Indexes, as proposed originally by McCarthy.

Norms: The test was standardized on a sample of 1,032 children stratified by race, geographic region, father’s occupational status, and, informally, urban-rural residency, in accordance with the 1970 U.S. Census data. The major problem which has affected test users’ confidence in the meaning of the scores is the exclusion of exceptional children from the standardization sample.

Suggested Uses: The MSCA is a useful aid in screening and diagnostic decisions.
Millon Adolescent Personality Inventory

**Purpose:** Designed as an objective personality measure for adolescents.

**Population:** Ages 13 to 18 years.

**Score:** 8 personality-style dimensions.

**Time:** Not reported.

**Authors:** Theodore Millon, Catherine J. Green, and Robert B. Meagher, Jr.

**Publisher:** NCS Professional Assessment Services.

**Description:** The Millon Adolescent Personality inventory (MAPI) is an objective personality measure of the true-false type that assesses a number of personality dimensions, expressed concerns, and behavioral correlates in adolescents aged 13 to 18 years.

**Scoring:** The inventory comprises eight scales measuring personality styles based on Millon’s theory of personality types; four scales designed to tap subjects’ expressed concerns, such as peer security and acceptance of sexual maturation; and four scales assessing subjects’ actual behavior. The MAPI yields normative scores that are adjusted for personality-trait-prevalence data on each of the personality style, expressed concern, and behavioral correlate scales. A narrative report interprets scores on each of the three sets of scales, identifies noteworthy responses to individual items, lists applicable DSM-III-R diagnoses, and discusses therapeutic implications for use in treatment planning.

**Reliability:** Two test-retest studies produced stability coefficients generally within the acceptable range. An analysis of the internal consistency of all 20 MAPI scales produced a median reliability coefficient of .74, with a range from .67 to .84. The findings are within the acceptable range for scales of this type.

**Validity:** Internal validity was established by employing statistical correlation of each of the 150 test items with total scores on each of the 20 scales, subsequent to completion of the initial scale-construction stage. Those test items that displayed a moderate correlation (usually .30 or higher) with any scale other than a theoretically incompatible one were added to that scale. Further validity research was undertaken subsequent to the completion of the final form of the MAPI. In general, the 20 individual scales of the MAPI display moderate correlations in the expected direction with relevant scales of the California Psychological Inventory, 16 PF, and the Edwards Personal Preference Schedule.

**Norms:** The test was normed on groups of “clinical” and “nonclinical” subjects ranging in age from 13 to 19 years. The normal group consisted of 1,071 males and 1,086 females enrolled in public and parochial schools in a number of cities in various parts of the United States. The clinical group consisted of 430 adolescents, of whom 325 were outpatients and 105 were inpatients. The sample was 84% white, 11% black, 3% Hispanic, 1% Oriental, and 1% other. The standardization sample roughly matched the estimated percentages for socioeconomic status groups within the general population.

**Suggested Uses:** The MAPI is recommended for use in clinical and research settings.
**Millon Clinical Multiaxial Inventory**

**Purpose:** Designed to assess presence and type of psychopathology in respondent.

**Population:** Adults.

**Score:** Produces 20 scores.

**Time:** (20-40) minutes.

**Author:** Theodore Millon.

**Publisher:** NCS Interpretive Scoring Systems.

**Description:** The Millon Clinical Multiaxial Inventory (MCMI) is a 175-item true-false inventory which appears to answer the question: “What type of chronic psychopathology does the respondent have?” The MCMI is a direct operationalization of Millon’s taxonomy proposed in Modern Psychopathology (1969). The text presents a taxonomy which is to account for (1) severity of the disorder, (2) covariation or clustering of signs, (3) a four by two (active-passive) matrix of disorders, (4) a circumplicial set of related pathognomic signs, (5) continuity between the premorbid personality and severe impairment, and (6) the separation between biological, situational, and personality factors.

**Scoring:** The results produce 20 scores: 8 basic personality styles (schizoid, avoidant, dependent, histrionic, narcissistic, antisocial, compulsive, passive-aggressive), 3 pathological personality syndromes (schizotypal, borderline, paranoid), 6 symptom disorders scales of moderate severity (anxiety, somatoform, hypomanic, dysthymia, alcohol abuse, drug abuse), 3 symptom disorder scales of extreme severity (psychotic thinking, psychotic depression, psychotic delusions), plus 2 additional correction scales which provide a means to identify and adjust possible test-taking distortion. The manual provides clear administration instructions although there is an intentional unavailability of scoring templates. The test uses base rate scores rather than standard scores.

**Reliability:** Reliability studies show a 1-week test-retest reliability coefficients of an .87 average for Basic Personality Scales, .85 for Pathological Personality Syndromes, and .81 for Symptom Disorder Scales. Test-retest coefficients for 5-week intervals average .82, .77, and .67 for the three sets of scales. Kuder-Richardson 20 coefficients average .83, .90, and .82 for Basic Personality Styles, Pathological Personality Syndromes, and Symptom Disorders Scales, respectively.

**Validity:** In regards to validity, the MCMI correlates with the Symptom Distress Checklist-90 (SCL-90), the Psychological Screening Inventory (PSI), and the MMPI (basic scales plus Wiggins Content Scales) in theoretically expected and clinically meaningful patterns.

**Norms:** Inpatient and outpatient clinical sample.

**Suggested Uses:** The MCMI is designed for diagnostic use in clinical and research settings.
Millon Clinical Multiaxial Inventory - II

Purpose: Designed as a clinical measure to assist with psychiatric screening and with clinical diagnosis.

Population: Adult clinical populations.

Score: 10 clinical personality pattern scores.

Time: (25) minutes.

Author: Theodore Millon.

Publisher: National Computer Systems, Inc.

Description: Based on Millon’s theory of personality and psychopathology, the brief Millon Clinical Multiaxial Inventory-II (MCMI-II) instrument provides a measure of 22 personality disorders and clinical syndromes for adults undergoing psychological or psychiatric assessment or treatment. Specifically designed to help assess both Axis I and Axis II disorders, the MCMI-II instrument can assist clinicians in psychiatric diagnosis, developing a treatment approach that takes into account the patient’s personality style and coping behavior, and guiding treatment decisions based on the patient’s personality pattern.

Scoring: The MCMI-II consists of 10 clinical personality pattern scales, 3 severe personality pathology scales, 6 clinical syndrome scales, 3 modifier indices, 1 validity index.

Reliability: The reliability of the MCMI II generally has been sound, with the Axis II scales showing the highest stability as predicted by Millon. Normal subjects also had noticeably higher stability coefficients than clinical subjects. Millon also tested the stability of high point and double-high-point configurations. He reports that high point codes are fairly stable over a month, with nearly two thirds of 168 subjects achieving the same scale high point. For double-high-point configurations, 25% achieve the same high scores with another 19% achieving the same two scales but in reverse order. Based on part of his normative sample, Millon reports quite high internal consistencies. The average of 22 clinical scales is .89, and the range is from .81 to .95.

Validity: Because of extensive item overlap, we cannot be sure of the factor structure of this instrument. But there are also overlaps based on the overlap of the constructs; that is, the personality disorders are by no means distinct entities.

Norms: Norms for the MCMI-II instrument are based on a national sample of 1,292 male and female clinical subjects representing a variety of DSM-III and DSM-III-R diagnoses. The subjects included inpatients and outpatients in clinics, hospitals, and private practices. The MCMI-II manual describes the distribution of gender, age, marital status, religion, and other factors within the sample.

Suggested Uses: The MCMI-II is used primarily in clinical settings with individuals who require mental health services for emotional, social, or interpersonal difficulties.
Millon Clinical Multiaxial Inventory - 3rd Ed.

**Purpose:** Designed to provide treatment information in the areas of personality disorders and clinical syndromes.

**Population:** Adult clinical population 18 and over with 8th grade reading

**Scales:** 28 clinical personality scores Modified Indices (4), Clinical Personality Patterns (11), Severe Personality Pathology (3), Clinical Syndromes (7), and Severe Clinical Syndromes (3)

**Time:** approx. 25 minutes

**Author:** Theodore Millon, Roger Davis, and Carrie Millon

**Publisher:** NCS Pearson

**Description:** The Millon Clinical Multiaxial Inventory-3rd. Ed. (MCMI-III) is an update of the MCMI-II which represents ongoing research, conceptual developments, and the changes in the DSM-IV. It is a standardized, self-report questionnaire assessing a wide range of information related to personality, emotionality, and test-taking attitude. The MCMI-III contains 175 items which produce 28 scales. Changes to the MCMI-II include addition of the Depressive and PTSD scales. There are 90 new items and 85 that remained the same maintaining the 175 total items of the MCMI-II. Most of the changes had to do with the severity of the symptoms to increase the ability to detect pathology. The items per scale were reduced, keyings were reduced from 953 to 440, and the possible ratings per item reduced from 3 to 2 choices. The instrument is brief in comparison to other personality inventories, it has a strong theoretical basis, administration and scoring are simple, and it has a multi-axial format.

**Scoring:** The MMCI-III consists of 28 personality scales, Modifying Indices (Disclosure, Desirability, Debasement, Validity), Clinical Personality Patterns (Schizoid, Avoidant, Depressive, Dependent, Histrionic, Narcissistic, Antisocial, Aggressive (Sadistic), Compulsive, Passive-Aggressive (Negativistic), Self-Defeating), Severe Personality Pathology (Schizotypal, Borderline, Paranoid), Clinical Syndromes (Anxiety, Somatoform, Bipolar: Manic, Dysthymia, Alcohol Dependence, Drug Dependence, Post-Traumatic Stress Disorder), Severe Clinical Syndromes (Thought Disorder, Major Depression, Delusional Disorder).

**Reliability:** For the MCMI-III the internal consistency measures have been strong. The alpha coefficients exceed .80 for 20 of the 26 scales. The Depression scale had a high of .90, and the Compulsive scale was low at .66. Test-retest reliability, interval of 5 to 14 days, was equally high with a median of .91.

**Validity:** Factor analysis supports the organization of the scales. Many correlations have been made between the MCMI-III and related instruments, for example, the Beck Depression Inventory correlated with the MCMI-III Major Depression (.74) and Dysthymia (.71). High correlations were also found between the MMPI-2 and the MCMI-III Major Depression (.71) and Dysthymia (.68). While there have been some surprising results, with moderate or low correlations, most of the findings have been in the expected direction.

**Norms:** The instrument was normed with psychiatric patients and uses a new weighted score, the Base Rate Score (BRS) that takes into account the prevalence of the specific disorder in the psychiatric population. The normative data and transformation scores are based entirely on clinical samples and are applicable only to individuals who evidence problematic emotional and interpersonal symptoms or who are undergoing professional psychotherapy or a psychodiagnostic evaluation.

**Suggested Use:** MCMI-III is used primarily in clinical settings with individuals who require mental health services for emotional, social, or interpersonal difficulties. One reviewer states that he considers this test as one of the greatest contributions made to the field during his professional life.
Minnesota Multiphasic Personality Inventory

**Purpose:** Designed as an objective personality test for the assessment of psychopathology.

**Population:** Adults.

**Score:** Reports scores for the various scales.

**Time:** (40-90) minutes.

**Authors:** Starke Hathaway and J. Charnley McKinley.

**Publisher:** University of Minnesota Press.

**Description:** The Minnesota Multiphasic Personality Inventory (MMPI) is an objective verbal inventory designed as an personality test for the assessment of psychopathology consisting of 550 statements, 16 of which are repeated. The replicated statements were originally included to facilitate the first attempt at scanner scoring. Though they are no longer needed for this purpose, they persist in the inventory.

**Scoring:** The inventory is scored in subunits, eight of which are conventionally termed clinical scales and provide the clinical profile. The clinical scales are Scale 1 (Hypochondriasis); Scale 2 (Depression); Scale 3 (Hysteria); Scale 4 (Psychopathic Deviate); Scale 6 (Paranoia); Scale 7 (Psychasthenia); Scale 8 (Schizophrenia); and Scale 9 (Hypomania). Two other scales were added from within the original item pool. Scale 5 (Masculinity-femininity) was developed along with the eight clinical scales. Shortly after, Scale 0 (Social Introversion) was added. Three additional measures were designed to estimate the validity of the clinical profile. The L (Lie) Scale has 25 statements each dealing with a common, relatively insignificant weakness to which most people are willing to confess. The F (Infrequency) Scale is made up of 64 statements that were answered in the keyed direction by less than 10% of the inventory’s original standardization group. The K Scale was designed to trap the respondent who attempts to conceal actual psychopathology.

**Reliability:** Test-retest reliabilities reported in the manual range from the .50s to the low .90s. Certain scales, such as Scale 2, are quite variable over time, whereas other scales supposedly more “characterological,” such as Scale 1, have much higher test-retest reliabilities.

**Validity:** The validity of the MMPI varies with the population examined and the questions to be answered. The inventory has been the subject of thousands of studies and seemingly works best with diagnosing those who are severely disturbed and are demographically most like the original Minnesota normative sample (i.e., white and middle-class). The inventory seemingly is less valid with groups divergent from this population, such as those from different races or cultures.

**Norms:** The criterion groups were selected from patients at the University of Minnesota hospitals.

**Suggested Uses:** Uses include screening, assessment, selection, and prediction applications in both research and clinical settings.
Minnesota Multiphasic Personality Inventory-2

**Purpose:** Designed as an objective personality test for the assessment of psychopathology.

**Population:** Adults.

**Score:** Reports scores for the various scales.

**Time:** (40-90) minutes.

**Authors:** Starke Hathaway and J. Charnley McKinley.

**Publisher:** University of Minnesota Press.

**Description:** The Minnesota Multiphasic Personality Inventory-2 (MMPI-2) is a broad-based test designed to assess a number of major patterns of personality, emotional, and behavioral disorders. It consists of 567 statements that a subject mark true or false. The test provides internal checks that determine if the general requirements have not been satisfied.

**Scoring:** The inventory is scored in subunits, eight of which are conventionally termed clinical scales and provide the clinical profile. The clinical scales are Scale 1 (Hypochondriasis); Scale 2 (Depression); Scale 3 (Hysteria); Scale 4 (Psychopathic Deviate); Scale 6 (Paranoia); Scale 7 (Psychasthenia); Scale 8 (Schizophrenia); and Scale 9 (Hypomania). Two other scales were added from within the original item pool. Scale 5 (Masculinity-femininity) was developed along with the eight clinical scales. Shortly after, Scale 0 (Social Introversion) was added. Three additional measures were designed to estimate the validity of the clinical profile. The L (Lie) Scale has statements dealing with a common, relatively insignificant weakness to which most people are willing to confess. The F (Infrequency) Scale is made up of statements that were answered in the keyed direction by less than 10% of the inventory’s original standardization group. The K Scale was designed to trap the respondent who attempts to conceal actual psychopathology.

**Reliability:** Reliability information on the new MMPI-2 basic scales was computed from test-retest data on 82 males and 111 females retested after a median of 7 days. Correlation coefficients ranged from .67 to .92 for males (median r = .82), and from .58 to .91 for the females (median r = .79).

**Validity:** Additional validity indicators have been developed for MMPI-2. The FB (Back F) scales identifies individuals who are likely to be completing the inventory in an invalid manner. Also, the VRIN (Variable Response Inconsistency) and TRIN (True Response Inconsistency) scales provide an index of a subject’s tendency to respond in ways that are either inconsistent or contradictory.

**Norms:** The restandardization sample for the MMPI-2, comprised of subjects between the ages of 18 and 90 (1,138 males and 1,462 females) was drawn from communities in California, Minnesota, North Carolina, Ohio, Pennsylvania, Virginia, and Washington, in addition to proportionate samples of individuals from a federal Indian reservation and various military bases.

**Suggested Uses:** Uses include screening, assessment, selection, and prediction applications in both research and clinical settings.
**Minnesota Multiphasic Personality Inventory - Adolescent**

**Purpose:** Designed for use with adolescents to assess a number of the major patterns of personality and emotional disorders.

**Population:** Ages 14-18.

**Scores:** Sixty-eight scale scores.

**Time:** Administration time not reported.

**Authors:** James N. Butcher, Carolyn L. Williams, John R. Graham, Beverly Kaemmer, Robert P. Archer (manual); Auke Tellegen (manual), Yossef S. Ben-Porath (manual), S.R. Hathaway (test booklet), and J.C. McKinley (test booklet).

**Publisher:** University of Minnesota Press.

**Description:** The MMPI-A is largely a parallel inventory to the MMPI-2, designed to assess psychopathology in adolescents.

**Scores:** The MMPI-A contains 68 scales: 16 Basic Scales (6 Validity scales and 10 Clinical Scales), 28 Harris-Lingoes Subscales, 3 Si Subscales, 15 Adolescent Content Scales, and 6 Supplementary Scales.

**Reliability:** The manual contains 40 alpha coefficients for assessing internal consistencies. Of the 40, 17 (43%) range from .75 to .91, 18 (45%) range from .55 to .68; and the remaining 5 (13%) range from .35 to .53. Intercorrelations among the clinical scales range from .00 to .85 in the normative sample.

**Validity:** To the extent that the MMPI-A retains the essence of the MMPI, the basic clinical scales come complete with documented empirical validity. The validity of the 15 new content scales is purported to obtain external validity coefficients equal to that of its parent measure.

**Norms:** The normative sample is diverse and is probably representative of the adolescent population of the United States. The clinical sample is less representative than the normative sample: All are from settings in Minneapolis, and the majority are from drug and alcohol treatment.

**Suggested Uses:** This inventory is useful for assessing adolescent psychopathology.
Minnesota Percepto-Diagnostic Test

Purpose: Designed to measure visual perception and visual-motor abilities.

Population: Ages 5 and over.

Score: Not indicated.

Time: (8) minutes.

Author: Gerald B. Fuller.

Publisher: Clinical Psychology Publishing Company, Inc.

Description: The Minnesota Percepto-Diagnostic Test (MPD), is a clinical and educational test constructed to measure visual perception and visual-motor abilities in children and adults. The MPD offers prompt, objective evidence of differential status, providing measures that may be used in conjunction with other information to 1) discriminate among visual, auditory, and mixed learning disabilities; 2) classify children’s behavioral problems as normal, emotionally disturbed, schizophrenic, or organic; and 3) diagnostically separate normal, brain-damaged and personality-disturbed adults.

Scoring: Scoring involves an evaluation of reproduced figures in terms of degrees of rotation, distortion, and separation. Scores are then transferred to a diagnostic summary sheet.

Reliability: Split-half reliabilities for the MPD range from .52 (for 12 year-old subjects) to .86 (for 5 year old subjects) with a median of .60. In terms of test-retest reliability, using a three-month retest interval, stability coefficients range from .53 with 9 year-old subjects to .70 with 20 year-old subjects. Using a one-year interval, coefficients range from .37 to .60. Parallel form reliabilities (uncorrected) have a median value of .47 (median corrected value of .84).

Validity: Using rotation scores with a sample of 657 adults, the MPD correctly classified 81% as normal (N = 267), personality-disturbed (N = 211), or brain-damaged (N = 179). With 1,872 children, rotation scores correctly classified 82% as normal (N = 1,287), emotionally disturbed (N = 339), or either brain-damaged or schizophrenic (N = 246). The MPD rotation scores could not distinguish between schizophrenic and brain-damaged children. With respect to children’s reading ability, for a sample of 703 children, 90% could be appropriately diagnosed as demonstrating good reading, primary reading retardation, secondary reading retardation, or brain-damaged reading. The manual cites investigations indicating 85% and 78% hit rates, respectively, in differentiating between organic and nonorganic children.

Norms: The normative sample includes 4,000 children and adolescents (aged 5-20 years) and 657 adults.

Suggested Uses: The MPD is recommended as a brief screening device of learning disabilities and behavior problems in clinical or research settings.
Myers-Briggs Type Indicator

Purpose: Designed to classify individuals according to Jungian theory.

Population: High school and college students.

Score: Two types of scores on the four given dimensions.

Time: Not reported.

Authors: Isabel Briggs Myers and Katharine C. Briggs.

Publisher: Consulting Psychologists Press, Inc.

Description: The Myers-Briggs Type Indicator (MBTI) is a forced-choice, self-report inventory that attempts to classify individuals according to an adaptation of Carl Jung’s theory of conscious psychological type. There is the assumption that human behavior, perceived as random and diverse, is actually quite orderly and consistent. This view supposes that the observed variability is due to “certain basic differences in the way people prefer to use perception and judgment.” The MBTI should be regarded “as affording hypotheses for further testing and verification rather than infallible expectations of all behaviors.”

Scoring: The MBTI classifies individuals along four theoretically independent dimensions. The first dimension is a general attitude toward the world, either extraverted (E) or introverted (I). The second dimension, perception, describes a function and is divided between sensation (S) and intuition (N). The third dimension, also a function, is that of processing. Once information is received, it is processed in either a thinking (T) or feeling (F) style. The final dimension is judging (J) versus perceiving (P).

Reliability: With test-retest intervals from five weeks to 21 months, reliability coefficients range from .73 to .83 for E-I, .69 to .87 for S-N, .56 to .82 for T-F, and .60 to .87 for J-P. Phi coefficient estimates measuring internal consistency range from .55 to .65 (E-I), .64 to .73 (S-N), .43 to .75 (T-F), and .58 to .84 (J-P). These reliabilities are similar to other self-report inventories.

Validity: The MBTI manual provides correlational data with the Allport-Vernon-Lindzey Study of Values (AVL), the Cray-Wheelwright Psychological Type Questionnaire (also known as the Jungian Type Survey), The Edwards Personal Preference Schedule (EPPS), the Personality Research Inventory (PRI), the Scholastic Aptitude Test (SAT), and the Strong Vocational Interest Blank (SVIB), among others. The Sixteen Personality Factors Test (16PF) and the Rokeach Dogmatism Scale have also been correlated with the MBTI. These numerous correlational studies indicate that “. . . a wealth of circumstantial evidence has been gathered and results appear to be quite consistent with Jungian theory.” Examination of data on individual MRTI scales demonstrates the behaviors and attitudes which the MBTI appears to tap, suggesting a strong argument for construct validity.

Norms: No general adult sample distribution is available; high school or college distributions are the comparison groups. Little empirical information is available on minorities or on blue-collar workers.

Suggested Uses: Recommended for use in research or clinical settings.
NEO Personality Inventory - Revised

**Purpose:** Designed to measure five major dimensions or domains of normal adult personality.

**Population:** Ages 17 and older.

**Scores:** 30 facet scores and 5 domain scores.

**Time:** (30-40) minutes.

**Authors:** Paul T. Costa, Jr. and Robert R. McCrae.

**Publisher:** Psychological Assessment Resources, Inc.

**Description:** The NEO-PI-R is the most recent version of Costa and McCrae’s instrument to assess normal adult personality using the five-factor model taxonomy of personality. It is one of the few commercially available test based on this model.

**Scoring:** The NEO-P-R assesses five major domains of personality: Neuroticism (N), Extroversion (E), Openness to Experience (O), Agreeableness (A), and Conscientiousness (C), each represented by six lower level facet scale scores. It is available in three formats: self-report and observer-report versions and the NEO-FFI, a 60-item short form of the instrument.

**Reliability:** Domain level reliabilities range from .86 to .95 for both the self and observer rating forms of this instrument. Facet level reliabilities are good ranging from .56 to .90 for both self- and observer-report forms of the NEO-PI-R. Short-term test-retest reliability has been found with the NEO-FFI and the NEO-PI-R. Long-term test-retest reliability has been shown for the N, E, and O domains for the previous version of this instrument.

**Validity:** There is strong consensual validity between self, peer, and spouse reports of the test. Construct, convergent, and divergent validity evidence for the scales has been collected by Costa and McCrae. NEO-PI-R scales correlated with analogous scales from other instruments.

**Norms:** Norms are based on a sample of 1,000 subjects (500 males, 500 females) selected from three large scale studies of the NEO-PI-R. The normative sample was stratified to match 1995 U.S. Census projections for age, gender, and race. Separate norms are also provided for college-aged samples.

**Suggested Uses:** It is recommended that these scales are useful tools for personality assessment and may provide a useful bridge between basic research in personality psychology and applied psychology.
Paulhus Deception Scales

Purpose: Designed to measure the tendency to give socially desirable responses to tests.

Population: Ages 16 years and older.

Scales: Impression Management, Self-Deceptive Enhancement

Time: (5-7) minutes.

Author: Delroy L. Paulhus

Publisher: Multi-Health Systems Inc

Description: The Paulhus Deception Scales (PDS), also known as Version 7 of the Balanced Inventory of Desirable Responding (BIDR), is a 40 item self-report questionnaire designed to measure the tendency to give socially acceptable or desirable responses (SDR). It evolved from the earlier development and revisions of the BIDR, and measures two distinct forms of SDR; self deception and impression management. According to theories of deception styles, “self-deception” represents an unconscious process to deny psychologically threatening thoughts and feelings reflective of psychoanalytic conflicts, and “other-deception” represents conscious distortion toward self-enhancement.

Scoring: The two scales of PDS are designed to capture two different styles of responses considered socially desirable. Impression Management (IM) involves conscious use of inflated self-descriptions, faking, or lying, and is thought to indicate hypersensitivity to situational self-presentation demands. Self-Deceptive Enhancement (SDE) intends to capture the tendency to give honest but inflated self-descriptions reflecting a lack of insight and an unconscious bias toward favorable self-portrayal. Item content of the two subscales was rationally composed to reflect and distinguish the two respective biases in self-report.

Reliability: Coefficient alpha for internal reliability for the PDS subscales and the total PDS score were satisfactory for all four samples. The coefficients for the SDE scales ranged from .70-.75 and IM and PDS total coefficients ranged from .81-.86.

Validity: In factor analysis, the SDE is strongly associated with other factor measures of desirable responding, and IM is grouped with the second factor. The IM scale correlates highly with a cluster of measures known as lie scales, and role playing measures. Attempts have been made to tie high SDE subject to narcissism or certain defense mechanisms. A number of studies of the convergent validity, structural validity, and discriminant validity of the PDS and its subscales were reported in the manual. There is a considerable body of literature using the PDS and the BIDR for clinical and forensic purposes that is not described in the manual, which provides additional confirmation of the value of the PDS as a measure of socially desirable responding. Overall, the studies cited and the validation process meet the psychometric standards for test validation as specified in the Standards for Educational and Psychological Testing.

Norms: The PDS was standardized using a large adult sample \( n = 441 \) from the general population, as well as samples from college student, military, and prison populations collected by the author and other researchers using the PDS. Information on the sample demographics and the sampling techniques used are not provided in the test manual and issues of sex or age differences in socially desirable responding are not addressed.

Suggested Use: The PDS is useful in identifying individuals who distort their responses and in evaluating the honesty of their responses, as it is administered concurrently with other instruments. The PDS should be regarded as a valuable tool to be used in any high-demand testing situation as a check on the validity of self-report test responses. However, some concerns have been raised about ongoing controversies in social desirability research and the need for more research to further validate the PDS in forensic and human resources settings. The indications that narcissistic tendencies are associated with a low IM-high SDE profile, or that a “repressor pattern” is associated with a high IM-high SDE profile, remain an empirical question. Nonetheless, the relationship between the various IM-SDE combinations and the respective response and personality styles generates useful and interesting hypotheses to be tested in research and clinical practice.
Peabody Individual Achievement Test

Purpose: Designed as a measure of academic achievement.

Population: Grades K-12.

Score: 6 scores.

Time: (30-40) minutes.

Authors: Lloyd M. Dun and Frederick C. Markwardt, Jr.

Publisher: American Guidance Service, Inc.

Description: The Peabody Individual Achievement Test (PIAT) is an individually administered measure of academic achievement that is norm-referenced. The test was designed to provide a wide-range screening measure in five content areas that can be used with students in kindergarten through the twelfth grade. The content areas covered by the PIAT are 1) mathematics, 2) reading recognition, 3) reading comprehension, 4) spelling, and 5) general information. The test materials are presented in two easel-kit volumes that present the stimulus items to the students on one side while the examiner can see both the stimulus items of the students and the examiner’s instructions on the reverse side. This easel-kit format facilitates the assessment process and is easy to work with for the examiner and the student.

Scoring: The PIAT produces six scores from corresponding domains: mathematics, reading recognition, reading comprehension, spelling, general information, and a total score.

Reliability: Test-retest reliability coefficients (Pearson product-moment correlations) were calculated based on sample retesting of 50-75 students in Grades K, 1, 3, 5, 8, and 12. There was a month interval between testings. The resultant reliability coefficients ranged from .42 in kindergarten for Spelling to .94 in the third grade for Reading Recognition. The overall median reliability coefficient was .78. In terms of median coefficient values the greatest confidence in stability is in the total test (.89) and Reading Recognition (.89) and least in Reading Comprehension (.64) and Spelling (.65).

Validity: It is believed that the rigorous item selection used in the development of the PIAT supports adequate content validity. Concurrent validity was calculated by comparing the scores of the PIAT to a measure of scholastic aptitude, the Peabody Picture Vocabulary Test (PPVT), Form A. The resultant product-moment correlation coefficients ranged from a median of .42 in kindergarten to a median of .69 in third grade. The range for the subtest coefficients ranged from .40 in Spelling to a median of .68 in General Information. The overall coefficient for the subtests of the PIAT with the PPVT, Form A, was .57.

Norms: The total sample was composed of 2,559 students. These students were randomly selected from public school classrooms and balanced according to sex (50% male, 50% female), age, race (84.4% white, 11.3% black, 4.3% other), and socioeconomic level.

Suggested Uses: The PIAT is recommended as a screening measure for academic achievement in educational, clinical, and research settings.
Peabody Individual Achievement Test -- Revised

**Purpose:** Designed to obtain a survey of an individual’s scholastic attainment.

**Population:** Grades K-12.

**Score:** Scores in six content areas.

**Time:** Untimed.

**Author:** Frederick C. Markwardt, Jr.

**Publisher:** American Guidance Service.

**Description:** The Peabody Individual Achievement Test (PIAT-R) is the revised version of the now classic Peabody Individual Achievement Test (PIAT), an individually administered, norm-referenced measure of academic achievement. The test was designed to provide a wide-range screening measure in six content areas that can be used with students in kindergarten through the 12th grade.

**Scoring:** The content areas covered by the PIAT-R are 1) General Information, 2) Reading Recognition, 3) Reading Comprehension, 4) Mathematics 5) Spelling and 6) Written Expression.

**Reliability:** Split-half reliability of the PIAT-R ranges from .84 for kindergarten Mathematics to .98 for third-grade Reading Recognition. The median split-half reliability for the total test is .98. With three exceptions, all at age 5, Kuder-Richardson reliabilities all cluster in the low to mid-90s, indicating a high degree of content homogeneity. Test-retest reliabilities for 2- to 4-week periods range in the low to upper .90s for composite correlations and in the mid-.80s to high .90s for individual subtests. The intrarater reliabilities for Level II Written Expression range from .49 for Grades 5 and 8 to .81 for Grade 11.

**Validity:** Content validity of the PIAT-R was established through the selection of items in the test. It is believed that the rigorous item selection used in the development of the PIAT-R supports adequate content validity. Construct validity was established by three methods: 1) by developmental changes, the extent to which test scores show an increase with age or grade; 2) by correlations with other tests, including the original FIAT (median r = .88 for the total test) and the Peabody Picture Vocabulary Test-Revised (median r = .72); and 3) factor analytic techniques.

**Norms:** The total sample was composed of 1,563 students from 33 communities, 100 to 150 from each grade (K-12), and an additional 159 from the kindergarten level. These students were selected randomly from public school classrooms (91.4%) and 11 private schools, and they were balanced according to sex (49.9% male, 50.1%), age, race (73.3% white, 14.3% black, 9.7% Hispanic, 3.2% other), and socioeconomic level.

**Suggested Uses:** Designed to assess student achievement in educational or research settings.
Peabody Picture Vocabulary Test -- Revised

**Purpose:** Designed primarily to measure a subject’s receptive (hearing) vocabulary for Standard American English.

**Population:** Ages 2.5 to 40.

**Score:** 1 score.

**Time:** (10-20) minutes.

**Authors:** Lloyd M. Dunn, Leota M. Dunn, Gary J. Robertson, and Jay L. Eisenberg.

**Publisher:** American Guidance Service.

**Description:** The Peabody Picture Vocabulary Test - Revised (PPVT-R) was designed as a measure of receptive language and has replaced the original Peabody Picture Vocabulary Test (PPVT) published in 1959. The PPVT-R retains many features of its predecessor. For example, it consists of two forms (L and M), allows a verbal or nonverbal response, is individually administered, and is untimed (although administration time typically requires only 15-20 minutes). The PPVT-R contains 350 items making it more reliable and discriminating than the PPVT, which had only 300 items.

**Scoring:** The score is obtained by subtracting errors from total ceiling score.

**Reliability:** The test manual reports internal consistencies from .61 to .88, and alternate form reliability values from .71 to .91 from the standardization sample. Also, when alternate-form equivalency is examined by comparing means, Form M typically produces slightly higher standard score equivalents than does Form L.

**Validity:** When PPVT-R standard scores are compared to PPVT IQS, WISC-R IQS, and Stanford-Binet IQS, the mean PPVT-R scores have been significantly lower. However, results from two studies indicate comparable mean standard scores between the PPVT-R and McCarthy Scales of Children’s Abilities. Correlations ranging from .16 to .78 have been obtained from studies examining the relationship between the PPVT-R standard score equivalents and various “g factor” scores such as Wechsler Full Scale IQS, Stanford-Binet IQS, and McCarthy Scales of Children’s Abilities GQIs; correlation coefficients ranging from .40 to .60, however, are more common. The modest coefficients of determination emerging from these correlations suggest limited shared variance; thus PPVT-R scores should not be interpreted as intelligence test scores. No predictive validity (or long-term temporal stability) data are available.

**Norms:** The PPVT-R was standardized in 1979 on a sample of 4,100 children and youth in the age range from 2 to 19 years with 100 children of each sex at each age level. It was also standardized on a sample of 828 adults with approximately 200 adults for each of the age ranges from 19 to 24, 25 to 29, 30 to 34, and 35 through 40 years. The sample of children was selected to approximate the 1970 U.S. Census data for sex and age, geographic, occupational background, racial-ethnic, and urban-rural population distributions.

**Suggested Uses:** Recommended uses include research and educational applications.
Peabody Picture Vocabulary Test - 3rd Edition

Purpose: Designed as a test of receptive vocabulary achievement and verbal ability.

Population: (English speaking) individuals ages 2:6 through 90+ years

Score: Standard scores range from 40 to 160.

Time: [11-12] minutes for most people.

Author: Lloyd M. Dunn and Leota M. Dunn

Publisher: American Guidance Service

Description: The Peabody Picture Vocabulary Test - Third Edition (PPVT-III) updates the PPVT of 1959 and the PPVT-R of 1981, and like them is an individually administered, un-timed, norm-referenced, wide-range test with two parallel forms. The 204 items on each form are grouped into 17 sets of 12 items of 4 black and white illustrations forming a picture plate. The original PPVT used 300 stimulus words, 150 on each form of the test. In the development of the PPVT-III item pool, national tryouts were used to eliminate stimulus words that were biased by gender, region, race or ethnicity. The test has two parallel forms, requires no oral or written responses and no reading by the examinee.

Scoring: Rapid and objective scoring is done while the test in being administered. Raw scores are converted into normative scores.

Reliability: The alpha reliabilities for the 25 standardized age groups were between .92 and .98 with a median reliability of .95 for both forms. The split-half reliabilities for the 25 age groups ranged from .86 to .97, with a median of .94 for both forms. The alternate forms reliabilities range from .88 to .96 with a median correlation of .94.

Validity: Correlations of the PPVT-III form A and B scores with scores of the WISC-III VIQ are .91 and .92; with the KAIT Crystalized IQ, .87 and .91; and with the K-BIT Vocabulary score, .82 and .80. Slightly lower correlations with the non-verbal scores were predicted; with the WISC-III PIQ .82 and .84; KAIT Fluid IQ, .76 and .85, and with the K-BIT matrices score, .65 and .62. Full scale IQ on the WISC-III was .90 and .90, KAIT Composite IQ was .85 and .91, K-BIT composite, .78 and .76.

Norms: The normative sample included 2725 persons. And while the original PPVT was standardized only on white children from Tennessee, the normative sample of the PPVT-III was selected to match the data of the 1994 US Census. The sample was stratified with each age group by gender, race/ethnicity, geographic region, and SES. Only individuals who were determined to speak and understand English were included in the testing.

Suggested Use: The PPVT-III is useful in testing preschool children for vocabulary acquisition as an indicator of linguistic and cognitive development. When English is the language of home, school, and community, it can be used as a screening device for verbal ability and for giftedness and for mental retardation. It is a useful test for people who perform poorly on group tests, or for nonreaders and for those with written language problems. For individuals with language impairment, it can provide a measure of linguistic potential, and it is useful for those who are autistic, withdrawn, or have psychotic problems, as speech is not required. Individuals with cerebral palsy or with visual-perceptual problems can also take this test.
Personal Assessment Inventory - IPAT Depression Scale

Purpose: Designed to assess a unitary factor of depression.

Population: Adults.

Score: Yields a total score.

Time: (10-20) minutes.

Authors: Samuel Krug and James E. Laughlin.

Publisher: Institute for Personality and Ability Testing.

Description: The developers of the IPAT Depression Scale report that they intended to develop a psychometrically sophisticated measure of depression that could be used with the IPAT Anxiety Scale Questionnaire. This 40-item depression scale was derived from factor analysis of the primary pathology factors of the Clinical Analysis Questionnaire. The authors claimed to have identified a unitary second-order depression factor, after which, they selected items from the item pool that best discriminated the depressive sample from normal and other clinical categories.

Scoring: The number of items endorsed that load for depression are summed and that score is interpreted in relation to suggested cut-off scores.

Reliability: The scale shows an internal consistency (alpha) range from .85 to .94 on various normal and clinical groups.

Validity: Reviewers have suggested that the validity data included in the manual are contaminated by the use of the same subjects employed in item selection. A correlation of .88 was obtained between the scale and the “pure depression factor in a sample of 1,904 normals and clinical cases” which comprised the CAQ norm groups. The correlation between the MMPI D scale and the depression scale is .31.

Norms: The handbook refers to a group of “67 clinically diagnosed depressives” together with 729 other “clinical cases” that were differentially diagnosed against 632 normals. It is not clear how these subjects were selected.

Suggested Uses: Due to the scale’s questionable reliability and validity, reviewers have recommended more extensive research on this measure before endorsing its clinical utility.
**Personality Assessment Inventory**

**Purpose:** Designed to provide information relevant to clinical diagnosis, treatment planning, and screening for psychopathology.

**Population:** Ages 18 years to adult.

**Scales:** Inconsistency, Infrequency, Negative Impression, Positive Impression, Somatic Complaints, Anxiety, Anxiety-Related Disorders, Depression, Mania, Paranoia, Schizophrenia, Borderline Features, Antisocial Features, Alcohol Problems, Drug Problems, Aggression, Suicidal Ideation, Stress, Nonsupport, Treatment Rejection, Dominance, Warmth.

**Total Time:** (40-50) minutes.

**Author:** Leslie C. Morey

**Publisher:** Psychological Assessment Resources

**Description:** The Personality Assessment Inventory (PAI), is a self-report inventory of adult psychopathology. It was designed as a multidimensional alternative to the Minnesota Multiphasic Personality Inventory (MMPI) for assessing abnormal personality traits. The PAI is a self-report questionnaire consisting of 344 items (scored on a 4-point ordinal scale: F = False; ST = Slightly True; MT = Mainly True; VT = Very True). The PAI includes current items, and avoids colloquial and slang expressions. Items considered potentially biased (on gender, ethnic, economic, religious or other grounds) were excluded. The PAI manual is both comprehensive and informative.

**Scoring:** The PAI has 22 non-overlapping scales which include 4 validity scales, 11 clinical scales, 5 treatment scales, and 2 interpersonal scales [10 scales are further subdivided into 31 conceptually distinct subscales]. Most scales consist of 8, 12, or 24 items with an average grade 4 reading level. Validity scales measure response Inconsistency, Infrequency, Negative Impression, and Positive Impression. Raw scores are plotted on the Profile Forms, yielding T scores (M=50,SD=10).

**Reliability:** Alpha coefficients of internal consistency for the 22 scales were median .81, median .82, and median .86 for the normative, college, and clinical samples. Inter item correlations were low, indicating independent content of most items within each scale Median test-retest coefficients over 2-4 weeks showed stability with median alphas exceeding .80.

**Validity:** Concurrent validity correlations of the PAI validity, clinical, treatment, and interpersonal scales with several other personality instruments (e.g., MMPI, STAI, Beck Scales, Wahler Physical Symptoms Inventory, Fear Survey Schedule) reveal many small to moderate coefficients, suggesting only relatively modest common variance. Exploratory factor analyses based on the scale and subscale intercorrelations for the standardization and clinical samples are methodologically questionable.

**Norms:** There were three samples, all comprised of at least 1,000 individuals; community-dwelling adults stratified on gender, race, and age according to 1995 U.S. Census projections, clinical patients; and college students (all samples comprised at least 1,000 individuals).

**Suggested Use:** The PAI is intended to provide information relevant to clinical diagnoses, treatment planning and screening for psychopathology.
Personality Research Form

**Purpose:** Designed to assess personality characteristics.

**Population:** Grades 7-16 and adults, college.

**Score:** 15 scores.

**Time:** (30-45) minutes.

**Author:** Douglas N. Jackson.

**Publisher:** Research Psychologists Press, Inc.

**Description:** Development of the Personality Research Form (PRF) was guided by the belief that more rigorous and valid assessment of personality characteristics could be achieved through the application of modern principles of personality and test theory. The goals established for the PRF were to develop an item pool and a set of personality scales relevant to normal human functioning in a wide variety of situations.

**Scoring:** There are 15 scores, which include: achievement, affiliation, aggression, autonomy, dominance, endurance, exhibition, harm avoidance, impulsivity, nurturance, order, play, social recognition, understanding, and infrequency.

**Reliability:** The odd-even reliabilities, adjusted using the Spearman-Brown correction, were calculated from the responses of 192 subjects. The reliability estimates for the personality scales ranged from .48 to .90, with a median reliability of about .78. The reliabilities of the Desirability (validity) Scale ranged from .59 to .66, and those of the Infrequency (validity) Scale ranged from .33 to .57. Test-retest reliability estimates ranged from .57 to .85, with a median of .77.

**Validity:** The PRF manual contains nine tables reporting correlations between scales on various forms of the PRF and scales on other tests. Reported are the relationships between the PRF-AA and the CPI and Strong Vocational Interest Blank (SVIB), the PRF-A and the Allport Vernon Lindzey Study of Values (SOV), and the PRF-E and the JPI, Jackson Vocational Interest Survey (JVIS), Bentler Psychological Inventory (BPI), Bentler Interactive Psychological Inventory (BIPI), and Cattell’s High School Personality Questionnaire (HSPQ). In general, the PRF scales have higher correlations with scales from the other instruments measuring the same or a similar construct (e.g., dominance-leadership) than with scales measuring different or antithetical constructs.

**Norms:** The normative sample consisted of 1,029 male and 1,002 female college students selected to represent a stratified random sample by regions of the United States.

**Suggested Uses:** The PRF is recommended for use in personality research and for measuring normal personality traits in settings such as schools, colleges, clinics, guidance centers, business, and industry.

The Porteus Maze Test

**Purpose:** Designed as a nonverbal test of performance intelligence.

**Population:** Ages 3 and over.

**Scores:** 2 scores, Test Age and qualitative or Q score.

**Time:** (15-60) minutes.

**Author:** Stanley D. Porteus.

**Publisher:** The Psychological Corporation.

**Description:** The Porteus Maze Test is a nonverbal test of performance intelligence. It is a graded set of paper forms on which the subject traces the way from a starting point to an exit; the subject must avoid blind alleys along the way. There are no time limits. The mazes vary in complexity from simple diamond shape for the average three-year-old to intricate labyrinths for adults. There are three sets of mazes: the original (the Vineland series), and two supplements, the Exten-
Scoring: Administration must be studied from Porteus’ instructions. The scoring instructions are also included in Porteus’ book. The scoring results in a Qualitative or Q score and a Test Age.

Reliability: This information is not included in the manual.

Validity: Many early studies of correlations with other intelligence tests ranged widely from moderate to high coefficients. Because the Porteus is a nonverbal performance test, one does not expect a high correlation with purely verbal tests, but the correlation should be at least moderately positive. Researchers note that a correlation coefficient of .50 is representative of studies with other IQ tests. Correlations with intelligence tests involving spatial abilities, such as the Kohs Blocks and Knox Cubes, are particularly high.

Norms: This information is not included in the manual.

Suggested Uses: Reviewers do not recommend this test for clinical use, but suggest that further research is necessary to validate the author’s claims.
Psychological Screening Inventory

Purpose: Designed as a brief mental health screening device.

Population: Ages 16 and over.

Score: Five scores.

Time: (15) minutes.

Author: Richard Lanyon.

Publisher: Research Psychologists Press, Inc.

Description: The Psychological Screening Inventory (PSI) is a 130-item, true-false, self-report questionnaire designed as a screening device "to be used in detecting persons who might profitably receive more intensive attention". It not intended to be a diagnostic instrument but to detect people over the age of 16 who should be referred for more extensive examination or therapy. The PSI can be administered either individually or in a group setting.

Scoring: The scales for the PSI are as follows: Discomfort (Di), Expression (Ex), Alienation (Al), Social Nonconformity (Sn), and Defensiveness (De). These five primary scales are transformed to T-Scores with a mean of 50 and a standard deviation of 10. The conversion from raw scores to T-Scores can be done by either looking up the values in a table in the manual or by plotting the scores directly onto a profile sheet. Separate norms exist for male and females.

Reliability: The internal consistencies of the five scales ranged from a low of .51(De) to .85 (Di). After ten days, the test-re-test reliability ranged from .73 (Al) to .89 (Sn); after one month, both Al and De correlated .66, whereas the reliability for Sn rose to .95.

Validity: Many studies have correlated the PSI scores with other personality instruments, such as the MMPI, California Psychological Inventory (CPI), the Eysenck Personality Inventory (EPI), and the Maudsley Personality Inventory (MPI). In general, the PSI appears to have good construct validity. The Alienation Scale correlates with MMPI Scales F, Sc, Pa and, to a lesser degree, with Pt Scales, which tap varying degrees of pathology. Social Nonconformity correlates with Fd on the MMPI and a separate scale of sensation seeking and has a strong negative correlation with the Marlowe-Crowne Social Desirability Scale (SDS). The Discomfort Scale correlates well with various indices of anxiety, such as the MMPI’s Pt Scale and the MPI’s Neuroticism Scale, whereas the Extraversion Scale has a strong negative correlation with Si and a strong, positive relationship with the MPI and EPI Extraversion Scale. The Defensiveness Scale is closely associated with the MMPI’s L and K Scales.

Norms: 100 males and 100 females, representative of the age distribution and mean educational level of the country, comprised the standardization group.

Suggested Uses: The PSI is recommended as a screening device in clinical and research settings.
Psychosocial Pain Inventory

Purpose: Designed to assess the influence of psychosocial factors in chronic pain syndromes.

Population: Adults.

Score: Total score.

Time: (120) minutes.

Author: Robert K. Heaton, Ralph A. W. Lehman, and Carl J. Getto.

Publisher: Psychological Assessment Resources, Inc.

Description: The Psychosocial Pain Inventory (PSPI) is a 25-item, eight-page structured interview designed to assess the influence of psychosocial factors in chronic pain syndromes. The instrument’s development is based on solid evidence that suggests that psychosocial factors contribute significantly to an individual’s reaction and adjustment to chronic pain. Items for the PSPI were derived from the research and theory.

Scoring: Psychosocial factors evaluated in this instrument include the following: 1) pain behavior (e.g., up time, time in bed), 2) social reinforcement, 3) life changes, 4) litigation, 5) financial status, 6) use of alcohol, 7) medication use, 8) coping strategies, 9) social environment, and 10) environmental stress. In addition, as a part of the assessment patients are asked questions concerning their personal and family histories, past and current medical histories, and reactions and adjustments to the pain and medical treatments.

Reliability and Validity: Essentially only one study reporting data regarding the Psychosocial Pain Inventory has appeared in the literature. No reliability studies have been completed by the authors. In addition, the validity data reported are sparse. Nevertheless, the PSPI does appear to have fairly good discriminant power. In a sample of 32 chronic pain sufferers researchers used the PSPI as an independent variable and report preliminary findings in which a stepwise-discriminant-function analysis was successful in predicting 94.4% of the patients who successfully completed medical treatment for pain and 85% of those 32 patients who did not respond to medical treatment.

Norms: The instrument was validated on 169 chronic pain patients (95 women and 24 men) seen at the Multidisciplinary Pain Clinic of the University of Colorado Health Science Center, with the average age being 45.3 years (SD = 12.5). Despite the average age of the normative sample, the instrument can be adapted to assess the functioning of all adult chronic pain sufferers.

Suggested Uses: The PSPI is recommended as a measure to assess the psychosocial factors influencing pain in research and clinical settings.
Questionnaire on Resources and Stress

Purpose: Designed to assess families’ coping and adaptational responses to a disabled family member.

Population: Any family member.

Score: Percentile norms.

Time: (60) minutes.

Author: Jean Holroyd.

Publisher: Clinical Psychology Publishing Company, Inc.

Description: The Questionnaire on Resources and Stress (QRS) for Families with Chronically Ill or Handicapped Members was developed for the quantitative assessment of families’ coping and adaptational responses to a disabled family member. The QRS consists of 285 self-administered, true-false items that form 75 rationally derived scales. These scales are grouped into three general response categories: Personal Problems (seven scales), Family Problems (three scales), and Problems of Index Case (five scales). The QRS is intended to measure sources of stress and variables that attenuate stress as well as family members’ responses to stress. A 66-item short form (QRS-SF) can be used for screening purposes.

Scoring: The 15 scales on the full-length QRS are grouped and identified as follows: Personal Problems scales--Poor Health/Mood, Excess Time Demands, Negative Attitude toward Index Case, Overprotection/Dependency, Lack of Social Support, Overcommitment/Martyrdom, and Pessimism; Family Problems scales--lack of Family Integration, Limits on Family Opportunity, and Financial Problems; and Problems of Index Case scales--Physical Incapacitation, Lack of Activities for Index Case, Occupational Limitations for Index Case, Social Obtrusiveness, and Difficult Personality Characteristics. An appendix in the manual lists percentile norms for special populations.

Reliability: Reliability reported in the manual is limited to Kuder-Richardson 20 estimates for internal consistency. The KR-20 reliability correlation was .96 for the full-length QRS. Scale coefficients ranged from as low as .24 to .88 on the QRS, scales with fewer items produced lower internal consistency coefficients. Test-retest reliability coefficients for the QRS are not available.

Validity: The manual reports that QRS content validity was established “by having 12 experts select the most relevant items from a large pool during initial questionnaire development.” Criterion and construct validity are not systematically addressed by the authors.

Norms: Norms for special populations of families with members with different handicapping conditions have been established for four major classifications: developmental disabilities, psychiatric problems, chronic medical illness, and neuromuscular disease. These norms are based on 329 cases, including 145 developmentally disabled children, 98 children with psychiatric problems, 49 children with medical illnesses (renal disease, leukemia, cystic fibrosis), and 37 children with neuromuscular diseases (Duchenne’s dystrophy, cerebral palsy).

Suggested Uses: The QRS is recommended for clinical and research settings.
Quick Neurological Screening Test, Revised Edition

**Purpose:** Designed for use in screening for early identification of disabilities.

**Population:** Ages 5 and over.

**Score:** Norms suggest cutoff scores.

**Time:** (20) minutes.

**Authors:** Margaret Mutti, Harold M. Sterling, and Norma V. Spalding.

**Publisher:** Academic Therapy Publications.

**Description:** The Quick Neurological Screening Test (QNST) is composed of 15 observed tasks that reportedly can be used as a screening test for learning disabilities. These tasks are very simple in nature and were adapted primarily from a typical pediatric neurological examination; however, a few tasks were derived from developmental scales or neuropsychological tests.

**Scoring:** Subjective scoring is required for the tasks, which include: handwriting ability, perceptual ability for numbers written on the palms of the hands, eye tracking, finger to nose coordination, rapidly reversing repetitive hand movements, tandem walk, and arm and leg extension. The test requires that the examiner be highly observant of the child’s behavior and make subjective ratings concerning the child’s performance. These subjective ratings are then compared to cutoff scores in the manual.

**Reliability:** Although subjective scoring is involved in the test, no direct measure of scorer reliability is presented in the manual. Indirect evidence suggests that there is some examiner bias. For example, in one study a test-retest reliability coefficient of .81 is reported after a month interval for 33 learning disabled children who were tested by a single examiner. A single examiner is likely to exercise the same scoring bias on two administrations. A lower reliability coefficient of .71 was reported in another study after a 1 month interval with two different examiners. Apparently one examiner administered the first test and another examiner the follow-up test. The difference between these two correlations may imply that individual examiners employ slightly different criteria in scoring even though both attempted to follow the instructions.

**Validity:** The QNST seems to be best for matching the findings of a standard pediatric neurological examination. In one study of over 550 subjects, 30% of which had positive neurological findings, the QNST was abnormally high in 98% of these. No patient had a positive neurological examination and a QNST in the normal range. A major problem with the QNST, however, is that a large unspecified number of subjects had abnormally high QNST scores and no positive finds on neurological examination.

**Norms:** 2,239 subjects from learning disabled and undifferentiated populations.

**Suggested Uses:** It is recommended that the QNST could be included as only one test in a battery of neuropsychological tests for learning disabilities.
Raven Standard Progressive Matrices

**Purpose:** Designed to measure a person’s ability to form perceptual relations.

**Population:** Ages 6 to adult.

**Score:** Percentile ranks.

**Time:** (45) minutes.

**Author:** J.C. Raven.

**Publisher:** U.S. Distributor: The Psychological Corporation.

**Description:** The Standard Progressive Matrices (SPM) was designed to measure a person’s ability to form perceptual relations and to reason by analogy independent of language and formal schooling, and may be used with persons ranging in age from 6 years to adult. It is the first and most widely used of three instruments known as the Raven’s Progressive Matrices, the other two being the Coloured Progressive Matrices (CPM) and the Advanced Progressive Matrices (APM). All three tests are measures of Spearman’s g.

**Scoring:** The SPM consists of 60 items arranged in five sets (A, B, C, D, & E) of 12 items each. Each item contains a figure with a missing piece. Below the figure are either six (sets A & B) or eight (sets C through E) alternative pieces to complete the figure, only one of which is correct. Each set involves a different principle or “theme” for obtaining the missing piece, and within a set the items are roughly arranged in increasing order of difficulty. The raw score is typically converted to a percentile rank by using the appropriate norms.

**Reliability:** Internal consistency studies using either the split-half method corrected for length or KR20 estimates result in values ranging from .60 to .98, with a median of .90. Test-retest correlations range from a low of .46 for an eleven-year interval to a high of .97 for a two-day interval. The median test-retest value is approximately .82. Coefficients close to this median value have been obtained with time intervals of a week to several weeks, with longer intervals associated with smaller values. Raven provided test-retest coefficients for several age groups: .88 (13 yrs. plus), .93 (under 30 yrs.), .88 (30-39 yrs.), .87 (40-49 yrs.), .83 (50 yrs. and over).

**Validity:** Spearman considered the SPM to be the best measure of g. When evaluated by factor analytic methods which were used to define g initially, the SPM comes as close to measuring it as one might expect. The majority of studies which have factor analyzed the SPM along with other cognitive measures in Western cultures report loadings higher than .75 on a general factor. Concurrent validity coefficients between the SPM and the Stanford-Binet and Weschler scales range between .54 and .88, with the majority in the .70s and .80s.

**Norms:** Norm groups included in the manual are: British children between the ages of 6 and 16; Irish children between the ages of 6 and 12; military and civilian subjects between the ages of 20 and 65. A supplement includes norms from Canada, the United States, and Germany.

**Suggested Uses:** Recommended uses include measurement of a person’s ability to form perceptual relations and reason by analogy in research settings.
Rogers Criminal Responsibility Assessment Scales

**Purpose:** Designed to assist psychologists involved in forensic practice.

**Population:** Criminals.

**Score:** 5 scales.

**Time:** Administration time not reported.

**Author:** Richard Rogers.

**Publisher:** Psychological Assessment Resources, Inc.

**Description:** The Rogers Criminal Responsibility Assessment Scales (R-CRAS) were developed by Richard Rogers to meet the needs of the psychologist involved in a forensic practice within the criminal court system. As professionals in forensic practice realize, the construct of legal insanity is, at best, poorly linked to the knowledge we have of human behavior. The legal profession has developed at least four different “standards” of legal insanity: the McNaghten standard, the Irresistible Impulse standard, the American Law Institute standard, and the Guilty But Mentally Ill standard. The R-CRAS was designed by using gradations of severity to standardize the evaluation of clinical information, such that the ambiguity of the decision on criminal responsibility is minimized.

**Scoring:** The five scales produced by the R-CRAS are: patient reliability, organicity, psychopathology, cognitive control, and behavioral control. Each of 30 items are scaled in gradation against the anchor of increasing severity (e.g., 0 for no information; 1 for not present or not applicable; 2 for clinically insignificant; and 3 through 6 for increasing levels of severity).

**Reliability and Validity:** A cross-validation study based upon 111 client-defendants estimated the kappa reliabilities for each of the five R-CRAS subscales; the results ranged from .68 and .63 (original sample and cross-validation sample, respectively) to 1.00. Interrater reliability of decisions with regard to insanity were .93 for the original sample and 1.00 for the cross-validation sample. This research also included a discriminant analysis in terms of clients evaluated as sane and insane against the five subscales. Rogers et al. reports a 72% correct classification of sane and 99% classification of insane clients.

**Norms:** The standardized sample included “sane” and “insane” client-defendants.

**Suggested Uses:** The R-CRAS is recommended for use as an assessment instrument in a forensic setting.
Rorschach Inkblot Test

**Purpose:** Designed as a projective technique to assess personality characteristics.

**Population:** Children and adults.

**Score:** 4-square (Exner).

**Time:** N/A.

**Author:** Hermann Rorschach.

**Publisher:** U.S. Distributor--Grune & Stratton, Inc.

**Description:** Hermann Rorschach, employing inkblots (i.e., using forms obtained through chance by folding over a piece of paper into the center of which ink had been dropped) to explore an aspect of personality, created the Rorschach. He standardized the inkblot procedure and synthesized the procedure with Jung’s work on the Word Association Test and Bleuler’s notions regarding personality assessment. When responses to inkblots had been used to assess imagination, emphasis was placed on the content of the responses. By contrast, Rorschach stressed not the content, but rather the formal properties of the response, and, as such, this enabled him to conceptualize the test as one of perception and not of imagination. By noting the intimate relationship between perceptual reactions and other psychological functions, Rorschach was able to conceptually place his technique in the middle of the assessment of total personality functioning.

**Scoring:** The foundation for Rorschach interpretation based upon the Comprehensive System is what Exner terms the “4-square,” which incorporates the basic scores and ratios thought to be characteristic of one’s problem-solving style. The four indices of the 4-square are 1) Erlebnistypus (EB, the ratio of human movement to weighted color responses); 2) Experience Actual (EA, the sum of human movement and weighted color responses); 3) Experience Base (eb, the ratio of nonhuman movement to shading and gray-black responses); and 4) Experience Potential (ep the sum of non-human movement, shading, and gray-black responses).

**Reliability and Validity:** Refer to Exner’s Comprehensive System publications for detailed information on the reliability and validity of this scoring system. There is also abundant literature on several other alternate scoring systems for the Rorschach.

**Norms:** N/A.

**Suggested Uses:** The Rorschach is recommended for projective personality assessment in clinical and research settings.
Rotter Incomplete Sentences Blank, Second Edition

**Purpose:** Designed “as a screening instrument of overall adjustment.”

**Population:** College students, adults, high-school students.

**Score:** Index of Overall Adjustment.

**Time:** (20-40) minutes.

**Authors:** Julian Rotter, Michael Lah, and Janet Rafferty.

**Publisher:** The Psychological Corporation.

**Description:** The Second Edition of the Rotter Incomplete Sentences Blank (RISB) is a projective measure of maladjustment with a semi-objective scoring system. This revised instrument provides direct information on personality conflicts. As scoring depends on intuitive clinical insights, cognizance of personality dynamics is essential for accurate interpretation. Although responses can also be scored qualitatively for projected motivational needs, as a general rule, interpretation of subjective scales is notoriously unreliable.

**Scoring:** Responses are rated on a 7-point ordinal scale (higher scores suggesting greater maladjustment) on the basis of omissions and incomplete responses, conflict responses, positive responses, or neutral responses. Overall scores generally range from 80 to 205 (on a scale from zero through 240). However, because of the diversity among individuals’ idiosyncratic responses, the RISB cannot readily be computer scored. This inevitably raises questions as to the objectivity and scoring consistency of RISB responses.

**Reliability:** Several studies reported in the RISB manual suggest uncertain reliability. Stability coefficients are reported as ranging from a low .38 (retest interval of 3 years) up to .82 (retest after only 1-2 weeks), so it cannot be assumed that the instrument is always reliable. Some inconsistency in scoring is evident because although interscorer reliabilities range from .72 to .99, item-rater reliabilities range from .44 to .93. Because the scoring examples are based solely on college samples, there is the further question as to their applicability for high school and adult groups. A positive feature of this test is its item homogeneity. The manual reports split-half estimates ranging form .74 to .86, and a Cronbach alpha coefficient of .69. This moderate level of item homogeneity suggests little item redundancy, and yet sufficient internal consistency to justify its use.

**Validity:** The RISB relies predominantly on face validity, so that item responses are readily amenable to distortion, depending on respondents’ lack of self-insight, and their conscious and unconscious motives. The instrument may act more effectively as a trait measure than as a state indicator of changes over time, and not be situationally sensitive. Despite these limitations, the RISB has received widespread use in both clinical and nonclinical settings.

**Norms:** College students.

**Suggested Uses:** Recommended uses of the RISB include screening, tracking changes in scores over time, group comparisons, and research.
The Short Employment Tests

**Purpose:** Designed to test applicants for clerical positions.

**Population:** Applicants for clerical positions.

**Scores:** Three scores.

**Time:** (15) minutes.

**Authors:** George Bennett and Marjorie Gelink.

**Publisher:** The Psychological Corporation.

**Description:** The Short Employment Tests (SET) is a group of 3 five-minute tests for applicants to clerical positions. According to the authors, SET is “not intended for use with candidates for administrative positions, nor with maintenance employees or factory production workers.”

**Scoring:** The battery consists of a Verbal (V) test of vocabulary and word knowledge, a 50 item test of general vocabulary; a Numerical (N) test of simple mathematical computations, a 90 item test of the four arithmetic operations; and a Clerical Aptitude (CA) test involving locating a name in a list and classifying a dollar amount next to that name. The tests are short and quickly administered and scored.

**Reliability:** The manual does not include reliability data on total scores, although subsequent reviewers have established these to be quite high.

**Validity:** The overall trend in the validity studies is that the total validity coefficients (combining all three tests) are higher than individual test validities. This emphasizes the importance of administering all three tests as opposed to administering only one or two of them. The highest correlations with similar sections of other tests are in the CA test, where all of the correlations are significant at the .01 level.

**Norms:** Normative data in the 1978 supplement to the original manual are frequently presented in applicant tables involving both sexes. This may result in test administrator confusion when referring to normative data.

**Suggested Uses:** The SET appears to be a usable employment device for the specific area of clerical applicants.
Sixteen Personality Factor Questionnaire

Purpose: Designed as an objective personality test.

Population: Ages 16 and above.

Score: Sten scores.

Time: 30-60 minutes.

Author: Raymond B. Cattell

Publisher: Institute for Personality and Ability Testing, Inc.

Description: The Sixteen Personality Factor Questionnaire (16PF) is an objective test of 16 multidimensional personality attributes arranged in omnibus form. In general, it provides normed references to each of these attributes (the primary scales). Conceptualized and initially developed by Raymond B. Cattell in 1949 as a broad, multipurpose measure of the "source traits" of individual personality, the 16PF is appropriate for a wide range of multifaceted populations. It provides a global representation of an individual’s coping style, the person’s reactive stance to an ever-fluid and transactional environment and that individual’s ability to perceive accurately certain specific environmental requisites for personal behavior.

Scoring: A subject’s raw score for each of the 16 primary factors is obtained through a weighted procedure where particular responses count as “1” or “2” summatively toward the final raw score. These weighted or unweighted sums are then compared to the desired normative score tables in the tabular supplement where a particular sten score is identified based on the magnitudinal range of the response and the individual normative demographics of the respondent. This sten score is entered on the profile form and subsequently depicted graphically for ease of interpretation.

Reliability: Reliability coefficients calculated by test-retest with short intervals (single or multiple day) demonstrate relatively acceptable coefficients, with only sporadic instances of a scale falling below a .70 magnitude. For stability coefficients, test-retest administrations conducted over long intervals (several weeks), magnitudes are expectedly reduced. Intercorrelations between primary factor scales generated from different test forms are seldom greater than .50 when Forms A and B are compared. Fewer coefficients of .50 or more magnitude exist for Forms C and D.

Validity: Forms A and B are reported to have the greatest total direct validity where each form has seven scales with validity coefficients of at least .70 magnitude. Indirect construct validities for Forms A, B, C, and D are also reported in the form of multiple correlation coefficients, representing the degree of relationship between each primary scale magnitude and the total remaining primary scale magnitudes in the 16PF. As might be anticipated, correlational coefficients fall below a .80 magnitude in only two instances: .63 for Shrewdness and .74 for Imagination.

Norms: The norms were constructed for high-school juniors and seniors, college students, and a general nation-wide population of age and income levels commensurate with the then current U.S. Bureau of Census figures.

Suggested Uses: The 16PF is recommended for use in personality assessment as part of a battery in clinical and research settings.
**Slosson Drawing Coordination Test**

**Purpose:** Designed to identify individuals with brain dysfunction or perceptual disorders involving hand-eye coordination.

**Population:** Children and Adults

**Scores:** Raw scores, Accuracy Score

**Time:** (10-15) minutes

**Author:** Richard L. Slosson

**Publisher:** Slosson Educational Publications

**Description:** The Slosson Drawing Coordination Test (SDCT) is a supplement to the Slosson Intelligence Test (SIT). It can be given individually or in groups.

**Scoring:** Each drawing is scored either “plus” when all lines are contained in the drawing, and or “minus” when the elements are elements are distorted. The raw score is the total number of “minus” drawings. The Accuracy score is a percentage score for errors at a certain age, and is obtained from a table in the manual.

**Reliability:** The reliability coefficient was .96 for test-retest reliability on a sample ranging from 4 to 52 years, tested at the beginning and the end of the same session. Inter-scorer reliability is high, as scoring rules have been simplified.

**Validity:** The goal of this test is to screen our individuals with serious forms of brain damage. Individuals with emotional or mental disturbances, those lacking motivation, or those with vision impairments may also show an abnormal degree of distortion on this test. There are also brain dysfunctions that do not involve eye-hand coordination, so this SDCT should be used in coordination with the SIT to strengths and weaknesses may be determined from the analysis of scatter. Qualitative validation research involved testing individuals who were known to be brain damaged.

**Norms:** Norms have been established for each age level and for each figure. Norms have been established for groups reflecting the individual's gender, age, dexterity, artistic ability, vision defects and disorganizing emotional or mental illness.

**Suggested Use:** The SDT can be used in conjunction with the Slosson Intelligence Test. It can be used to identify individuals with various forms of brain dysfunction of perceptual disorders where eye-hand coordination is involved.
**Slosson Intelligence Test**

**Purpose:** Designed for use as a “quick estimate of general verbal cognitive ability.”

**Population:** Ages 4–0 and over.

**Scores:** Total score only.

**Time:** (10–20) minutes.

**Authors:** Richard L. Slosson, Charles L. Nicholson (revision), and Terry H. Hibpshman (revision).

**Publisher:** Slosson Educational Publications, Inc.

**Description:** The manual states that the purpose of the Slosson Intelligence Test (SIT) is to serve as a “quick estimate of general verbal cognitive ability” or “index of verbal intelligence.” Although reviewers have described it as a brief screening measure of verbal crystallized intelligence, the manual presents appropriate cautions about interpretation of the SIT as a screening measure, suggesting at several points that follow-up assessment is necessary to corroborate SIT results.

**Scoring:** The SIT is easy to administer and score. The test contains 187 untimed items assessing the cognitive domains of vocabulary, general information, similarities and differences, comprehension, quantitative ability, and auditory memory. All the items are presented in question and answer format. The raw score is computed by adding the highest item in the basal to the number of correct responses after the basal.

**Reliability:** Kuder-Richardson 20 reliability coefficients by age level range from .88 to .97, indicating a high degree of inter-item consistency. Test-retest reliability is reported to be .96, based on a weak sample size of 41 and a one-week administration interval. Split-half reliability, calculated using the Spearman-Brown correction and the Rulon procedure, was .97 for the entire sample.

**Validity:** Concurrent criterion-related validity is based on correlations between the SIT total standard score and the WAIS-R and the WISC-R IQS. In a study of 10 subjects, significant correlations were found between TSS and three IQ scores on the WAIS-R. Comparisons with the WISC-R were made utilizing 234 subjects between the ages of 6 and 16. At each of four age levels the TSS correlated significantly with each of the WISC-R IQS.

**Norms:** The sample approximates the percentages found in the United States in terms of geographic region, occupational category, educational level, gender, and race. Minorities are under represented in the standardization sample as are those living in areas with populations below 5,000 and above 500,000.

**Suggested Uses:** It is recommended that the SIT is suitable for screening purposes, although reviewers recommend using the short form of other comprehensive intelligence test batteries that are more psychometrically sound.
Slosson Oral Reading Test

**Purpose:** Designed as a “quick estimate to target word recognition levels for children and adults.”

**Population:** Preschool-Adult.

**Scores:** Total score only.

**Time:** (3-5) minutes.

**Authors:** Richard L. Slosson and Charles L. Nicholson.

**Publisher:** Slosson Educational Publications, Inc.

**Description:** The Slosson Oral Reading Test (SORT) is designed to assess a subject’s “level of oral word recognition, word calling or reading level.” This instrument is not a a diagnostic measure nor does it measure all aspects of reading such as word knowledge and comprehension. It is a “quick screening test to determine a student’s reading level.”

**Scoring:** A basal level is attained when a subject can pronounce all 20 words in a group. A ceiling is reached when none of the 20 words in a group can be pronounced correctly. Basic administration and scoring procedures are printed on each test protocol. Raw scores, grade and age equivalents, percentile rank, standard scores, and confidence levels can also be determined and recorded directly on the protocol.

**Reliability:** Internal consistency and test-retest stability all yield coefficients above .95. The high reliability can be attributed to the SORT containing a large number of items, a good sampling of test items, and test specificity as only oral reading is being measured.

**Validity:** The SORT has been administered concurrently with several tests of reading recognition and reading comprehension. Passage Comprehension from the Woodcock-Johnson Test of Achievement and Reading Comprehension from the Peabody Individual Achievement Test correlate with the SORT .68 and .83, respectively. The SORT is also correlated highly (.87) with the Slosson Intelligence Test.

**Norms:** The SORT was co-normed with the Slosson Intelligence Test. The sample included 1,331 subjects, ranging from preschool to adults. Although the author suggest that the sample represents the U.S. population, close inspection reveals that large differences exist between sample and census data on geographic location and occupational status. There were also large differences in the number of subjects at different age levels.

**Suggested Uses:** Although the primary use of the SORT is as a screening instrument, the author also suggests the SORT may be used to assess a student’s progress, determine a student’s grade level in reading, and to determine if a student is in need of further diagnostic assessment.
**Stanford-Binet Intelligence Scale: Form L-M, Third Edition**

**Purpose:** Designed to assess cognitive ability.

**Population:** Ages 2 to adult.

**Score:** Scaled scores, area scores, and a Composite Score.

**Time:** (30-90) minutes.

**Authors:** Lewis M. Terman and Maud A. Merrill.

**Publisher:** The Riverside Publishing Co.

**Description:** The third edition of the Stanford-Binet Intelligence Scale by Lewis M. Terman and Maud A. Merrill was published in 1972. It is a wide-range individual test, assessing intelligence from age two through the superior adult level. It is an age scale, requiring subjects to solve problems, give definitions, memorize new material, and use some visual-motor skills at various age levels.

**Scoring:** This test has twelve subtests: Vocabulary, Comprehension, Absurdities, Verbal Relations, Pattern Analysis, Copying, Matrices, Paper Folding and Cutting, Quantitative, Bead Memory, Memory for Sentences, Memory for Digits. The examiner must establish a basal age of the child and continue testing until a ceiling is reached. Interpretation is based primarily on objective scoring.

**Reliability:** Reliability of the scale varies at different ages and different IQ ranges. From age 2.5 to 5.5, the reliability coefficients range from .83 for IQS 140-149 to .91 for IQS 60-69. For ages 6 to 13, the coefficients are .91 to .79 respectively, and for ages 14-18 the coefficients range from .95 to .98 respectively.

**Validity:** Validity of the scale depends on three sources: 1) the choice of items according to mental age on the 1937 scale assures that the new scale is measuring the same thing that was measured in the original 1916 scale; 2) regular increases in mental age from one age to another agreed with increases in percent passing from one chronological age to the next in both forms of the 1937 scale; and 3) biserial correlations were computed for each item of Forms L and M of the 1937 scale. The retention of an item for the 1960 scale partly depended on its correlation with the total score. The mean biserial correlation for the 1960 scale is .66 (range from .61 at young age levels to .73 for the adult levels).

**Norms:** The standardization group consisted of a representative sample of 2,100 children, with approximately 100 subjects at each Stanford-Binet year level. Unlike the 1960 norms, which did not include nonwhites in the standardization group, the 1972 norms contained nonwhites (including black and Spanish-surnamed individuals) and whites. Subjects were, however, excluded from the normative sample if English was not the primary language spoken in the home.

**Suggested Uses:** Used for assessment in educational, research, and clinical settings.

**Purpose:** Designed for the testing of cognitive abilities.

**Population:** Ages 2 to 23.

**Score:** Scaled scores, area scores, and a Composite Score.

**Time:** (30-90) minutes.

**Authors:** R.L. Thorndike, E.P. Hagen, and J.M. Sattler.

**Publisher:** Riverside Publishing Co.

**Description:** A three-level hierarchical model was used to guide the construction of the Stanford-Binet Intelligence Scale: Fourth Edition (SB: FE). The model postulates (a) g (a general intelligence factor) at the highest level of interpretation; (b) crystallized, fluid, and short-term memory factors at the second level; and (c) more specific factors such as verbal reasoning, quantitative reasoning, and abstract visual reasoning— at the third level. The Composite Score reflects the highest level and is considered to be the best estimate of g in the scale.

**Scoring:** Raw scores are converted into three types of standard scores: standard age scores (or scaled scores) for the subtests (M=50, SD=8), area scores (M=100, SD=16), and a Composite Score (M=100, SD=16). The fifteen subtests include: Vocabulary, comprehension, Absurdities, Verbal Relations, Pattern Analysis, Copying, Matrices, Paper Folding and Cutting, Quantitative, Number Series, Equation Building, Bead Memory, Memory for Sentences, Memory for Digits, and Memory for Objects.

**Reliability:** The Composite Score of the SB: FE has excellent reliability. Internal consistency reliabilities for the Composite Score range from .95 to .99 over the 17 age groups. The median Composite Score reliability is .97. The median subtest reliabilities range from a low of .73 for Memory for Objects to a high of .94 for Paper Folding and Cutting. Subtest reliabilities differ somewhat according to age group, ranging from a low of .66 for Memory for Objects at age 10 years to a high of .96 for Pattern Analysis at ages 18-23.

**Validity:** The Technical Manual for the SB: FE presents several studies investigating the Scale’s criterion validity. Comparisons were made with various other tests, including the Stanford-Binet: Form L-M, WISC-R, WPPSI, WAIS-R, and K-ABC, for both normal and exceptional populations. In the 13 studies reported in the manual, correlations between the SB: FE and these criterion measures ranged from a low of .27 to a high of .91 (Mdn r = .80). The median r of .80 supports the concurrent validity of the SB: FE.

**Norms:** The standardization sample consisted of 5,013 individuals in 17 age groups. The number of individuals ranged from 194 in the 18-0 to 23-11 age group to 460 in the 5-0 to 5-11 group. The sample was selected so as to be representative of the U.S. population according to 1980 census data. Stratification variables included geographic region, community size, ethnic group, age, gender, and socioeconomic status.

**Suggested Uses:** Designed for use in an educational, clinical, or research setting.
Stanford-Binet Intelligence Scale: 5th Edition (SB5)

**Purpose:** Designed for the testing of cognitive abilities.

**Population:** Individuals ages 2 through 85+ years.

**Score:** Verbal, Performance, and Full Scale Scores.

**Time:** Varies depending on age and functional level of examinee. Most items are untimed adding to the overall time for administration.

**Author:** Gale H. Roid.

**Publisher:** Riverside Publishing.

**Description:** This fifth edition blends features of earlier editions of the SB with recent improvements in psychometric design. Point-scale format subtests, designed to measure behavior at every age, and used in the 1986 edition are combined with the age-scale or functional-level design of the earlier editions (1916-1937). Two routing subtests identify the developmental starting points of the examinee, and the items can be tailored to cognitive level, resulting in greater precision in measurement. The fifth edition maintains many of the same subtests and items of previous editions, and includes a global g factor and several broad factors at the second level as in the Stanford-Binet Fourth Edition (SB: FE). Changes in the SB5 distinguishing it from the SB: FE, include some modernization of both artwork and content and some additions and enhancement of content. The SB5 now has five factors, (Fluid Reasoning, Knowledge, Quantitative Reasoning, Visual-Spatial Processing, and Working Memory) as opposed to the four of the SB: FE. Many toys and objects have been returned from earlier editions, as they are helpful with early childhood assessment. Unique to the SB5 is the use of a nonverbal mode of testing covering all five cognitive factors. The range of the scales has been extended to more accurately measure both higher and lower areas of functioning. Changes in the Item and Record Forms are reported to have made the scale more useful for clinical, forensic, school and vocational applications and interpretations. This edition also allows for evaluation of the abilities of elderly examinees.

**Scoring:** Raw scores are converted into scaled scores (M = 10, SD = 3) using age appropriate tables. The scaled scores are summed for Nonverbal, Verbal and Full Scale IQ as well as for the five factor index scores. These are all normalized standard scores (M = 100, SD = 15). Percentile rank equivalent and the confidence intervals are also obtained. Computerized scoring is faster, provides greater consistency of raw score conversion and is recommended for use whenever available.

**Reliability:** Using the split half method, and correcting with the Spearman-Brown formula, reliability coefficients were extremely high for the Full Scale Score (.98). The Nonverbal (.95) and Verbal (.96), showed excellent stability, and the Abbreviated Battery (.91) is also considered excellent as it contains only two subtests. The five factor index scores were all above 90, and were higher than the subtest scales, which were however comparable to other cognitive tests with ranges from .84 to .89.

**Validity:** As with the SB: FE, several studies were done to investigate the validity of the SB5. The first such studies explored the relationship between the SB5 and the SB: FE and the Form L-M. There was a high correlation (.90) between the SB5 Full Scale and the SB: FE Composite Scale. The higher scores of the Composite Scale are explained by the differences in the SD used and the countrywide changes in the IQ (0.3 points per year, as documented by Flynn (1985, 1987). The difference in the Full Scale Score of the SB5 and the one score of the Form L-M was in the direction, but not as great as predicted. Differences in scoring and nonverbal aspects of the tests, as the Full Scale and Verbal Score correlations are high (.85 and .88). Comparisons were also made with other tests such as the WPPSI-R, the WISC-III, the WAIS-III, and the Woodcock-Johnson III test of Cognitive Ability and W-J III Test of Achievement. Correlations ranged from .78 to .84. for Full Scale or Verbal IQ and comparable scores. This extensive analysis revealed a high correlation between the composite IQ scores of the SB5 and the composite scores of previous SB edition and all of the major IQ batteries used for all populations.

**Norms:** The standardization consisted of 4,800 subjects ages 2 to 85+. Care was taken to assure that the sample was as representative of the US population Census 2001. Stratification variables were age, sex, race / ethnicity, geographic region, and socioeconomic level.

**Suggested Use:** The SB5 are used in diagnosis of mental retardation, learning disabilities, and developmental cognitive delays in young children and for placement in academic programs for the intellectually gifted.
Stanford Diagnostic Mathematics Test (Red and Green Levels)

**Purpose:** Designed to identify specific mathematical concepts in which a student is lacking.

**Population:** Grades 1.5 to 6.5.

**Score:** Four scores.

**Time:** (95-110) minutes.

**Authors:** Leslie S. Beatty, Richard Madden, Eric F. Gardner, and Bjorn Karlsen.

**Publisher:** The Psychological Corporation.

**Description:** The Stanford Diagnostic Mathematics Test (SDMT) was designed to identify those specific mathematical concepts and skills on which a student is making less than satisfactory progress. Covering the mathematical content of grades 1 through 8, the test is divided into four levels, (two grades per level) and three areas: Number System and Numeration, Computation, and Applications. It may be administered to single individuals or to groups, and permits both norm-referenced and content-referenced interpretations.

**Scoring:** The three subtests produce raw scores that can be converted into percentile ranks, stanines, grade equivalents, and scaled scores for each of the three subtests and for the total score.

**Reliability:** Reliability estimates for the subtests are presented as KR20 coefficients and alternate-form coefficients. The reliability values are acceptably high, but most users will probably be more interested in focusing on the Domain and Cluster scores for diagnostic purposes. The manual indicates that the Domain scores may be interpreted if considerable caution is used; however, few Cluster scores can be interpreted with an acceptable degree of confidence, since not many of the reliability values even reach .70.

**Validity:** Evidence for criterion-related validity consists of correlations between the subtests and the corresponding Mathematics Tests of the Stanford Achievement Test, all of which exceed .66. These values constitute evidence of reasonable validity for the subtests.

**Norms:** The national standardization sample was stratified on socioeconomic status, district enrollment, and geographical area. A comparison of the sampled districts with the total U.S. public schools on geographical area and district enrollment shows favorable agreement, but no SES data are presented in the manuals.

**Suggested Uses:** Recommended for use in assessing mathematical achievement in educational and research settings.
State-Trait Anxiety Inventory

**Purpose:** Designed to study anxiety.

**Population:** Grades 9-16 and adults.

**Score:** 2 scores: state anxiety and trait anxiety.

**Time:** (10-20) minutes.

**Author:** Charles D. Spielberger, Richard L. Gorusch, and Robert E. Lushene.

**Publisher:** Consulting Psychologists Press, Inc.

**Description:** The State-Trait Anxiety Inventory (STAI) was initially conceptualized as a research instrument for the study of anxiety in adults. It is a self-report assessment device which includes separate measures of state and trait anxiety. According to the author, state anxiety reflects a "transitory emotional state or condition of the human organism that is characterized by subjective, consciously perceived feelings of tension and apprehension, and heightened autonomic nervous system activity." State anxiety may fluctuate over time and can vary in intensity. In contrast, trait anxiety denotes "relatively stable individual differences in anxiety proneness . . ." and refers to a general tendency to respond with anxiety to perceived threats in the environment.

**Scoring and Norms:** Scores on the STAI have a direct interpretation: high scores on their respective scales mean more trait or state anxiety and low scores mean less. Both percentile ranks and standard (T) scores are available for male and female working adults in three age groups (19-39, 40-49, 50-69), male and female high school and college students, male military recruits, male neuropsychiatric patients, male medical patients, and male prison inmates.

**Reliability:** The stability of the STAI scales was assessed on male and female samples of high school and college students for test-retest intervals ranging from one hour to 104 days. The magnitude of the reliability coefficients decreased as a function of interval length. For the Trait-anxiety scale the coefficients ranged from .65 to .86, whereas the range for the State-anxiety scale was .16 to .62. This low level of stability for the State-anxiety scale is expected since responses to the items on this scale are thought to reflect the influence of whatever transient situational factors exist at the time of testing.

**Validity:** Correlations are presented in the manual between this scale and other measures of trait-anxiety: the Taylor Manifest Anxiety Scale, the IPAT Anxiety Scale, and the Multiple Affect Adjective Check List. These correlations are .80, .75, and .52, respectively.

**Suggested Uses:** Recommended for studying anxiety in research and clinical settings.
Study of Values

Purpose: Designed to measure six basic personality-related motives or interests.

Population: Grades 10-16 and adults.

Score: 6 scores.

Time: (20) minutes.

Authors: Gordon W. Allport, Philip E. Vernon, and Gardner Lindzey

Publisher: Riverside Publishing Co.

Description: The Study of Values attempts to measure the relative strength of six basic, personality-related motives or interests: theoretical, economic, aesthetic, social, political, and religious. The basis for this classification was presented by Eduard Spranger in Types of Men. In this work Spranger developed the position that the personalities of human beings can best be understood by studying their values.

Scoring: The test generates six scores in these domains: theoretical, economic, aesthetic, social, political, and religious.

Reliability: Corrected Spearman-Brown formula, split-half reliabilities ranged from a low of .84 (Theoretical) to high of .95 (Religious). With regard to test-retest reliability, the manual presents two studies. In one, carried out in 1951 with 34 subjects and a one-month interval between administrations, the correlation coefficients ranged from .77 (Social) to .12 (Economic). In the second, carried out in 1957 with 53 subjects and a two-month interval between administrations, the correlation coefficients ranged from .84 (Economic) to .93 (Religious).

Validity: With regard to validity, the evidence presented in the manual is based primarily on group differences. For example, that college students majoring in business administration will have relatively high Economic scale scores while theology students will have relatively high Religious scale scores are expectations that the manual’s empirical data support. Males and females have characteristically different value profiles. Value scores have been correlated with measures of creativity, church attendance, descriptions of self, and numerous other characteristics.

Norms: The normative sample consisted of a heterogenous sample 1,816 college students.

Suggested Uses: The Study of Values test is recommended primarily for research settings.
Symptom Checklist-90-Revised

**Purpose:** Designed to reflect psychological symptom patterns.

**Population:** Adults 13 years and older with a sixth grade reading level.

**Scales:** Somatization (SOM), Obsessive-Compulsive (O-C), Interpersonal Sensitivity (I-S), Depression (DEP), Anxiety (ANX), Hostility (HOS), Phobic Anxiety (PHOB), Paranoid Ideation (PAR), Psychoticism (PSY), and Additional Items, the Global Severity Index (GSI), The Positive Symptom Distress Index (PSDI), and the Positive Symptom Total (PST).

**Time:** (12-15) minutes for most individuals, (2-5) minutes for instructions.

**Author:** Leonard R. Derogatis

**Publisher:** National Computer Systems, Inc.

**Description:** The Symptom Checklist-90-Revised (SCL-90-R) is a 90-item self-report symptom inventory. It is a measure of current psychological symptom status with a time reference of “the past 7 days including today.” The SCL-90-R replaces some of the items on the Anxiety and the Obsessive-Compulsive dimensions of the SCL-90 that were flawed psychometrically. Ambiguities in the test instructions and in the definition of the distress continuum were also modified.

**Scoring:** Scores for each of the nine factors are the average rating given to the symptoms of that factor. The remaining seven items do not measure any particular factor, but are evaluated qualitatively. Three “global” scores are also obtained. The GSI is the average rating given to all 90 items. The PST is the number of symptoms complained of (i.e., the number of items rated higher than zero). The PSDI is the average rating, from 1 to 4, given to those symptoms which are complained of (i.e., not rated “0”). Raw scores for each of the primary symptoms are converted into standardized scores. There are three global indices; the Global Severity Index (GSI), The Positive Symptom Distress Index, (PSDI), and the Positive Symptom Total (PST).

**Reliability:** The internal consistency coefficient alphas for the nine symptom dimensions ranged from .77 for Psychoticism, to a high of .90 for Depression. Test-retest reliability coefficients range between .80 and .90 after one week of therapy.

**Validity:** The few validity studies of the SCL-90-R demonstrate levels of concurrent, convergent, discriminant, and construct validity comparable to other self-report inventories. Results of studies in which the SCL-90-R was used as a change measure (e.g., studies of pharmacological effects, stress) provide support for the SCL-90-R, most frequently for the three global scores, and with more scattered findings regarding one or another of the nine dimensions.

**Norms:** Gender specific norms were done for four groups: Norm A (1002 adult psychiatric outpatients), Norm B (974 adult non-patients), Norm C (423 adult psychiatric inpatients), and Norm E (806 adolescent non-patients).

**Suggested Use:** The SCL-90-R serves as an outcome measure to assess treatment effectiveness, measures of change with treatment, psychopharmacology outcomes and for research studies.
**System of Multicultural Pluralistic Assessment**

**Purpose:** Designed to assess the educational needs of children in a culturally non-discriminatory manner.

**Population:** Ages 5 to 11.

**Score:** Provides assessment in three areas: medical, social systems, and pluralistic systems.

**Time:** (4-5) hours.

**Authors:** Jane R. Mercer and June F. Lewis.

**Publisher:** The Psychological Corporation.

**Description:** The System of Multicultural Pluralistic Assessment (SOMPA) was designed to be responsive to the mandates of PL 94-142, that is, to assess the educational needs of children ages 5 to 11 years of age in a racially and culturally non-discriminatory manner. Its construction and recommended usage are derived from an explicit ideological view as to what American society should be like, i.e., culturally pluralistic. Thus, unlike most psychometric devices, the SOMPA authors openly express their ideological assumptions, e.g., intelligence and/or learning potential are distributed equally among all ethnic and racial groups, and hopes that its use will promote cultural pluralism.

**Scoring:** The medical perspective focuses on the presence or absence of organic pathology. Six measure are employed: Physical Dexterity Tasks (sensory motor coordination), The Bender Visual Motor Gestalt Test (perceptual and neurological factors), The Health History Inventories, Weight by Height norms (nutritional or developmental problems), Vision (the Snellen Test), and Auditory Acuity (national norms).

The social systems perspective deals with how the child’s behavior conforms to the social norms of the various social groups in which the child participates. Two instruments are used: the Adaptive Behavior Inventory for Children (ABIC) and the Wechsler Intelligence Scale for Children-Revised (WISC-R).

The third perspective, the pluralistic model, is unique to the SOMPA. Its procedures are designed to yield an index of the child’s intelligence or Estimated Learning Potential (ELF) through a “corrected” WISC-R score which is based on a comparison of how well the child performs on that test with other children who have had similar learning opportunities (i.e., from a similar socio-cultural background).

**Reliability and Validity:** The reliability and validity for the separate measures comprising the SOMPA appear to be satisfactory.

**Norms:** SOMPA measures were standardized on 2,100 California children ranging in age from five to eleven. The measures comprising the SOMPA were individually standardized upon creation. Consult their respective manuals for details.

**Suggested Uses:** The SOMPA is a comprehensive child assessment battery that is recommended for educational and research purposes.
Szondi Test

Purpose: Designed to assess genic relationship.

Population: Ages 4 and over.

Score: N/A.

Time: (10-15) minutes repeated six to ten times with one day intervals.

Author: Lipot Szondi.

Publisher: U.S. Distributor: Grune & Stratton, Inc.

Description: The Szondi Test is a projective technique based on a person’s reaction to a series of 48 photographs of psychotic patients. The photographs were chosen in accordance with the principle of genic relationship; that is, the person assumedly selects a photograph which portrays a psychiatric disorder also inherent in the subject’s own familial genealogy.

Scoring: This test yields eight factors and 4 vectors (each vector is a total of 2 factors). They are as follows; homosexual, sadistic, sexual vector, epileptic, hysterical, paroxysmal vector, catatonic, paranoid, schizophrenic vector, depressive, manic, and contact vector.

Reliability and Validity: The manual contains no information on reliability and validity.

Norms: The manual contains no information on normative samples.

Suggested Uses: Recommended uses include projective assessment in a clinical setting.
Tell-Me-a-Story

**Purpose:** Developed to identify strengths and deficits in cognitive, affective, intrapersonal and interpersonal functioning.

**Population:** ages 5 to 18

**Scales:** Quantitative Scales (Cognitive Functions, Personality Functions, Affective Functions), Qualitative Indicators (Affective Functions, Cognitive Functions)

**Time:** 45-60 minutes (short form); 120 minutes (long form)

**Author:** Giuseppe Constantino, Robert G. Malgady, and Lloyd H. Rogler

**Publisher:** Western Psychological Services

**Description:** The Tell-Me-a-Story (TEMAS) is a multicultural thematic apperception test designed for use with minority and non-minority children and adolescents with a set of stimulus cards and extensive normative data for each group. The stimulus card are structured to elicit specific responses, and are in color to facilitate verbalization and projection of emotional states. It differs from the TAT in that 1) it focuses on personality functions as manifested in internalized interpersonal relationships rather than on intrapsychic dynamics, 2) the TEMAS consists of 23 chromatic pictures while the TAT has 19 achromatic pictures and one bland card, 3) TEMAS attempts to elicit meaningful stories indicating conflict resolutions of bipolar personality functions, while the TAT uses ambiguous stimuli to elicit meaningful stories, 4) The TEMAS stimuli represent the polarities of negative and positive emotions cognitions and interpersonal functions, while the TAT is primarily weighted to represent negative emotions, depressive mood, and hostility. The TEMAS stimulus cards are culturally relevant, gender sensitive, and have diminished ambiguity.

**Scoring:** The children’s stories are scored in the record booklet using specific criteria representative of cognitive, affective and intrapersonal and interpersonal functioning. The scales and functions scored are Quantitative Scales (Cognitive Functions [Reaction Time, Total Time, Fluency, Total Omissions], Personality Functions [Interpersonal Relations, Aggression, Anxiety-Depression, Achievement Motivation, Delay of Gratification, Self-Concept, Sexual Identity, Moral Judgment, Reality Testing], Affective Functions [Happy, Sad, Angry, Fearful]), Qualitative Indicators (Affective Functions [Neutral, Ambivalent, Inappropriate Affect], Cognitive Functions [Conflict, Sequencing, Imagination, Relationships, Total Transformations, Inquiries, Omissions and Transformations scores for each of the following: Main Character, Secondary Character, Event, Setting]).

**Reliability:** Internal consistency reliability estimates were reported by the manual for the 34 functions (Cognitive, Personality, and Affective) suggested by the authors. The Cronbach’s alpha ranged from .31 to .98 with half below an acceptable level (.70), even for projective tests. Test-retest reliabilities were low where the highest function correlation was \( r = .53 \) and 26 of the 34 functions showed correlations near zero. Because the interrater reliabilities were typically moderate to high, the test-retest correlations suggest an extremely situational nature to this instrument. The functions proposed by the authors might vary with time, but should be more stable within individuals than illustrated by the test-retest data reported. One reviewer’s conclusion is that the TEMAS does not measure the constructs it proposes to measure, or does so with a large proportion of each score attributable to error. However, establishing reliability and subsequent validity has been a common problem for projective testing in general.

**Validity:** The TEMAS manual implies criterion-related validity using a rather complex regression analysis that concludes 6 to 22% of the variance of therapeutic outcomes is predicted using four standardized measures and four ratings of graduate students as criteria. The analysis reported is not sufficient evidence to conclude that correlational validity has been established. In independent research, the verbal responses of Hispanic children were greater to the TEMAS than to the TAT, however there is a lack of differential effectiveness of the minority cards. Additionally, greater verbal response does not necessarily indicate that the scores for Black or Hispanic students were more valid measures of personality than the scores that might have been obtained on the TAT. The TEMAS is based on a reasonable premise (i.e., existing projective techniques may be culturally insensitive) and the measure has promising construct related validity, but it has not yet established criterion-related validity and has yet to show that it is better than the TAT or similar measures at assessing the personality characteristics of minority children. The well-established TAT is still the suggested standard for thematic projective investigations. The reliability of the TEMAS is not acceptable when estimated with both internal
consistency and test-retest methods. The TEMAS is not an improvement over other projective tests and may have poorer psychometric properties than existing measures. The TEMAS cards may be useful for research purposes so direct comparisons to the TAT and other personality instruments are encouraged. Clinical or predictive uses of the TEMAS are still questionable and caution is recommended, even when minority children are the clients.

**Norms:** The normative sample included 2,000 individuals ages 11 to 85+ years, spanned the age range representative of the US population census for gender, socioeconomic status (examinee or parental education levels), and race or ethnic group.

**Suggested Use:** The TEMAS is of clinical use to better understand the cognitive, affective and personality functions, give problem-specific information in order to develop a more accurate treatment plan, to assess therapeutic progress and outcome.
Tennessee Self Concept Scale

**Purpose:** Designed to summarized an individual’s self-worth.

**Population:** Ages 12 and over.

**Score:** 2 scoring systems yield different types and amount of scores.

**Time:** (10-20) minutes.

**Author:** William H. Fitts.

**Publisher:** Western Psychological Services.

**Description:** The Tennessee Self Concept Scale (TSCS) consists of 100 self-descriptive items by means of which an individual portrays what he or she is, does, likes, and feels. The scale is intended to summarize an individual’s feeling of self-worth, the degree to which the self-image is realistic, and whether or not that self-image is a deviant one. As well as providing an overall assessment of self-esteem, the TSCS measures five external aspects of self-concept (moral-ethical, social, personal, physical, and family) and three internal aspects (identity, behavior, and self-satisfaction). In addition, crossing the internal and external dimensions results in the mapping of 15 “facets” of self-concept.

**Scoring:** The TSCS allows two scoring systems. The counseling form yields 14 profiled scores: self-criticism, 9 self-esteem scores (identity, self-satisfaction, behavior, physical self, moral-ethical self, personal self, family self, social self, and total), 3 variability of response scores (variation across the first 3 self-esteem scores, variation across the last 5 self-esteem scores, and total), and a distribution score. The clinical and research form yields 29 profiled scores: the 14 scores in the counseling form and the following 15: response bias, net conflict, total conflict, 6 empirical scales (defensive positive, general maladjustment, psychosis, personality disorder, neurosis, personality integration), deviant signs, 5 scores consisting of counts of each type of response made.

**Reliability:** The reliability estimates for all TSCS scales are retest coefficients based on a sample of 60 college students over a two-week period and in general range from .60 to .90.

**Validity:** The manual presents correlations between the TSCS scales and those of the MMPI, the Edwards Personal Preference Schedule, and several other well-known measures. A great many of these correlations are significant; so many in fact where the MMPI is concerned it appears the two inventories must be nearly completely overlapping. Researchers have reviewed a number of factor-analytic studies and concluded that none support Fitts’ specific hypothesis of 15 dimensions of self-concept.

**Norms:** The primary norm group for the TSCS was a sample of 626 people who varied in age from 12 to 68 years. The group was composed of approximately equal members of men and women and ranged over a variety of educational, social, and economic levels. There is a reasonable degree of racial and geographic diversity, but younger white subjects, especially students, are over represented.

**Suggested Uses:** Recommended uses for the TSCS include the assessment of self-concept in clinical, counseling, and research settings.
Test Anxiety Inventory

**Purpose:** Designed to measure test anxiety.

**Population:** High school and college students.

**Score:** Percentile ranking.

**Time:** (8-10) minutes.

**Author:** Charles D. Spielberger.

**Publisher:** Consulting Psychologists Press, Inc.

**Description:** The Test Anxiety Inventory (TAI) is a self-report inventory designed to measure test anxiety (TA) as a situation-specific personality trait. The TAI consists of 20 items or statements, and the respondents indicate on a four point Likert-type scale how often they experience the feeling described in each statement. The TAI provides a measure of total TA (TAI-T) as well as measures of two TA components—worry (W) and emotionality (E).

**Scoring:** The TAI is a brief instrument occupying one side of a page on which the 20 items are printed. The total TAI score (TAI-T) is based on all 20 items. Eight of the items measure the W component and 8 items measure the E. Four items that load on both subscales contribute to the TAI-T score but are not scored on either the W or E subscales. Percentile ranks are calculated from the raw scores.

**Reliability:** Test-retest reliabilities for TAI-T are reported for groups of high school, college, and graduate students over time periods ranging from two weeks to six months. Reliability was in the range of .80 to .81 for two-week to one-month periods with all groups. After six months, the reliability was .62 for a group of high school students. The alpha coefficients for TAI-T ranged from .92 to .96; for the subscales, alphas ranged from .83 to .91 for TAI-W, and from .85 to .91 for TAI-E.

**Validity:** The relationship between the TAI and its subscales with other anxiety measures (e.g., Sarason’s Test Anxiety Scale (TAS), Liebert & Morris’ Worry and Emotionality Questionnaire (WEQ), the STAI State and Trait Anxiety scales, and the STAI State Anxiety scale administered under examination stress conditions) all provide evidence of convergent validity. The correlation between the TAI-T score and the TAS was sufficiently high (.82 to .83) to suggest that the two scales measure essentially the same construct.

**Norms:** The normative sample consisted of high school and college students.

**Suggested Uses:** The TAI is recommended for use in research and clinical settings.
Test of Language Development

Purpose: Designed to assess language development in children.

Population: Ages 4-8.

Score: Five types of scores.

Time: (50-65) minutes.

Authors: Donald D. Hammill and Phyllis L. Newcomer.

Publisher: PRO-ED.

Description: The rationale for the Test of Language Development (TOLD) as reported by the authors serves four purposes: 1) to identify children who are significantly below their peers in language, 2) to determine children’s specific strengths and weaknesses, 3) to document children’s language progress as a consequence of special intervention programs, and 4) to serve as a measurement device in research involving language behavior.

Scoring: The TOLD’s seven subtests are: Picture Vocabulary, Oral Vocabulary, Grammatic Understanding, Sentence Imitation, Grammatic Completion, Word Discrimination, and Word Articulation. The TOLD yields five different types of scores: raw scores, language ages, percentiles, standard scores, and quotients for composite scores.

Reliability: Two studies report split-half coefficients greater than .80 for all subtests except Picture Vocabulary, where coefficients ranged between .61 and .72. Similarly, low coefficients were obtained for two subtests, Oral Vocabulary and Grammatic Understanding, at the five- and eight-year-old groups. Test-retest Reliability: For the TOLD, 21 children were tested twice over a period of five days. Pearson product-moment coefficients were computed on the raw scores, with coefficients being above the .80 level for all measures.

Validity: For criterion-related validity, the authors compared the TOLD test to existing criterion measures. The resulting correlation coefficients between the TOLD and several other measures were quite low. For example, the Grammatic Understanding subtest had a correlation coefficient of .13 at the four-year-old level and .47 at the eight-year-old level when compared to the Northwestern Syntax Screening Test’s Reception subtest.

Norms: The TOLD was standardized on 1,836 children, with 198 subjects being represented at the smallest age level. Subjects were selected from all regions of the United States and controlled for sex differences, residence, race, geographic area, an occupation of parents.

Suggested Uses: Recommended uses include language assessment in clinical, educational, and research settings.
Test of Language Development, Second Edition

Purpose: Designed to assess language development in children.

Population: Ages 4-12.

Score: Five types of scores.

Time: (50-65) minutes.

Authors: Donald D. Hammill and Phyllis L. Newcomer.

Publisher: PRO-ED.

Description: The rationale for the Test of Language Development (TOLD) as reported by the authors serves four purposes: 1) to identify children who are significantly below their peers in language, 2) to determine children’s specific strengths and weaknesses, 3) to document children’s language progress as a consequence of special intervention programs, and 4) to serve as a measurement device in research involving language behavior.

Scoring: The test is available in two editions: Primary (TOLD-P) and Intermediate (TOLD-I). The TOLD-P’s seven subtests are: Picture Vocabulary, Oral Vocabulary, Grammatic Understanding, Sentence Imitation, Grammatic Completion, Word Discrimination, and Word Articulation. The TOLD-I’s five subtests are: Sentence Combining, Characteristics, Word Ordering, Generals, and Grammatic Comprehension. Both the TOLD-P and TOLD-I yield five different types of scores: raw scores, language ages, percentiles, standard scores, and quotients for composite scores.

Reliability: On the TOLD-P, two studies report split-half coefficients greater than .80 for all subtests except Picture Vocabulary, where coefficients ranged between .61 and .72. Similarly, low coefficients were obtained for two subtests, Oral Vocabulary and Grammatic Understanding, at the five- and eight-year-old groups. For the TOLD-I, internal consistency coefficients were greater than .80. Test-retest Reliability: For the TOLD-P, 21 children were tested twice over a period of five days; on the TOLD-I, 30 children were tested after one week. In both cases, Pearson product-moment coefficients were computed on the raw scores, with coefficients being above the .80 level for all measures.

Validity: For criterion-related validity, the authors compared the two TOLD tests to existing criterion measures. The resulting correlation coefficients between the TOLD-I and the Test Of Adolescent Language were adequate. For the TOLD-P several of the coefficients were quite low. For example, the Grammatic Understanding subtest had a correlation coefficient of .13 at the four-year-old level and .47 at the eight-year-old level when compared to the Northwestern Syntax Screening Test’s Reception subtest.

Norms: The TOLD-P was standardized on 1,836 children, with 198 subjects being represented at the smallest age level, whereas the TOLD-I was standardized on 871 children, with 166 being the least number per age level. Subjects were selected from all regions of the United States and controlled for sex differences, residence, race, geographic area, an occupation of parents.

Suggested Uses: Recommended uses include language assessment in clinical, educational, and research settings.
Thematic Apperception Test

**Purpose:** Designed as a projective measure.

**Population:** Ages 4 and over.

**Score:** N/A.

**Time:** (100-200) minutes in 2 sessions one day apart.

**Author:** Henry A. Murray.

**Publisher:** Harvard University Press.

**Description:** The Thematic Apperception Test (TAT) is, along with the Rorschach, among the most widely used, re-searched, and taught projective tests in existence. It consists of a series of pictures of relatively ambiguous scenes to which subjects are requested to make up stories or fantasies concerning what is, has, and is going to happen, along with a description of the thoughts and feelings of the various characters depicted. The test protocol thus provides the examiner with a rich source of data, based on the subject’s perceptions and imagination, for use in the understanding of the subject’s current needs, motives, emotions, and conflicts, both conscious and unconscious. Its use in clinical assessment is generally part of a larger battery of tests and interview data.

**Scoring:** The data from the TAT can be scored according to a variety of existing quantitative systems. However, more commonly in clinical use the stories are interpreted in accord with general principles of inference derived from psycho-dynamic theory.

**Reliability and Validity:** The manual provides no information on reliability or validity, although the various scoring systems have independent psychometric data.

**Norms:** N/A.

**Suggested Uses:** The TAT is recommended as a projective method of personality assessment.
Vineland Adaptive Behavior Scales

Purpose: Designed to assess handicapped and non-handicapped persons in their personal and social functioning.

Population: Birth to 18 years and low functioning adults.

Score: 13 scores.

Time: 2 editions: Interview Edition (20-60) minutes, Interview Ed., Expanded Form (60-90) minutes.

Authors: Sara S. Sparrow, David A. Balla, and Dominick V. Cicchetti.

Publisher: American Guidance Service.

Description: The Vineland Adaptive Behavior Scales (VABS) were designed to assess handicapped and non-handicapped persons from birth to adulthood in their personal and social functioning. Following Edgar Doll’s original conceptualization of adaptive behavior as multidimensional in structure and his measurement of the behaviors by areas, the VABS is organized around four Behavior Domains: Communication, Daily Living Skills, Socialization, and Motor Skills.

Scoring: Standard score equivalents for domain raw scores and Adaptive Behavior Composite Standard scores were developed by the Angoff and Robertson procedure already familiar to users of the Kaufman Assessment Battery for Children. The manual also provides percentile ranks and stanines (for the domain and Composite scores), adaptive levels (by percentile groups), age equivalents (by raw score conversions) and maladaptive levels (for the Maladaptive Behavior domain).

Reliability: Split-half and test-retest reliability coefficients for the Composite scores are good, ranging from median values of .83 for the Motor Skills domain to .94 for the Composite. Interrater coefficients are lower for the same measures: .62 to .78. When broken down by subdomains, the coefficients fluctuate a great deal and some are quite low.

Validity: Selected standardization subgroups were compared on the original Vineland, the ABIC, the K-ABC, the PPVT-R, and the VABS. These concurrent measures exhibited low to moderate correlations, with generally higher coefficients obtained when the comparisons were made on subjects with handicapping conditions.

Norms: Standardization sampling followed 1980 census data and included 3,000 subjects from birth through 18-11 equally divided by sex.

Suggested Uses: The VABS is recommended for assessment in clinical and research settings.
Wahler Physical Symptoms Inventory

**Purpose:** Designed to measure the degree of physical or somatic complaints endorsed by an individual.

**Population:** Ages 16 and over.

**Score:** Raw and decile scores.

**Time:** Administration time not reported.

**Author:** H.J. Wahler.

**Publisher:** Western Psychological Services.

**Description:** The Wahler Physical Symptoms Inventory (WPSI) is an instrument designed to measure the degree of physical or somatic complaints endorsed by an individual. Wahler designed the inventory to specifically include those complaints considered to be exclusively somatic in composition, eliminating items of a psychological nature. H. J. Wahler developed the questionnaire in the late 1960s in an attempt to aid professionals in their delineation or differential diagnosis of physical and/or psychological problems.

**Scoring:** The examiner subtracts the number of omits and double ratings from the total number of items; then, all of the remaining ratings are summed and then divided by the number of items on which they are based (total number minus omitted and double-marked items). This final number constitutes the WPSI score. In order to make these scores more meaningful, raw scores can be converted into deciles, and a decile table is provided in the WPSI manual. The deciles and their corresponding raw scores are derived from normative data based on the psychiatric patient’s responses to the WPSI.

**Reliability:** Internal consistency is excellent, ranging from .85 to .94. Wahler also measured the reliability of the WPSI based on test-retest administrations. Because the instrument was constructed to measure the presence of somatic complaints at the time of testing, it is conceivable and highly likely that complaints would change over time, given the remission or acquisition of physical problems. It is not surprising, then, that reliability coefficients varied considerably with the passage of time (ranging from .45 to .94).

**Validity:** The author based concurrent validation on scores obtained from groups who would be “expected” to report and emphasize somatic symptomatology. Not surprisingly, those individuals with physical disabilities, psychiatric problems, and those who had applied for Social Security disability demonstrated greater complaining behavior when compared to the student samples. Wahler did obtain correlations between the WPSI and all subscales of the MMPI, most of which were significant at the ≤ .05 level. However, the strongest relationship was between the MMPI Hs and Hy subscales and the WPSI.

**Norms:** Norm groups included college, physically disabled, and people applying for Social Security compensation populations.

**Suggested Uses:** Recommended for use as a screening instrument in clinical and research settings.
Wechsler Memory Scale –3rd Edition (WMS-III)

**Purpose:** Designed to assess learning, memory, and working memory.

**Population:** Individuals in the age range of 16-89 years.

**Score:** Eight primary indexes and four supplemental auditory process composites.

**Time:** (30 - 35) minutes, 25-30 min. between tests, (15-20) min. for optional subtests.

**Author:** David Wechsler.

**Publisher:** The Psychological Corporation.

**Description:** This third edition updates the WMS-R and provides subtest and composite scores that assess memory and attention functions using both auditory and visual stimuli. There are now eight Primary Indexes (Auditory Immediate [was Verbal], Visual Immediate [was Visual], Immediate Memory [new], Auditory Delayed [new], Visual Delayed [new], Auditory Reception Delayed [new], General Memory [only delayed subtest scores], and Working Memory), which constitute Immediate Memory, General [Delayed] Memory, and Working Memory [was Attention/Concentration]. This edition retains the index score configuration of the WMS-R, but scale content, administration and scoring procedures have been changed. There is one slight change to the optional Information and Orientation subtest and one item has been deleted and six items added to the Mental Control subtest. The Figural Memory, Visual Paired Associated, and card B of Visual Reproduction subtests have been dropped. There are slight wording and administration changes in Logical Memory, and added subtests help to assess various aspects of visual memory. Requires strict and discreet timing for certain subtests.

**Scoring:** Requires verbatim recording of responses for subjective scoring on many sub-tests.

**Reliability:** The reliability coefficients for the WMS-III Primary subtests and Primary Indexes were on average found to be higher than for the WMS-R. Internal consistency reliability coefficients ranged for .70s to the .90s.

**Validity:** Correlation with the WMS-R was not direct because of the many changes in the scales. The Verbal Memory of the WMS-R had a .72 correlation coefficient with the Auditory Immediate, .68 with the Auditory Delayed, and .65 with General Memory of the WMS-III. The General Memory of the WMS-R and the Auditory Immediate .73, Auditory Delay .69, and general memory .67 of the WMS-III. As expected the correlations were lower for visually presented material with .34 for verbal memory and visual memory indexes. When correlated with the Children’s Memory Scale, the WMS-III auditory indexes correlated highest with the corresponding CMS indexes. Studies comparing the WMS-III and the WIAT show highest correlations between the WMS-III auditory indexes and working memory indexes and the WIAT subtests and composites, similar to results found with the CMS and the WIAT. In comparing the WMS-III with the WAIS-III there is a pattern of the auditory memory correlating more strongly with the VIQ and the visual memory measures correlating more strongly with the PIQ as an indication of convergent and divergent validity; while they are related, they measure different constructs.

**Norms:** The normative sample for the WMS-R included six age groups of 50 subjects each for a total sample of 300 aged 16 to 74 years. In contrast the WMS-III standardization sample, stratified to be representative of the general population included 1250 individuals aged 16-89 years.

**Suggested Use:** The WMS-III provides a more detailed analysis of memory function. When used in conjunction with the WAIS-III, meaningful comparisons between intellectual ability and memory functions can be made.
Wechsler Adult Intelligence Scale

**Purpose:** Designed as a general test of cognitive ability for adults.

**Population:** Ages 16 and above.

**Score:** Verbal IQ, Performance IQ, and Full Scale IQ.

**Time:** (90-120) minutes.

**Author:** David Wechsler.

**Publisher:** The Psychological Corporation.

**Description:** The Wechsler Adult Intelligence Scale (WAIS) is a general test of cognitive ability, which Wechsler defined as, “... the global capacity of the individual to act purposefully, to think rationally, and to deal effectively with his environment.” In keeping with this definition of intelligence as an aggregate of mental aptitudes or abilities, the WAIS consists of 11 subtests divided into two parts, verbal and performance.

**Scoring:** The WAIS consists of six verbal subtests and five performance subtests. The verbal tests are: Information, Comprehension, Arithmetic, Digit Span, Similarities, and Vocabulary. The Performance subtests are: Picture Arrangement, Picture Completion, Block Design, Object Assembly, and Digit Symbol. The scores derived from this test are a Verbal IQ (VIQ), a Performance IQ (PIQ), and a Full Scale IQ (FSIQ). The FSIQ is a standard score with a mean of 100 and a standard deviation of approximately 15.

**Reliability:** Split-half reliability coefficients (corrected) of .97 for FSIQ, .96 for VIQ, and .93 for PIQ are reported by Wechsler. Test-retest coefficients from .84 to .90 for intervals of 0 to 520 weeks are also reported. Intercorrelations of the subtests and correlations of each subtest with FSIQ indicate both uniqueness and communality for the subtests.

**Validity:** Predictive validity of the WAIS yields correlations of .60 in high school and a little lower for college. Factor analytic studies and confirmed good construct validity generally showing three to five factors, and usually two of these can be related to the VIQ and PIQ.

**Norms:** The standardization sample included 2,700 subjects intended to be representative of the general American population. The 1950 census was used to select a sample stratified by age, sex, occupation, geographical distribution, education, and urban-rural residence. About 10% of the sample were non-white.

**Suggested Uses:** The WAIS is recommended for use in clinical, educational, and research settings.
Wechsler Adult Intelligence Scale—Revised

**Purpose:** Designed as a comprehensive test of cognitive ability for adults.

**Population:** Ages 16 and over.

**Score:** Verbal IQ, Performance IQ, and Full Scale IQ.

**Time:** (60-90) minutes.

**Author:** David Wechsler.

**Publisher:** The Psychological Corporation.

**Description:** The Wechsler Adult Intelligence Scale—Revised (WAIS-R) is a general test of intelligence, which Wechsler defined as, "... the global capacity of the individual to act purposefully, to think rationally, and to deal effectively with his environment." In keeping with this definition of intelligence as an aggregate of mental aptitudes or abilities, the WAIS-R consists of 11 subtests divided into two parts, verbal and performance.

**Scoring:** The WAIS-R consists of six verbal subtests and five performance subtests. The verbal tests are: Information, Comprehension, Arithmetic, Digit Span, Similarities, and Vocabulary. The Performance subtests are: Picture Arrangement, Picture Completion, Block Design, Object Assembly, and Digit Symbol. The scores derived from this test are a Verbal IQ (VIQ), a Performance IQ (PIQ), and a Full Scale IQ (FSIQ). The FSIQ is a standard score with a mean of 100 and a standard deviation of approximately 15.

**Reliability:** Corrected split-half reliability coefficients for Verbal IQ (.95 to .97) and Full Scale IQ (.96 to .98), and their respective standard errors of about 2 points, are quite acceptable. The reliability of Performance IQ is excellent, averaging .93, although the value of .88 at ages 16 to 17 is not ideal. Subtest reliability coefficients average values exceed .80 for 9 of the 11 subtests. Only Picture Arrangement (.74) and Object Assembly (.68) fall short of expectations. Test-retest reliability coefficients affirm the excellent reliability of the Verbal and Full Scales, and show Performance IQ to be quite acceptable (.89 to .90). Test-retest coefficients for the subtests confirm the reliability of all tasks except Object Assembly and Picture Arrangement.

**Validity:** The manual for this revised edition does not provide new validity data, but refers interested persons to the manual for the original WAIS.

**Norms:** The total sample, comprising 1,880 individuals, was carefully stratified on the variables of sex, race [white-non-white], geographic region, occupational group, educational attainment, and urban-rural residence. The nine age groups ranged in size from 160 to 300.

**Suggested Uses:** The WAIS-R is recommended for use in clinical, educational, and research settings.
**Wechsler Adult Intelligence Scale – 3rd Edition (WAIS-III)**

**Purpose:** Designed as a comprehensive test of cognitive ability for adults.

**Population:** Ages 16-89.

**Score:** Verbal IQ, Performance IQ, and Full Scale IQ.

**Time:** (60 - 90) minutes.

**Author:** David Wechsler.

**Publisher:** The Psychological Corporation.

**Description:** The Wechsler Adult Intelligence Scale–Third Edition updates the WAIS-R, represents contemporary updated norms, and extends the age range. Some items have been modified, and there is greater discrimination for individuals in the mild to moderate mental retardation range as the floor has been extended. The artwork has been updated, but has been criticized as being distracting and overly detailed and discriminating against disadvantaged individuals, while the colors are thought to be potentially unfair to color blind individuals. There is less emphasis on timed performance, the new Matrix Reasoning subtest provides a better measure of fluid intelligence, the Letter-Number Sequencing subtest measures working memory, and the Symbol Search subtest (adapted from the WISC-III) measures processing speed. These changes strengthen the theoretical basis and statistical linkage to other measures of achievement and cognitive functioning.

**Scoring:** There are 14 subtests making up the Verbal and Performance Scales with 7 subtests each, however three of the subtests are supplemental or optional. Letter-Number Sequencing is a supplementary subtest for the Verbal Scale but must be administered for computing the Working Memory Index Score. Matrix Reasoning and Symbol Search are additions to the Performance Scale, and are also parts of the Perceptual Organization and the Processing Speed Indexes respectively. Eleven subtests are used to compute the IQ scores. There are four index scores, Verbal Comprehension, Perceptual Organization, Working Memory and Processing Speed Indexes also comprised of eleven subtests. While scoring rules and reverse rules are implemented consistently, and improvements in the presentation, scoring criteria and queries and prompts, are noted, some items seem to be marked arbitrarily for query, and complex judgments are sometimes required for scoring.

**Reliability:** Extensive testing of reliability was an emphasis of the WAIS-III. The split-half reliability coefficients are outstanding for the Full Scale IQ, the Verbal IQ, and the VCI, and are excellent for the Performance IQ, POI, and WMI. For most of the subtests the split-half reliabilities are excellent, however for the Object Assembly and the Picture Arrangement is below .75 for many of the age levels. There is less than excellent test-retest stability for the Letter-Number Sequencing, Picture Arrangement, or Object Assembly, and the PO and PS factors do not appear as separate constructs for the oldest age group. The WAIS-III helps to provide information about the interrelationships of a broad array of cognitive abilities as it is statistically linked to the WIAT and the WMS-III.

**Validity:** Evidence of concurrent validity of the WAIS-III and WAIS-R, WISC-III, WIAT, SB: FE has been provided, and the criterion-validity of .88 for the SB: FE and a range of .53-.81 with WIAT composites.

**Norms:** The normative sample was stratified for many key variables and was consistent with the latest census data. Oversampling was done for research on educational level and cognitive abilities, and to perform item bias analyses for African-American and Hispanic individuals. Another strength of the WAIS-III standardization sample was that the FSIQ was extended to 45 to 155 from 46 to 150.

**Suggested Use:** The WAIS-III can be used for assessment of learning disabilities. A sample of LD adults administered the WAIS-III and the WMS-III had low scores on the ACID profile, (Arithmetic, Coding [Digit Symbol-Coding], Information, and Digit Span). Discrepancies among index score were found with LD adult population using the WAIS-III, and may be a stronger way to identify LD. In LD groups, WMI←VCI, and PSI ←POI. Therefore combining the WMI and the PSI as in the SCALD profile (Symbol Search, Coding, Arithmetic, Letter-Number Sequencing, Digit Span) may be useful for investigating LD in adults. The WAIS-III is used for understanding ADHD, for assessment of mental retardation, and interpretation of age related differences in ability, such as age trends in working memory and intelligence.
Wechsler Bellevue Intelligence Scale

**Purpose:** Designed as a general test of cognitive ability for children and adults.

**Population:** Ages 10 and over.

**Score:** Verbal, Performance, and Full Scale IQ.

**Time:** (40-60) minutes.

**Author:** David Wechsler.

**Publisher:** The Psychological Corporation.

**Description:** The Wechsler Bellevue Intelligence Scale (WBIS) is a general test of intelligence, which Wechsler defined as, “... the global capacity of the individual to act purposefully, to think rationally, and to deal effectively with his environment.” In keeping with this definition of intelligence as an aggregate of mental aptitudes or abilities, the WBIS consists of 11 subtests divided into two parts, verbal and performance.

**Scoring:** The battery of tests consists of six Verbal and five non-verbal or Performance tests as follows: an information test, a general comprehension test, a memory span test (digits forward and backward), an arithmetical reasoning test, a similarities test, a vocabulary test, a block design test, an object assembly test and a digit symbol test.

**Reliability:** Reliability coefficients for the subtests range from .62 to .88. The Verbal IQ, Performance IQ, and Full Scale IQ have reliability coefficients of .84, .86, and .90, respectively.

**Validity:** The author investigates correlations between the WBIS and other tests of cognitive ability. The coefficients are as follows: Stanford-Binet, 1937 rev., .62; Otis, .73; Raven Progressive Matrices, .55, and the Army Alpha test, .53.

**Norms:** The WBIS was standardized on 1750 subjects of both sexes, ages 7-69. This sample’s demographic characteristics are not representative of the United State population.

**Suggested Uses:** Because of its dated normative and psychometric data, the WBIS is recommended for teaching purposes only.
Wechsler Individual Achievement Test - 2nd Edition

**Purpose:** Designed to assess individual achievement.

**Population:** Children, adolescents, college students, and adults, or age 4 through 85.

**Scales:** Reading, Mathematics, Written and Oral Language, and 9 sub-test scores

**Time:** 45 minutes for Pre K-K, 90 min. Grades 1-6, 90 - 120 min. Grades 7-16

**Author:** David Wechsler.

**Publisher:** The Psychological Corporation

**Description:** The Wechsler Individual Achievement Test - Second Edition (WIAT-II) revises the Wechsler Individual Achievement Test (WIAT; The Psychological Corporation, 1992). The areas covered by the test remain the same, but the depth and range of subject matter has been increased, with some new items, and updates of content to reflect changes in curriculum standards and incorporate cutting-edge research in the acquisition and assessment of educational skills. Specifically, the Listening Comprehension and Oral Expression subtests have been modified to coincide with skills demanded of the student in the classroom. The Reading Comprehension subtest now includes questions geared to a greater level of understanding, and assess reading rate and oral reading. Phonological decoding has been added to evaluate “word attack” skills, and there are several ways to demonstrate writing skills. An attempt has been made to establish a greater link between assessment and intervention by better analysis, which looks at both the product and the process involved in a skill. The age range of the WIAT has been extended in both directions including preschoolers for early assessment, and college students and adults in need of academic skills evaluation.

**Scoring:** Scoring for the WIAT-II has been made more clear and consistent with instructional practice. Standard scores are by age or by grade, (M=100, SD=15) with a range of 40 to 160 and seasonal tables for pre-K - 8.

**Reliability:** Internal consistency reliability estimates of the WIAT-II subtests are generally high (above .85) with the exception of the Written Expression and Listening Comprehension subtests in the school-aged sample and the Written Expression and Oral Expression subtests in the college-adult sample. The reliability estimates of these subtests were only somewhat lower (above .70). Internal consistency reliability of the Composite scores was very high (above .90) in both samples with the exception of the Oral Language Composite, which was above .85. In the school-aged sample, test-retest correlations for the subtests (across intervals of approximately 10 days) were consistently above .85 and test-retest correlations for the Composite scores were above .90. Test-retest correlations were somewhat lower in the college-adult sample, with correlations between .75 and .85 in Reading Comprehension, Written Expression, Oral Expression, and the composite scores for Written Language and Oral Language.

**Validity:** The corresponding subtests of the WIAT and the WIAT-II are strongly correlated (above .80) in the school-aged sample for those subtests with minimal content changes. The correlations were lower for subtests that had changed the most: Reading Comprehension (r = .74), Written Expression (r = .68), Listening Comprehension (r = .68), and Oral Expression (r = .62) subtests. Although it is evident that the refined WIAT-II subtests are assessing academic skills in different ways than traditional achievement tests, it is not yet fully clear whether the WIAT-II’s substantial content revisions will yield better, more usable indices of students’ academic abilities.

**Norms:** The WIAT-II was standardized on 5586 individuals with two standardization samples drawn for PreK-12 (ages 4-19) and for the college-adult population. Both standardization samples were stratified on the basis of grade, age, sex, race-ethnicity, geographic region, and parent education level, based on the data from the 1998 Bureau of the Census.

**Suggested Use:** The WIAT-II is useful in schools, clinics, private practices and residential treatment facilities. In conjunction with direct observation, history and additional measures the WIAT-II can assist with diagnosis, eligibility, placement, decisions regarding interventions.


**Wechsler Intelligence Scale for Children--Revised**

**Purpose:** Designed as a comprehensive measure of cognitive ability for children.

**Population:** Ages 6-16.

**Score:** Verbal, Performance, and Full Scale Scores.

**Time:** (50-75) minutes.

**Author:** David Wechsler.

**Publisher:** The Psychological Corporation.

**Description:** The Wechsler Intelligence Scale for Children-Revised (WISC-R) is a general test of intelligence, which Wechsler defined as, “…the global capacity of the individual to act purposefully, to think rationally, and to deal effectively with his environment.” In keeping with this definition of intelligence as an aggregate of mental aptitudes or abilities, the WISC consists of 13 subtests divided into two parts, verbal and performance.

**Scoring:** The WISC-R is a collection of 13 distinct subtest divided into two scales - a Verbal Scale and a Performance Scale. The six Verbal Scale tests use language-based items, whereas the seven Performance Scales use visual-motor items that are less dependent on language. Five of the subtest in each scale produce scale-specific IQS, and the 10 subtest scores produce a Full Scale IQ.

**Reliability:** Each of the three IQ scales has an internal consistency reliability coefficient of .89 or above in the standardization group over the entire age range covered by the scale. Average internal consistency reliability coefficients, based on the 11 age groups, are .96 for the Full Scale IQ, .94 for the Verbal Scale IQ, and .90 for the Performance Scale IQ. The average subtest reliability coefficients range from a low of .70 for Object Assembly to a high of .86 for Vocabulary. The average reliability coefficients range from .77 to .86 (Mdn = .80) for the Verbal Scale subtests and from .70 to .85 (Mdn = .72) for the Performance Scale subtests. Test-retest stability coefficients were .95 for the Full Scale IQ, .93 for the Verbal Scale IQ, and .90 for the Performance Scale IQ.

**Validity:** Because the WISC-R overlaps with the WPPSI in the age range of 6-0-0 to 6-7-15, either of the two tests can be used to evaluate children in this age range. The correlations were .80 for the Verbal Scales, .80 for the Performance Scales, and .82 for the Full Scales. In a study in which 5 to 6-year-old middle-class children were administered both tests, with a one-year test-retest interval, the WISC-R yielded IQS that were 5 points lower on the average than those of the WPPSI. Correlations between the two tests were .81 for the Verbal Scales, .80 for the Performance Scales, and .94 for the Full Scales. Four studies comparing the WISC-R and the Stanford-Binet: Fourth Edition are reported in the Technical Manual of the Fourth Edition. Correlations ranged from .66 to .83 between the WISC-R Full Scale IQ and the Fourth Edition composite.

**Norms:** The WISC-R was standardized on a sample of 2,200 American children selected as representative of the population on the basis of the 1970 U.S. Census.

**Suggested Uses:** It is suggested that the WISC-R is an appropriate instrument for practitioners and clinical researchers in assessing children’s intelligence.
Wechsler Intelligence Scale for Children - Third Edition

**Purpose:** Designed as a “measure of a child’s intellectual ability.”

**Population:** Ages 6-0 to 16-11.

**Scores:** Verbal, Performance and their related subscale scores, and a Full Scale IQ.

**Time:** (50-75) minutes.

**Author:** David Wechsler.

**Publisher:** The Psychological Corporation.

**Description:** The WISC-III is the third generation of the Wechsler Intelligence Scale for Children. Its predecessor, the WISC-R, is the most popular and widely researched test of children’s intelligence in history.

**Scoring:** The WISC-III is a collection of 13 distinct subtest divided into two scales - a Verbal Scale and a Performance Scale. The six Verbal Scale tests use language-based items, whereas the seven Performance Scales use visual-motor items that are less dependent on language. Five of the subtest in each scale produce scale-specific IQS, and the 10 subtest scores produce a Full Scale IQ.

**Reliability:** Subtest reliabilities (expressed as internal consistencies for all but the speeded subtest of Symbol Search and Coding) are moderate to excellent (.61 to .92). The consistency of IQS and Indices is very good to excellent (.80 to .97). Subtest stability coefficients, based on 353 children subdivided into three age groups, are adequate (.56 to .89). IQ and Index stability is mostly good to excellent (.74 to .95; only one coefficient below .80). Interrater reliabilities for selected Verbal Scale subtest are excellent (all greater than .92).

**Validity:** The manual reports strong correlations between WISC-III metrics and comparable metrics from the WPPSI-R, WISC-R, WAIS-R, Otis-Lennon School Ability Test, and Differential Ability Scales (rs between WISC-III IQS and comparable composites range from .59 to .92). Additionally, independent studies report correlations between WISC-III IQS/Indices and comparable metrics from other batteries are well within acceptable limits. The WISC-III manual reports appropriate correlations with achievement, and studies published since test publication also report appropriate IQ-achievement correlations in children representing normal, referred, learning disabled, severely emotionally disturbed, language/speech impaired, and hearing-impaired/deaf clinical categories. The WISC-III exhibits little predictive bias when used with males, females, whites, blacks, and Hispanics.

**Norms:** The normative sample is large (N=2,200) and representative of 1988 U.S. Census data.

**Suggested Uses:** It is suggested that the WISC-III is an appropriate instrument for practitioners and clinical researchers in assessing children’s intelligence.
Wechsler Intelligence Scale for Children – 4th Edition (WISC-IV)

Purpose: Designed to as a “measure of a child’s intellectual and cognitive ability.”

Population: Children, aged 6 years 0 months through 16 years 11 months.

Score: Four Index Scales and a Full Scale Scores.

Time: (65 - 80) minutes for most children.

Author: David Wechsler.

Publisher: The Psychological Corporation.

Description: This fourth edition updates the WISC-III and provides subtest and composite scores representing intellectual functioning in general and specific cognitive abilities. The changes in the WISC-IV represent current research on cognitive development, intellectual assessment and cognitive processes. The revisions also include updated norms, additional subtests, and emphasis on scores reflecting discrete areas of cognitive functioning. New subtests include Word Reasoning, Matrix Reasoning, Picture Concepts, Cancellation, and Letter-Number Sequencing. Former WISC-III subtests that were dropped include Mazes, Object Assembly, and Picture Arrangement.

Scoring: Requires Subtest raw scores are converted into scaled scores that are summed into four index scores. The Verbal Comprehension Index (VCI) is composed of three subtests that are mostly verbal, the Perceptual Reasoning Index (PRI) uses three subtests that rely less on verbal skills, the Working Memory Index (WMI) consists of items requiring recall and repetition of letters and numbers, and the Processing Speed Index (PSI) uses non-verbal, timed search and coding tests. The optional subtests are not included in the index scores unless they replace a core test.

Reliability: The subtest reliability coefficients for internal consistency ranged from .79 to .90 with a median of .86. These coefficients showed substantial improvement from those of WISC-III subtests. The index scores reliability coefficient ranged from .88 PSI to .97 FS with a median of .92. These are identical to or slightly higher than WISC-III corresponding scales.


Norms: One of the major goals was to update the norms to be more representative of the relevant population. The normative sample included 2200 children aged 6:0-16:11 and additional samples from special groups. The sample was stratified on demographic variables of age, sex, race/ethnicity, parent education level, and geographic region based on the March 2000 U.S. census data.

Suggested Use: The WISC-IV is an appropriate instrument for practitioners and clinical researchers in assessing children’s intelligence and general cognitive functioning. When used with other assessment tools, it can be useful in identifying giftedness, mental retardation, and cognitive strengths and weaknesses. The test results can be useful in treatment planning, in placement and provision of clinical or educational services, and can add important information to a neuropsychological evaluation.
Weschler Preschool and Primary Scale of Intelligence-Revised

Purpose: Designed as a comprehensive measure of cognitive ability for preschoolers.

Population: Ages 4.0-6.5.

Score: Verbal, Performance, and Full Scale IQs.

Time: (50-75) minutes.

Author: David Wechsler.

Publisher: The Psychological Corporation.

Description: The Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R) is one of the major instruments for assessing the cognitive ability of young children. Like its predecessor, the WPPSI-R is separate and distinct from, although similar in form and content to, the WISC-R.

Scoring: The WPPSI-R contains 12 subtests, 6 in the Performance Scale and 6 in the Verbal Scale. Five of the six subtests in each scale are designated as the standard subtests. They are Object Assembly, Geometric Design, Block Design, Mazes, and Picture Completion in the Performance Scale and Information, Comprehension, Arithmetic, Vocabulary, and Similarities in the Verbal Scale. The optional subtests are Animal Pegs in the Performance Scale and Sentences in the Verbal Scale. The WPPSI-R employs the Deviation IQ (M = 100, SD = 15) for the Verbal, Performance, and Full Scale IQs and scaled scores (M = 10, SD = 3) for the subtests. A raw score is first obtained on each subtest and then converted to a scaled score within the examinee’s own age group through use of a table in the WPPSI-R manual.

Reliability: The WPPSI-R Performance, Verbal, and Full Scale IQs have excellent reliability in eight of the nine age groups covered by the test. From ages 3 through 6.5 years, the reliabilities for each of the three IQS range from .90 to .97. Across the nine age groups, the average internal consistency reliabilities are .92 for the Performance Scale IQ, .95 for the Verbal Scale IQ, and .96 for the Full Scale IQ. Test-retest reliabilities for a period of approximately 3 to 7 weeks for Performance, Verbal, and Full Scale IQs were .87, .89, and .91, respectively.

Validity: Reviewers note that because of the similarity of content, much of the research on the validity of the original WPPSI is pertinent to the WPPSI-R. These studies indicate that the WPPSI has adequate construct, concurrent, and predictive validity for many types of normal and handicapped children in the age range from 4 to 6.5 years.

Norms: The WPPSI-R was standardized on 1,700 children, 100 boys and 100 girls in each of eight age groups from 3 to 7 years and one group of 50 boys and 50 girls from 7 years. The 1986 U.S. census data were used to select representative children for the normative sample.

Suggested Uses: It is suggested that the WPPSI-R is an appropriate instrument for practitioners and clinical researchers in assessing children’s cognitive ability.
Wechsler Preschool and Primary Scale of Intelligence - 3rd Edition

Purpose: Designed to measure intellectual abilities and academic achievement.

Population: Children, aged 2:6 years through 7:3 years.

Scales: Verbal, Performance, Full Scale IQs, Processing Speed Quotient, and General Language Composite

Time: 30-35 minutes for children aged 2:6-3:11 and 40-50 minutes for ages 4-7:11.

Author: David Wechsler

Publisher: The Psychological Corporation

Description: The Wechsler Preschool and Primary Scale of Intelligence - Third Edition is a revision of the WPPSI-R and has extended the age range, updated the norms, added new subtests and composite scores, and claims to have a developmentally appropriate structure based on contemporary intelligence and cognitive development theory. The artwork has been updated, and some the test material has been made more child-friendly and engaging. Some modifications in the administration and scoring make the scales easier to use.

Scoring: The WPPSI-III contains 4 core subtests for ages 2:6 to 3:11; Receptive Vocabulary, Information, (VIQ) Block Design, Object Assembly (PIQ), and 7 for ages 4:0 to 7:3; Information, Vocabulary, Word Reasoning (VIQ), Block Design, Matrix Reasoning, Picture Concepts (PIQ), and Coding (FSIQ). Supplemental tests for 2:6 to 3:11 group are Picture naming and Receptive Vocabulary (GLC). For the 4 to 7:3 group the supplementary tests are Symbol Search, Comprehension, Picture Completion, Similarities, Object Assembly (PSQ), and the optional subtests are Receptive Vocabulary and Picture Naming (GLC). The WPPSI-III employs the Deviation IQ (M=100, SD=15) for the Verbal, Performance and Full Scale IQS, and scaled scores (M=100, SD=3) for the subtests.

Reliability: The reliability coefficient of the WPPSI-III subtests range from .83 to .95. The reliability coefficients for the composite scales ranged from .89 to .96. Test-retest reliabilities for a mean interval of 26 days for the 2:6 to 3:11 year old group Verbal, Performance, Full and General Language scores were .90, .84, .92 and .92 respectively. For the 4 to 7:3 year old group for Verbal, Performance, Processing Speed, Full and General Language were 92, .87, .93, .92, and .90 respectively.

Validity: The evidence of previous research that because of the similarity of content, much of the research on the validity of the WPPSI- R The scores derived from the WPPSI-R correlate well with the WPPSI, WISC-R, Stanford Binet (4th ed.), and McCarthy Scales (rs between WPPSI-R FSIQs and other test composites range from .74 to .90). The correlation between the WPPSI-R FSIQ and the Kaufman-Assessment Battery for Children (K-ABC) Mental Processing Composite is low (.49), but the K-ABC has consistently yielded lower correlations with other intelligence tests as well. These results imply good criterion validity for the WPPSI-R. There are also studies showing the discriminant validity of the WPPSI-R with gifted, mentally deficient, learning disabled, and speech-language impaired children.

Norms: The normative sample included 1700 children in nine age groups. The sample was representative of the US population of children aged 2:6 to 7:3 for sex, race/ethnicity, parental education level and geographic region.

Suggested Use: The WPPSI-III for assessment of general intellectual functioning as part of an assessment to identify giftedness, delays, or mental retardation. Results can be useful in guiding clinical or school related placements.
Whitaker Index of Schizophrenic Thinking

**Purpose:** Designed to measure degrees of schizophrenic thinking

**Population:** Mental Patients 16 years and older with 8th grade education and IQ of 80

**Score:** Similarities, Word Pairs, New Inventions, Total

**Time:** 20 minutes for each Form

**Author:** Leighton C. Whitaker

**Publisher:** Western Psychological Services

**Description:** The Whitaker Index of Schizophrenic Thinking (WIST) is offered as a measure of schizophrenic thinking, defined as thinking that is “illogical,” “impaired,” and “unwitting.” The WIST has two main uses: to differentiate persons with schizophrenia from those without schizophrenia, and to serve as an objective measure of degree of thinking impairment. There are two nonequivalent alternate forms for the test: Form A which has item content designed to arouse emotion or anxiety and Form B has more neutral item content. Each form has twenty-five multiple-choice items, nine “similarities” items that require selecting the answer that is most similar to a given word, nine “word pair” items requiring selection of an answer that is most similar in meaning to a given pair of words, and seven “new inventions” items requiring selection of the most likely consequence of a new invention.

**Scoring:** Each item has a range of scores from 0 to 4 representing the degree of schizophrenic thinking. The score is the total of the weighted scores for each wrong answer. All individual scores are summed and recorded as the TOTAL. The subscales for Similarities, Word Pairs and New Inventions are summed for the total WIST score. The time taken for the subject to respond is also recorded and summed to obtain the WIST Index.

**Reliability:** The manual contains evidence for reliability including internal consistency measures on an “early version of Form A” (not further specified) yielding a Kuder-Richardson Formula 20 reliability coefficient of .77 on unweighted scores. Hoyt reliability coefficients on Forms A and B were approximately .80 using unweighted scores. Forms A and B differ as described above, and are noted to be noncomparable for the purpose of reliability estimation. Test-retest reliability is considered impractical because the subject learns the correct answer during the inquiry. Additional evidence of WIST reliability would be desirable.

**Validity:** Analysis showed that use of the Index (Score plus Time) value and a cutoff of 20 on Form A, 17 on Form B produced maximum differentiation of diagnosed schizophrenics and non-schizophrenics, at approximately 80% efficiency. Such accuracy clearly makes the WIST useful for research applications, but presents difficulty if the WIST is used for clinical assessment and individual decision making. Although Whitaker argues the WIST measures a particular type of thinking disorder specific to schizophrenia, others argue that the WIST reflects generalized cognitive deficit. In another study, 20 schizophrenics (10 paranoid, 10 non-paranoid) and 10 non-schizophrenic patients were assessed, but the WIST Index did not differentiate well between schizophrenic and non-schizophrenic patients. However, the WIST Index had a significant negative correlation with the Shipley Institute of Living Scale, which measures general cognitive functioning. In a study with anti-psychotic drugs, no significant relationship between drug dosage level and cognitive performances were found.

**Norms:** The WIST standardization sample included 38 hospitalized acute and 44 chronic schizophrenic patients, 55 hospitalized non-schizophrenics, and a “normal” group of 50 including 26 maintenance workers and 24 college students. Careful patient selection is described.

**Suggested Use:** The WIST can be useful in psychiatric emergency rooms, hospitals and clinics, community mental health center, college or university health centers and counseling centers, and recruitment centers for police or the military. The WIST can be used to evaluate symptomatic status of patients previously diagnosed as schizophrenic.
Wide Range Achievement Test

Purpose: Designed to measure reading recognition, spelling, and arithmetic computation.

Population: Ages 5-11, 12 and over.

Score: 3 scores: Spelling, Arithmetic, Reading.

Time: (20-30) minutes.

Authors: Joseph F. Jastak and Sarah Jastak.

Publisher: Jastak Associates, Inc.

Description: The Wide Range Achievement Test (WRAT) is a brief achievement test measuring reading recognition, spelling, and arithmetic computation. There are two levels; level I is normed for children ages 5-0 to 11-11; level II is normed for children aged 12 through adults aged 64.

Scoring: Norms provided for the 1978 edition include standard scores with a mean of 100 and a standard deviation of 15, percentile scores, and grade levels. The standard scores are scaled based on the norm group; the grade levels are arbitrarily assigned and can be interpreted only as rough references to achievement level. Only standard scores should be used for comparisons among scores.

Reliability: The manual reports split-half reliabilities of .98 for Reading at both levels, .94 for Arithmetic at both levels, .96 for Spelling I, and .97 for Spelling II. During the norming study, both levels of the WRAT were administered to children ages 9 through 14. Since there is overlap in skills tested between the high end of level I and the low end of level II, this provides another estimate of the reliability of both. On Reading and Spelling, split-half reliabilities ranged from .88 to .94 for different age groups; on Arithmetic they ranged from .79 to .89. These results indicate that overall the reliability of the WRAT is excellent.

Validity: The test most similar to the WRAT is the Peabody Individual Achievement Test (PIAT), another short, individually administered test which covers comparable material. In general the WRAT correlates very highly with the PIAT. The WRAT correlates moderately with various IQ tests, in the range of .40 to .70 for most groups and most tests.

Norms: The 1978 WRAT norms are based on 15,200 subjects for seven states. According to the manual, no attempt was made to make the sample representative of national characteristics. The manual states that minorities were represented, but gives no data on their representation. The sample was stratified by age, sex, and approximately by ability.

Suggested Uses: Recommended uses for the test described in the manual include comparing achievement of one person to another, determining learning ability or learning disability, comparing codes with comprehension in order to prescribe remedial programs, and informally assessing error patterns to plan instructional programs.
**Wide Range Achievement Test -- Revised**

**Purpose:** Designed to "measure the codes which are needed to learn the basic skills of reading, spelling, and arithmetic."

**Population:** Ages 5-0 to 11-11, 12-0 to 75.

**Score:** 3 scores: Reading, Spelling, and Arithmetic.

**Time:** (15-30) minutes.

**Authors:** Sarah Jastak, Gary S. Wilkinson, and Joseph Jastak.

**Publisher:** Jastak Associates, Inc.

**Description:** The Wide Range Achievement Test--Revised (WRAT-R) is the sixth edition of the popular test that was first published in 1936. Like the earlier versions, the WRAT-R contains three subtests: Reading (recognizing and naming letters and words), Spelling (writing symbols, name, and words), and Arithmetic (solving oral problems and written computations). The authors of the WRAT-R stress that the test is designed to measure basic school codes rather than comprehension, reasoning, and judgement processes.

**Scoring:** The manual contains information to transform the raw scores into standard scores.

**Reliability:** The Rasch analysis provided person-separation and item-separation values and these are given as evidence of internal consistency. Traditional internal-consistency data, such as split-half or alpha coefficients, are not provided and are needed. Test-retest reliability coefficients range from .79 (Level 2 Arithmetic) to .97 (Level 1 Spelling), but no indication of the time interval between test and retest is given and the coefficients are based on small samples for only a few age groups in the schoolage range.

**Validity:** WRAT-R subtests have moderately high correlations with Woodcock-Johnson achievement subtests. The manual supplies two sources of evidence for construct **Validity:** person and item-separation values and increasing raw scores with age. The manual reports very high correlations (i.e., .91 to .99) between the WRAT and WRAT-R, but the correlations were calculated using the WRAT-R norm sample and an "arbitrary" sample of previously administered WRATs. It appears that the correlations are not based on a sample of subjects who took both tests. Spruill and Beck found WRAT/WRAT-R correlations of .98, .97, and .71 for the Reading, Spelling, and Arithmetic subtests, respectively.

**Norms:** A stratified national sampling plan was used for the standardization of the WRAT-R. A total of 5,600 subjects, or 200 subjects in each of 28 age groups from 5 to 74 years, composed the sample. Stratification variables included age, sex, race, geographical region, and metropolitan versus nonmetropolitan residence.

**Suggested Uses:** General uses for the test described in the manual include comparing achievement of one person to another, determining learning ability or learning disability, comparing codes with comprehension in order to prescribe remedial programs, and informally assessing error patterns to plan instructional programs.
Wonderlic Personnel Test

**Purpose:** Designed to measure general mental ability for aid in personnel selection.

**Population:** Adult employment applicants.

**Score:** Yields one total score.

**Time:** 12 [20] minutes.

**Author:** E.F. Wonderlic.

**Publisher:** E.F. Wonderlic & Associates, Inc.

**Description:** The Wonderlic Personnel Test (WPT), so named to reduce the possibility that job applicants will think they are taking an intelligence test, was originally a revision of the Otis Self-Administering Tests of Mental Ability. The WPT is a 50-item, 12-minute omnibus test of intelligence. The items and the order in which they are presented provide a broad range of problem types (e.g., analogies, analysis of geometric figures, disarranged sentences, definitions) intermingled and arranged to become increasingly difficult. The WPT exists in 16 forms, and was designed for testing adult job applicants in business and industrial situations.

**Scoring:** The WPT yields one final score which is the sum of correct answers.

**Reliability:** The manual reports odd-even reliabilities, which are not appropriate for speeded tests; however, it also reports test-retest reliabilities of .82 to .94, and interform reliabilities of .73 to .95.

**Validity:** Correlations with educational level and/or academic achievement are between .30 and .80.

**Norms:** White adults across all occupational categories.

**Suggested Uses:** Suggested uses include as measure of general ability for employment screening, although it should be used with caution.
**Word Association Test**

**Purpose:** Designed to reveal associative connections between stimuli words and responses.

**Population:** Adults.

**Score:** N/A.

**Time:** N/A.

**Authors:** D. Rapaport, M. Gill, and R. Schafer.

**Publisher:** International Universities Press, Inc.

**Description:** The word association method is a psychological test intended to reveal associative connections between stimulus words and free verbal responses. Subjects are instructed to give discrete verbal responses to common stimulus words (usually nouns or verbs). Because verbalizations reflect ideation, it is assumed that the systematic study of associative thought can reveal information about an individual’s personality characteristics, areas of emotional disturbances, and the like. Although the method has no necessary tie to any particular theory, it is usually employed as a projective technique.

**Scoring:** There are no formal scoring procedures employed with this method.

**Reliability and Validity:** The manual provides no reliability and Validity information. Very few validity and reliability studies have been undertaken because internal clinical judgment constitutes such an integral part of the approach.

**Norms:** Norms are included in the manual for normal college students and schizophrenics.

**Suggested Uses:** The Word Association Test is recommended as part of a comprehensive test battery in clinical and research settings.
Wide Range Achievement Test -- Revised

**Purpose:** Designed to “measure the codes which are needed to learn the basic skills of reading, spelling, and arithmetic.”

**Population:** Ages 5-0 to 11-11, 12-0 to 75.

**Score:** 3 scores: Reading, Spelling, and Arithmetic.

**Time:** (15-30) minutes.

**Authors:** Sarah Jastak, Gary S. Wilkinson, and Joseph Jastak.

**Publisher:** Jastak Associates, Inc.

**Description:** The Wide Range Achievement Test--Revised (WRAT-R) is the sixth edition of the popular test that was first published in 1936. Like the earlier versions, the WRAT-R contains three subtests: Reading (recognizing and naming letters and words), Spelling (writing symbols, name, and words), and Arithmetic (solving oral problems and written computations). The authors of the WRAT-R stress that the test is designed to measure basic school codes rather than comprehension, reasoning, and judgement processes.

**Scoring:** The manual contains information to transform the raw scores into standard scores.

**Reliability:** The Rasch analysis provided person-separation and item-separation values and these are given as evidence of internal consistency. Traditional internal-consistency data, such as split-half or alpha coefficients, are not provided and are needed. Test-retest reliability coefficients range from .79 (Level 2 Arithmetic) to .97 (Level 1 Spelling), but no indication of the time interval between test and retest is given and the coefficients are based on small samples for only a few age groups in the schoolage range.

**Validity:** WRAT-R subtests have moderately high correlations with Woodcock-Johnson achievement subtests. The manual supplies two sources of evidence for construct validity: person and item-separation values and increasing raw scores with age. The manual reports very high correlations (i.e., .91 to .99) between the WRAT and WRAT-R, but the correlations were calculated using the WRAT-R norm sample and an “arbitrary” sample of previously administered WRATs. It appears that the correlations are not based on a sample of subjects who took both tests. Spruill and Beck found WRAT/WRAT-R correlations of .98, .97, and .71 for the Reading, Spelling, and Arithmetic subtests, respectively.

**Norms:** A stratified national sampling plan was used for the standardization of the WRAT-R. A total of 5,600 subjects, or 200 subjects in each of 28 age groups from 5 to 74 years, composed the sample. Stratification variables included age, sex, race, geographical region, and metropolitan versus nonmetropolitan residence.

**Suggested Uses:** General uses for the test described in the manual include comparing achievement of one person to another, determining learning ability or learning disability, comparing codes with comprehension in order to prescribe remedial programs, and informally assessing error patterns to plan instructional programs.
**Wide Range Achievement Test - Third Edition (WRAT-3)**

**Purpose:** Designed to assess the cognitive ability of children.

**Population:** All individuals aged 5-75.

**Score:** Absolute, Standard, and Grade scores for reading, spelling and arithmetic.

**Time:** (15-30) minutes for each of the 3 forms.

**Author:** Joseph Jastak / Gary Wilkenson.

**Publisher:** Wide Range, Inc.

**Description:** The re-standardization of the WRAT serves to expand the scope of the test to include pre and post testing while maintaining an ease and reliability of previous editions. The WRAT-3 uses a single level format, and has two alternate forms, the BLUE and TAN, which provide the traditional three subtests of previous editions. The alternate forms may be used in combination with one another for a more qualitative assessment of academic skill, or singularly, leaving the other form to be used for testing at a later date. There are three subtests contained on each of the alternate forms. The reading subtest includes the recognition and naming of letters and pronunciation of words out of context. In the spelling subtest, the examinee is asked to write his or her own name, and then to write letters and words as they are dictated. The Arithmetic involves counting, reading number symbols, solving oral problems, and doing written computations.

**Scoring:** The norms for all tests on the Blue, the Tan, and Combined forms provide raw scores, standard scores with a mean of 100 and a standard deviation of 15, absolute scores and grade scores. The raw score is a basic unit of test measure, and has limited interpretive use. The absolute scores are an interval scale obtained through a Rasch analysis. These scores are suitable for use in statistical studies, pre-post testing, local norming and cut-off levels. The standard scores are scaled on the norm. They are also an interval scale, with equal gradient so they can be used for parametric statistical analyses. The grade scores are defined by mean performance for a particular grade level. The scores are ordinal and have limited use for interpretation of test results.

**Reliability:** Median test coefficient alphas for the nine tests of WRAT-3 range from .82 to .95. The range is higher for the three Combined tests (Both BLUE and TAN cards). Alternate form correlations of Reading (.92), Spelling (.93), and Arithmetic (.89) also support the reliability of the measure.

**Validity:** The WRAT-3 intents to measure basic academic skills. There is a high correlation with the WRAT-R (.79-.92) The WRAT correlated moderately well with the WAIS-R. (.66, .66, and .73) for the Reading, Spelling and Arithmetic Combined scores and the Full Scale WISC-R score.

**Norms:** A stratified national sampling included nearly 5000 individuals provided data for standardization of the WRAT-3.

**Suggested Use:** The WRAT-3 is used to measure the basic codes used to learn reading, spelling, and arithmetic. When used with a measure of general intelligence that has the same SD, the WRAT-3 can be useful in determining learning ability or disability.
**Wide Range Assessment of Memory and Learning**

**Purpose:** Designed to evaluate a child’s ability for learning and memorizing information.

**Population:** Ages 5 through 17

**Scales:** Verbal Memory Index, Visual Memory Index, Learning Index, General Memory Index

**Time:** 45-60 minutes for Core Battery; 83-102 minutes for Expanded Battery.

**Authors:** David Sheslow & Wayne Adams

**Publisher:** Jastak Associates Inc

**Description:** The Wide Range Assessment of Memory and Learning (WRAML) is designed to assess memory and learning functions across the school years.

**Scoring:** There are nine subtests each yielding a norm-referenced score. Scores on three subtests are combined to give a Verbal Memory Index, a Visual Memory Index, and a Learning Index. The scaled scores for these three indexes are then summed to yield a General Memory Index. Four of the nine subtests (Verbal Learning, Story Memory, Sound Symbol, and Visual Learning) ask for both immediate and delayed recall. Interpretations are provided, based on the age of the child tested, of the difference between the immediate and delayed score. Thus, the nine subtests of the WRAML yield a total of 18 scores. The GMI and Verbal, Visual, and Learning Indexes can be computed in percentiles and standard scores. Individual subtests yield scaled scores.

**Reliability:** For the nine subtests, the reliability coefficients are usually between .80 and .85. When subtests are combined, reliabilities range from approximately .90 to .96. Test reliability is as high for younger as for older children.

**Validity:** Construct validity, used in Rasch measurement, indicates excellent item definitions of variables measured and internal consistency. Comparisons with the WMS-R for adolescents (16-17 year). The WRAML appears superior to the WMS-R for use with adolescents. The WRAML is well validated, and is widely used in research. Information is also included concerning the standard error of measurement for each subtest and index for each age group, along with correlations between scores on the WRAML and other standardized instruments such as the McCarthy Memory Index, Stanford Binet Short-Term Memory, and the Wechsler Memory Scale.

**Norms:** The test was normed and standardized based on samples of children from 5 to 16 years of age. There were approximately 112 children in each subgroup (half-year intervals). The total norming group consisted of 2,363 individuals. The norming samples are representative of the US population with regard to gender, geographic region, and parental occupation.

**Suggested Use:** The major use for the WRAML is clinical in terms of providing incremental information in making an individual diagnosis. The WRAML is used to evaluate learning and school-related problems. It is helpful in evaluating the effects of a language disability or problems with verbal memory retrieval, in identifying the inefficient or disorganized memory strategies of a bright but under-achieving student, and in pointing to the functional inefficiency of memory in a child with attention deficit. It can be used to assess memory following head injury.
Wonderlic Personnel Test

**Purpose:** Designed to measure general mental ability for aid in personnel selection.

**Population:** Adult employment applicants.

**Score:** Yields one total score.

**Time:** 12 (20) minutes.

**Author:** E.F. Wonderlic.

**Publisher:** E.F. Wonderlic & Associates, Inc.

**Description:** The Wonderlic Personnel Test (WPT), so named to reduce the possibility that job applicants will think they are taking an intelligence test, was originally a revision of the Otis Self-Administering Tests of Mental Ability. The WPT is a 50-item, 12-minute omnibus test of intelligence. The items and the order in which they are presented provide a broad range of problem types (e.g., analogies, analysis of geometric figures, disarranged sentences, definitions) intermingled and arranged to become increasingly difficult. The WPT exists in 16 forms, and was designed for testing adult job applicants in business and industrial situations.

**Scoring:** The WPT yields one final score which is the sum of correct answers.

**Reliability:** The manual reports odd-even reliabilities, which are not appropriate for speeded tests; however, it also reports test-retest reliabilities of .82 to .94, and interform reliabilities of .73 to .95.

**Validity:** Correlations with educational level and/or academic achievement are between .30 and .80.

**Norms:** White adults across all occupational categories.

**Suggested Uses:** Suggested uses include as measure of general ability for employment screening, although it should be used with caution.
Word Association Test

Purpose: Designed to reveal associative connections between stimuli words and responses.

Population: Adults.

Score: N/A.

Time: N/A.

Authors: D. Rapaport, M. Gill, and R. Schafer.

Publisher: International Universities Press, Inc.

Description: The word association method is a psychological test intended to reveal associative connections between stimulus words and free verbal responses. Subjects are instructed to give discrete verbal responses to common stimulus words (usually nouns or verbs). Because verbalizations reflect ideation, it is assumed that the systematic study of associative thought can reveal information about an individual’s personality characteristics, areas of emotional disturbances, and the like. Although the method has no necessary tie to any particular theory, it is usually employed as a projective technique.

Scoring: There are no formal scoring procedures employed with this method.

Reliability and Validity: The manual provides no reliability and Validity information. Very few validity and reliability studies have been undertaken because internal clinical judgment constitutes such an integral part of the approach.

Norms: Norms are included in the manual for normal college students and schizophrenics.

Suggested Uses: The Word Association Test is recommended as part of a comprehensive test battery in clinical and research settings.
Woodcock Johnson III - Tests of Achievement

**Purpose:** Designed to measure intellectual abilities and academic achievement.

**Population:** Ages 2 to 90+

**Scales:** Reading, Oral Language, Mathematics, Written Language, Knowledge, Supplemental Scores, Intra-Achievement Discrepancies, and Total Achievement

**Time:** 60-70 minutes.

**Authors:** Richard Woodcock, Nancy Mather, Kevin McGrew

**Publisher:** Riverside Publishing

**Description:** This Woodcock Johnson III - Tests of Achievement (WJ III ACH) updates and expands the WJ-R Tests of Achievement (WJ-R ACH) with the intention of improving the diagnostic power of the instrument. The WJ III ACH is designed to identify and describe an individual's current strengths and weaknesses. Three oral tests, a diagnostic spelling test and a measure of phonological awareness have been added to evaluate fluency in reading and in math. The other WJ III battery, the WJ III Test of Cognitive Abilities (WJ III COG) is designed to measure general and specific cognitive functions. When these two batteries are administered together, they allow the tester to investigate over-underachievement and to examine patterns of intra-individual discrepancies among cognitive or achievement areas. The batteries of the WJ III COG and WJ III ACH are further classified as standard or extended.

**Scoring:** Scoring is to be completed during testing to determine basal and ceiling levels. Incorrect responses are recorded verbatim. Raw scores are totaled and converted into age and grade equivalents, percentile ranks, and discrepancy scores with use of the Scoring Tables. Compuscore and Profiles Program are used for all other scoring. The WJ III ACH includes the basic interpersonal communication skill (BICS) scores, and the cognitive-academic language proficiency (CALP) score. These scores are helpful for determining the ability of nonnative speakers of English to function in informal settings and academic settings (e.g., college lectures), respectively.

**Reliability:** The median reliability coefficient alphas for all age groups for the standard battery of the WJ III ACH for tests 1 through 12 ranged from .81 to .94. For the Extended battery, median coefficients ranged from .76 to .91. The reliability scores for the WJ III meet or exceed standards. The median cluster reliabilities are mostly .90 or higher, and the individual test reliabilities are mostly .80 or higher, and can be used for decision making purposes with support from other sources.

**Validity:** The technical manual presents a considerable amount of evidence supporting the validity of scores from the test, noting that the earlier versions of the battery have also been shown to have validity. The WJ III ACH content is similar to other achievement tests in subject areas and to established practice in schools. Growth curves of cluster scores in the technical manual illustrate expected developmental progressions, with steep growth from age 5 to 25, with a decline thereafter. The most extensively presented data focus on validity evidence from confirmatory factor analyses of test scores from participants age 6 to adult. These analyses show that the seven-factor CHC cognitive model is more closely related to the observed internal structure than six other models derived from other commonly used cognitive batteries. The internal correlations of the entire battery are consistent with relations between areas of achievement and between areas of achievement and ability clusters.

**Norms:** The normative samples (pre-school, school aged, college and adult) were selected to be representative of the US population, based on geographic distribution, community size and type, and SES, occupational status, sex, race.

**Suggested Use:** The WJ III Tests of Achievement has uses for educational, clinical or research purposes, for diagnosis, educational programming, guidance, program evaluation.
Woodcock Johnson III - Tests of Cognitive Skills

**Purpose:** Designed to measure, together with the WJ III Ach, intellectual abilities and academic achievement.

**Population:** Ages 2 to 90+

**Scales:** Verbal Ability, Thinking Ability, Cognitive Efficiency, Supplemental, Intra-Cognitive Discrepancies, Predicted Achievement and General Intellectual Ability (GIA)

**Time:** 60-70 minutes.

**Authors:** Richard Woodcock, Nancy Mather, Kevin McGrew

**Publisher:** Riverside Publishing

**Description:** The Woodcock Johnson III - Tests of Cognitive Skills (WJ III COG) updates and expands the WJ-R Tests of Cognitive Ability (WJ-R COG) and is designed to measure general and specific cognitive functions. The other battery the WJ III ACH is designed to identify and describe an individual’s current strengths and weaknesses. When these two batteries are administered together, they allow the tester to investigate over-underachievement and to examine patterns of intra-individual discrepancies among cognitive or achievement areas. The batteries of the WJ III COG and WJ III ACH are further classified as standard or extended. New tests, clusters and interpretive procedures have been added with the intention of improving the diagnostic power of the instrument and increasing the breadth and depth of coverage. The clusters are grouped into three cognitive areas, Verbal Ability, Thinking Ability, and Cognitive Efficiency. Oral language tests have been moved to the WJ III ACH.

**Scoring:** Requires scoring during testing to determine basal and ceiling levels. Raw scores are totaled and converted into age and grade equivalents, percentile ranks, and discrepancy scores with use of the Scoring Tables. Compuscore and Profiles Program are used for all other scoring. The WJ III COG provides two indices of general cognitive functioning (i.e., intelligence) by means of the General Intellectual Ability (GIA) score and the Brief Intellectual Ability (BIA) score.

**Reliability:** The median reliability coefficient alphas for all age groups for the standard battery of the WJ III COG for tests 1 through 10 ranged from .81 to .94. For the Extended battery, median coefficients ranged from .74 to .97. The reliability scores for the WJ III meet or exceed standards. The median cluster reliabilities are mostly .90 or higher, and the individual test reliabilities are mostly .80 or higher, and can be used for decision making purposes with support from other sources.

**Validity:** The technical manual presents a considerable amount of evidence supporting the validity of scores from the test, noting that the earlier versions of the battery have also been shown to have validity. Test content on the WJ III COG has emerged from previous versions, is similar to the content found on other well-established cognitive measures, or is based on sound experimental instruments.

**Norms:** The normative samples (pre-school, school aged, college and adult) were selected to be representative of the US population, based on geographic distribution, community size and type, and SES, occupational status, sex, race.

**Suggested Use:** The WJ III Tests of Cognitive Ability has uses for educational, clinical or research purposes. As a diagnostic tool, it can determine specific strengths and weaknesses, and can help associate certain factors on future development. Intra-ability, intra-individual and ability/achievement discrepancies can be useful in determining the existence of specific disabilities, and in selection for LD programs. The test results may be helpful in educational programming, individual program planning, guidance, in assessing growth, in research and evaluation and it is an excellent instrument for introducing individual assessment in college and university courses. Psychometric training.
**Word Association Test**

**Purpose:** Designed to reveal associative connections between stimuli words and responses.

**Population:** Adults.

**Score:** N/A.

**Time:** N/A.

**Authors:** D. Rapaport, M. Gill, and R. Schafer.

**Publisher:** International Universities Press, Inc.

**Description:** The word association method is a psychological test intended to reveal associative connections between stimulus words and free verbal responses. Subjects are instructed to give discrete verbal responses to common stimulus words (usually nouns or verbs). Because verbalizations reflect ideation, it is assumed that the systematic study of associative thought can reveal information about an individual’s personality characteristics, areas of emotional disturbances, and the like. Although the method has no necessary tie to any particular theory, it is usually employed as a projective technique.

**Scoring:** There are no formal scoring procedures employed with this method.

**Reliability and Validity:** The manual provides no reliability and Validity information. Very few validity and reliability studies have been undertaken because internal clinical judgment constitutes such an integral part of the approach.

**Norms:** Norms are included in the manual for normal college students and schizophrenics.

**Suggested Uses:** The Word Association Test is recommended as part of a comprehensive test battery in clinical and research settings.