Cate Crow

TEST RESOURCE GUIDE

VOLUME VII

ASSESSMENT OF COGNITION AND AFFECT

1999 Edition

BOARD OF EDUCATION OF THE CITY OF NEW YORK

PREFACE

This volume is one in a series of guides which contains a compendium of reviews of measurement instruments. Other volumes in the series which are now available are:

Volume I Introduction

Volume I explains the nature of New York City's assessment needs and the criteria by which published instruments are evaluated. All test users are urged to read this volume to become aware of the terms and concepts of tests and measurements and the standards of instrument selection in New York City Public Schools.

Volume II Preschool Assessment and Kindergarten Supplement

Volume II contains reviews of published instruments which are used in the assessment of preschool children. Tests that are used for children no earlier than kindergarten are included in the supplement. All test reviews focus on the preschool or kindergarten populations.

Volume III Vocational Assessment

Volume III contains reviews of published instruments related to vocational assessment -- interest inventories, measures of attitudes, values, knowledge, and skills including work samples. Most of the instruments focus on the adolescent.

Volume IV Assessment of School Achievement

Volume IV focuses on the assessment of achievement of academic subjects and how students learn in elementary and secondary schools. Tests in languages other than English are included in this volume.

Volume V Communication/Language Assessment and Tests of Language Proficiency

Volume V consists of reviews of standardized tests that are intended to measure communication and language skills in English and other languages, primarily Spanish. Tests of English proficiency, language dominance, and proficiency in languages other than English are also reviewed.

Volume VI Assessment of Adaptive Behavior

Volume VI consists of reviews of standardized instruments that are intended to measure nonacademic behaviors including adaptive living skills.

Planned volumes include:

Volume VIII	Tests for Severely Impaired Students and Assessment of Psychomotor Abilities
Volume IX	Alternative Assessment Techniques
Volume X	Observation Techniques

Until these volumes are published, use the updated 1987 edition of the *Test Resource Guide* for tests in categories not recently reviewed.

It should be noted that tests included in the *Test Resource Guide* are not necessarily recommended for use in the assessment of New York City Public School students. A listing indicates only that the instrument has been reviewed. Users of this *Guide* are encouraged to read the entire review, attending, in particular, to the Special Alerts Section which provides information on recommended uses of the instrument. Tests in the *Guide* should be used consistent with guidelines in the Special Alerts/Comments section. Tests that are not in the *Guide* should be selected on the basis of information offered in Volume I and the test manual and their use should be consistent with the manual and their scores interpreted with care.

To maintain the timeliness of the *Test Resource Guide*, additional test reviews are produced and distributed in an addendum with an updated index. Addendum reviews should be placed in their appropriate volume and the newest index, always dated, used to locate a specific instrument. Addenda that have been published include an Addendum to Volume II which has been incorporated in the 1994 edition, a re-printing, and a 1997 Addendum containing reviews of tests for each volume that has already been published, i.e., II, III, IV, and V.

Copies of the *Test Resource Guide* are available from CSEs and the Division of Student Support Services, NYC BOE.

Volume VII, *Assessment of Cognition and Affect*. This volume consists of

- an Introduction which discusses concepts relevant to assessing the domains of cognition and affect;
- reviews of congnitive measures presented in alphabetical order;
- reviews of affective neasures presented in alphabetical order
- an Appendix which includes a Glossary, Updating Information, and a List of Publishers;
- a Selected Bibliography;
- two indexes: (1) an alphabetical listing of tests reviewed in this volume and (2) a listing of all tests reviewed in the first seven volumes in the *Test Resource Guide* series.

Some instruments assessing cognitive and affective behavior have been included in volume II, Preschool Assessment. Reviews of those instruments included in volume VII address the school age population only. Tests for kindergartners and preschoolers that do not go beyond these two levels are not included. *Test Resource Guide* users are advised to see Volume II for tests focusing on these populations.

INTRODUCTION

Volume VII of the *Test Resource Guide* consists of reviews of standardized measures of cognition and affect that may be used to provide information for determining eligibility for special education and for planning the education of students who have an educational disability. The instruments are not necessarily direct measures of educational performance and needs, but they provide information about the student to help the assessment professional recommend appropriate placement and instructional interventions. Although cognitive and affective assessments describe a student's present behavior, most of the instruments are predictive in nature; i.e., they are intended to provide information about future behavior.

ORGANIZATION OF VOLUME VII ASSESSMENT OF COGNITION AND AFFECT

This volume is divided into two sections, each preceded by a brief introduction describing the kinds of instruments and procedures used to assess the domain. The first section, Cognition, consists of reviews of cognitive instruments, The second, Affect, consists of reviews of affective measures. The instruments reviewed in each section are arranged in alphabetical order by title.

Each review consists of the author's statement of the purpose of the instrument, a description including the various scales, administration and scoring, a review of its psychometric properties, and a section entitled Special Alerts/Comments in which the assessment professional is advised of the appropriateness of the test for special populations and the use of resulting scores. Most of the comments in this section reflect the experience of New York City assessment professionals who have reviewed and used many of these instruments.

Inclusion in the *Test Resource Guide* does NOT signify approval. A review simply means that a test specialist and assessment professional have evaluated the instrument. Users of the *Guide* should read the entire review focusing on the Special Alerts Section which provides information on recommended uses of the instrument. Instruments in the *Guide* should be used consistent with guidelines in the Special Alerts/Comments section. Those that are not in the *Guide* should be used consistent with the Manual and interpreted with care. Note that information about unreviewed instruments should be sent to the Division of Student Support Services as per updating instructions in the Appendix.

TESTING IN THE ASSESSMENT PROCESS

The exploration of the cognitive and affective domains relies to a great extent on the integration of standardized test results with clinical judgment. The interpretation of assessment results requires extensive knowledge about human behavior and a solid understanding of psychometrics.

Psychological tests cannot be properly applied outside the context of psychology. Familiarity with relevant behavioral research and psychometrics as well as intensive training and supervised experience are needed (Anastasi & Urbina, 1997). Perhaps in no other area of psychoedcational assessment is this caution so necessary as in the affective and cognitive domains.

Many tests in these domains are described as "C Level" instruments which demand specific user qualifications for purchase such as a "PhD-level degree in psychology or education or the equivalent in a related field with relevant training in assessment or verification of licensure or certification by an agency" (Psychological Corporation, 1998, 56) requiring training and experience in assessment as outlined in the 1985 APA standards. Sometimes the doctoral level degree is omitted and user requirements cite only the completion of "a recognized graduate training program in psychology with appropriate coursework and supervised practical experience in the administration and interpretation of clinical assessment instruments" (AGS, 1998, 78). NYC does not support the use of tests by individuals not specifically trained for their administration and interpretation.

Constructs, concepts, and procedures relevant to the domain are discussed at the beginning of each test review section. See pages C1 to C4 for cognitive assessment and pages A1 to A3 for assessment of affect.

CHARACTERISTICS OF COGNITIVE AND AFFECTIVE ASSESSMENTS

There are overall standards applicable to assessment instruments which should be considered in selecting an instrument and interpreting results. Although many factors are related to these issues, most important are norms, reliability, and validity. Instruments should have representative norms, acceptable reliability for individual decision making, and validity for the purpose for which the test has been developed and is used. These measurement principles "have meaning and force outside of measurement wherever evaluative judgment and decisions are made" (Messick, 1994, 13). Additional characteristics, recommended by Ollendick and Hersen (1984), are a multimethod approach, empirical validation, and sensitivity to change.

Norms

Assessment professionals need to know about an instrument's norms to understand the publisher's assumptions about the standardization sample and the generalizability of the scores. Norms are usually described in detail by the publisher in the manual. Included are the number and nature of the students in the standardization sample, the types of schools in which they were enrolled (public/private, elementary/secondary), geographic area of the country, type of community (city/rural/suburban), socioeconomic status, racial/ethnic composition, and a comparison of these factors to U.S. census data. Sometimes information is given on the number of students with disabilities included in the sample and the nature of their disability. For measures in the affective domain, scores are sometimes based on testing clinical populations and samples; they should be completely described as indicated above.

In general, the larger the standardization sample, the greater the probability of ruling out the influence of chance factors on test scores. However, the quality of the sample is exceedingly important -- i.e., geographic distribution (New York City, New York State, northeast), socioeconomic status (not predominantly middle or upper class), racial/ethnic composition (inclusion of African-American, Latino, Asian students). For tests in languages other than English, the countries in which the test was developed and normed must be reported. If bilingual students are in the standardization sample, their language proficiency in English and in their native/home language and their country of origin must be included. Instruments that provide usable information should include subjects in the standardization sample with the same demographic characteristics as the student being assessed.

The point at which norms become outdated is more judgmental than empirical and depends in part on the skill being assessed (Salvia & Ysseldyke, 1995). Norms that are older than seven years may be unrelated to "today's" students in terms of school life and life styles. Norms older than fifteen years are clearly outdated. The concept of recency of norms is particularly relevant to cognition where some researchers have indicated large increases in intelligence test scores (Detterman & Thompson, 1997; Flynn, 1987; Williams, 1997).

Additional information about norms can be found in Volume I, Introduction of the *Test Resource Guide* and this volume's Glossary.

Reliability

Without reliability, a student could obtain a score one day and a widely variant one the next. We need to be reasonably sure that

the test yields a similar score each time, given the absence of instruction, illness, or other factors.

Reliability is related to the age of the test-taker. Young children's scores are generally much less reliable than those of older ones even on the best developed instruments. The larger the number of items in a scale, the greater the reliability because of the dependence of reliability on the length of the test. Reliability coefficients obtained from large numbers of diverse populations will tend to be higher than those calculated for one grade.

In reviewing the reliability of cognitive and affective instruments, the assessment professional should be especially concerned with test-retest and interrater reliability. The first, relating to stability, helps the assessment professional apply test findings to the long term. On the other hand, interrater reliability requires training for test administrators so that similar findings are reported for an individual whose behavior is being observed by different test administrators on the same measure. Assessment professionals should note that most test publishers report internal consistency as the sole measure of reliability. Alone it is insufficient for determining the reliability of an instrument.

Reliabilities for cognitive instruments are generally high. Empirical data have also shown them to be consistent across many demographic variables such as race, sex, and socioeconomic status (Kaufman & Reynolds, 1984, 205). Affective measures, on the other hand, seldom have high reliability coefficients, but these are also generally consistent across race, sex, and socioeconomic status.

The standard error of measurement, which enables the practitioner to establish a range of points around the reported score within which the "true" score lies, is also useful for analysis and interpretation of scores. A "true" score cannot be obtained in calculating test scores because of errors of measurement; the aim of those who develop tests is to minimize the errors and realize a narrow range of scores around the obtained score.

Assessment instruments with reliability coefficients of .90 or higher are deemed adequate for decision making for individuals. Lower coefficients preclude use of scores for interpreting an individual student's cognitive or affective behavior; such instruments should be used only qualitatively.

More detailed discussions of reliability can be found in Volume I, Introduction of the *Test Resource Guide* and in the Glossary of this volume.

Validity

Validity is the "appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores" (AERA, 1985,9). Guiding principles of validity include: (1) the concept of validity as a non-unitary construct; (2) the need to validate the uses and interpretations of test results rather that the instrument itself; (3) the definition of validity as a matter of degree rather than an all-or-none judgment; and (4) the need for multiple types of evidence regarding the validity of a particular use or interpretation (Linn, 1994). Generally validity is "meaningful only as it pertains to the *particular* use for which the test results are intended" (Worthen, Borg, & White, 1993, 178). It is clear, therefore, that the assessment professional must have a definitive reason for using the instrument and relate its reported validity for that purpose. Although "the burden of validating tests still rests upon the test publisher, ...the person making the inference on the basis of test scores, is responsible for knowing the validity of the use" (Geisinger, 1990, 10).

The various types of validity -- content, concurrent, predictive, construct, consequential -- are all important for cognitive and affective instruments. Least important is *content validity* which refers to the relationship of the test items to the subject being tested. In the domains of cognition and affect, content is most closely related to constructs being measured. ***Construct validity*** is dependent upon the user's definition of what is being measured.

The impact of the test, i.e. ***consequential validity***, should be considered in selecting an instrument. Assessment professionals should be aware of how the results affect different groups of students, particularly in regard to race/ethnicity, socioeconomic status, gender, language, or disability. "Appraisals ... of the potential social consequences of a proposed [instrument's] use and of the actual consequences when used, provide a consequential basis for test use" (Messick, 1984. 231).

Concurrent and predictive validity, also called criterion related validity, are statistical approaches in which the test is related to a criterion, usually another test. Predictive validity is most important in cognitive or performance measures because these instruments are used to help determine future status of an individual assuming no further intervention. "Intelligence tests are, outside of achievement tests, the best available predictors of academic achievement. Measures of general intelligence also turn out to be very good predictors of success in most job training and vocational training programs. ... Intelligence tests also predict ... other criteria" (Kaufman & Reynolds, 1984, 205) such as success in psychotherapy, rehabilitative success with traumatic brain injury, etc. Assessment professionals should

consider the criterion measure as well as the magnitude of the resulting coefficient of correlation in interpreting the predictive or concurrent validity of an instrument.

Detailed discussions of validity can be found in Volume I, Introduction, of the *Test Resource Guide* and this volume's Glossary.

Test Bias

Court cases on the use of intelligence tests in student assessment in the 1970's and early 1980's resulted in contradictory and inconsistent decisions (MacMillan & Balow, 1991). In Larry P. v. Riley, Judge Peckham indicated that the individually administered intelligence test was biased against African-American students. After ruling that these tests were prohibited in the decision making process for special education for African-American students, he allowed that they may be used for this same ethnic group in qualifying for gifted and talented programs. Cases in federal circuits in states other than California found against the African-American plaintiffs in favor of the use of these instruments.

Standardized instruments are generally sensitive to gender, ethnicity, socioeconomic class, and linguistic and cultural differences (Kaufman & Reynolds, 1984). Consequently test bias studies are infrequently undertaken. However, Suzuki and Valencia (1997) reported that test bias studies among Puerto Rican, Asian-American, and Native American students have rarely been undertaken.

Nonetheless, test development procedures by the major test publishers include both statistical and nonstatistical approaches to the study of test bias. Frequently minorities are over-sampled to avoid test bias and to provide greater representation of these groups in the standardization sample. As a result, standardized instruments generally do not systematically discriminate against any specific group (Figueroa, 1991). Low scores on a cognitive measure may indicate lack of experience rather than systematic bias (Thorndike, 1997).

The assessment professional needs to recognize the student's educational and cultural experiences when selecting cognitive or affective measures and interpreting results to avoid test bias. Bias can inadvertently intrude into the assessment process for students of diverse linguistic and cultural backgrounds. Criteria such as the number of years the student has spent in school, not only in an English language school, are critical in interpreting test results for that student.

See Volume I, Introduction, of the *Test Resource Guide* for additional information.

Use of Tests with Second Language Learners and Culturally Diverse Populations

Even on an unbiased test, the interpretation of the performance of students from culturally and linguistically diverse populations is fraught with problems. Frequently it is recommended that cognitive tests which are primarily nonverbal be used so that minimal emphasis is placed on English language skills. Naturally it is virtually impossible to create a nonverbal affective measure.

The standardization samples for most cognitive and affective instruments have experiential backgrounds that differ significantly from most culturally and linguistically diverse populations. This means that it is inappropriate to use information from the normative sample to calculate scores. Since scores cannot be reported, the interpretation of the student's performance must be qualitative and descriptive, and even then caution must be observed (Greenfield, 1997). Such an analysis should take into consideration the degree of acculturation of the student and family and the number of years the student has spent in an English language school system and type of program (e.g. ESL, bilingual, monolingual). How the assessment professional uses a test helps determine whether a test in English, the second language, can be used (Geisinger, 1992).

Results on cognitive and affective measures for linguistically and culturally diverse students should not be used for predicting future performance (Figueroa, 1991; Gronlund & Linn, 1980; Mehrens & Lehman, 1987). Sometimes they may be useful for providing information about current functioning. Qualitative reporting of results is generally preferable. A qualitative understanding of the student's learning style, strategies, and work habits is essential in an appropriate comprehensive report. Assessment professionals should determine the ability of the student in the native/home language and in English before interpreting test results.

Test Administration

When the assessment professional is planning to report scores, STANDARDIZED ADMINISTRATION PROCEDURES MUST BE FOLLOWED. Most norm referenced tests (NRTs) are accompanied by instructions in the manual which clearly delineate examiner behavior in administering the instrument. If such instructions are not followed, score interpretations are not meaningful. Any modification of the testing situation requires that test results be viewed and reported only qualitatively.

Although some cognitive and affective instruments may be administered by persons other than psychologists, most require scoring and interpretation of results by psychologists or persons

who have had extensive, in-depth training in individual psychological testing and personality and behavioral studies.

Test Scores

Standard scores and percentiles are most commonly used metrics for cognitive and affective assessments. Cognitive instrument results are generally reported as standard scores known as IQs, although other terminology to indicate a composite score may be used. These are on a scale with a mean of 100 and a standard deviation of 15 (Wechsler scales) or 16 (Stanford Binet scores). Scores on such scales are not necessarily reflective of general ability unless the test author has indicated this purpose in writing in the test manual.

Intelligence was deemed a function of heredity (Plomen, DeFries, McClearn, & Rutter, 1997; Plomen, Fulker, Corley, & DeFries, 1997), and any change over time was ascribed to limitations of the test. Extensive research on the stability of intelligence test performance has shown that the IQ is "quite stable" over elementary, secondary school, and college (Anastasi & Urbina, 1997, 324). Nevertheless, decades of research on IQ have documented that scores of intelligence, i.e., IQs, can change rather dramatically as a result of environment (Anastasi & Urbina, 1997; Ceci, 1991). Thus, an IQ, as other scores, should be interpreted in light of the many different aspects of student behavior and with the expectation that it can be modified rather than representing a trait immutable to change.

Functional level is an inappropriate concept for interpreting cognitive and affective scores. Similarly age and grade levels are inadequate and irrelevant for these constructs.

A detailed discussion of test scores is found in volume I, Introduction of the *Test Resource Guide*.

Use of Test Results

The use of results from cognitive and affective assessments should be guided by the concepts and ideas presented fully in Volume I, Introduction. Assessment professionals are aware that test scores indicate performance at the time of testing in relation to the test norms; they must also remember that the scores do not indicate why the performance occurred as it did. The reasons require determining the individual's "antecedent context" and "anticipated context" (Anastasi, 1992, p.2). Additionally assessment professionals should bear in mind that many environmental factors impact upon cognitive and affective assessment (Thorndike and Hagen, 1977). The following characteristics are examples of factors that should be investigated along with test performance: home language, values placed on book learning, family stability, separation/alienation

from the dominant culture, economic disadvantagement. The best use of test results is in an assessment package that includes multiple sources of information over time and across settings.

The usefulness of cognitive and affective assessments depends on the interpretations by psychologists who use the information to explain school and educational behaviors. Consequently, results from cognitive and affective testing should be presented in a comprehensive psychological report at conferences of parents, teachers, and supervisors in discussions of student learning and achievement and placement and instructional planning.

Computer-generated reports in addition to computerized scoring are available for some instruments. Assessment professionals should be aware that they are responsible for the accuracy of scoring and the accuracy, appropriateness, and relevance of reports even when they are generated by a computer program.

It should be noted that scores should not be reported for many students. Instruments that are unreliable are included in this restriction. Scores must not be reported for students for whom the normative sample is not representative, e.g. culturally and linguistically diverse students. Reporting scores for students with specific disabilities such as deafness or hard of hearing, and medical or physical disabilities should be done with caution and sensitivity. In these cases qualitative and descriptive information about student performance on the tasks covered by the test should take the place of a score. Assessment professionals are reminded that assessment tools that are not valid should not have been administered at all. In all instances, the student's behavior in classroom activities and other school programs is a better indicator of the student's performance than an unreliable, invalid score from a test with unrepresentative norms.

Each test review includes information on score reporting in the Special Alerts section.

REFERENCES

AGS. (1998) *Clinical catalog*. Circle Pines, MN: Author, 78.

American Educational Research Association, American Psychological Association, & National Council of Measurement in Education. (1985). *Standards of educational and psychological testing*. Washington, DC: APA.

Anastasi, A. (1992). Introductory remarks. In K.F. Geisinger (Ed.), *Psychological testing of Hispanics* (1-15). Washington, D.C.: APA.

Anastasi, A., & Urbina S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.

TITLE *Differential Ability Scales (DAS)* **AGE** *2-6 to 17-11 yrs.*
TIME *35-85 mins.*
AUTHOR *C.D. Elliot* **TYPE** *NRT*
PUBLISHER *Psychological Corporation*
COPYRIGHT DATE *1990* **SCORES** *Standard scores, %iles, GEs, AEs*

The focus of this review is on the school-age population.

PURPOSE

The DAS is intended to assess cognition for classification and diagnosis. Results provide a profile of specific strengths and weaknesses.

DESCRIPTION

The DAS is an individually administered test that contains both ability and achievement subtests. The cognitive battery includes 17 subtests grouped into two overlapping levels - preschool and school. Not all subtests are administered to every student. A general cognitive ability score (GCA), based on six core subtests at each level, is reported. In addition, cluster scores for verbal and nonverbal ability may be obtained.

The GCA is an estimate of reasoning and conceptual abilities which reflects Spearman's g. Other scores are cluster, specific ability, and diagnostic.

NORMS

The standardization sample included 3475 children, 175 at each half year interval 2-6 to 4-11 and 200 cases each year for ages 5 to 17-11. The sample was stratified on age, sex, race/ethnicity, parental education, and geographic region to match 1988 Census data. Students with learning disabilities, speech and language impairments, mental retardation, emotional disturbance, and mild visual, hearing, or motor impairments were included. Item bias was examined through the oversampling of ethnic groups. NYC students participated in the standardization, which occurred in 1988-1989. The norms are still relevant but by the year 2000 will be severely outdated.

RELIABILITY

Average internal consistency reliabilities of subtests ranged from .70 to .92. GCA values ranged from .90 to .95 across levels. Test-retest reliability over a two- to seven-week interval varied between .85 and .92 for GCA among three age groups of students, approximately 100 per group. Gridley and McIntosh (1992) indicated that the reliability estimates were generally of high enough magnitude to make possible interpretations at a number of levels. They also reported that DAS subtest reliabilities were superior to those in other batteries. SEMs for GCA standard scores were reported between 3.25 and 4.63.

VALIDITY

Content validation was supported by test development procedures and a series of confirmatory and exploratory factor analyses which lent support to the internal structure of the DAS (Keith, 1990). The hierarchical model was presented as evidence of construct validity (Elliot, 1990). Factor analyses of the DAS, WJ-R, and DTLA-3 on results from testing 100 third, fourth, and fifth graders indicated very strong factors structures for a one-factor, two-factor, and three-factor model for the DAS as well as on the seven-factor *Gf-Gc* model (McGhee, 1993).

Concurrent validity of the cognitive components of the DAS was explored by correlation between the GCA and corresponding scores of the K-ABC and Wechsler scales. In general the coefficients were moderately strong to strong for younger students (.67 to .89 up to age seven) and higher for older students (.84 to .91 for 8- to 15-year-olds). Some studies reported lower composite scores for the DAS than with the other instruments (Platt, Kamphaus, Keltgen, & Gilliland, 1991). However, the technical manual reported studies with students with disabilities which indicated scores "as expected" (p.276).

The central construct underlying the DAS is problem solving. The GCA is an exceptionally strong measure of *g* (McGhee, 1993; Platt, Kamphaus, Keltgen, & Gilliland, 1991).

SPECIAL ALERTS/COMMENTS

The DAS provides an overall indication of intellectual abilities which may be useful in determining intellectual/cognitive functioning from preschool through high school. Rather than being a downward extension of an adult intelligence test, the DAS was created from a developmental and educational perspective which resulted in a child-centered and age appropriate instrument (Gridley & McIntosh, 1992).

The Manual includes a comprehensive review of psychometric theory and ability testing including its history, use, and, in particular, theoretical constructs.

The DAS is a C level test.

Scores as low as 75 can be obtained with the DAS; this is lower than those yielded by other intelligence tests (Platt, Kamphaus, Keltgen, & Gilliland, 1991).

Item bias has been avoided to a large extent through the work of a Bias Review Panel and statistical analyses. However, composite scores for cognitive development, including the GCA "contain verbal subtests are therefore appropriate only for students who are proficient in English (Manual, p.37). Therefore, the DAS should not be used with LEP students and scores should nor be reported for them. Assessment professionals should be aware, however, of a Special Nonverbal Scale that is available for LEP students or those with disabilities which warrant the use of a nonverbal scale; results from such testing are reported by a Special Nonverbal Composite score. Scores should be reported in ranges for linguistically and culturally diverse students.

The use of the DAS with students who are deaf or hard of hearing has not been demonstrated. Caution should be observed when using this instrument with these students.

Scores may be reported for students for whom the DAS is appropriate. Score ranges are recommended for students with linguistic and cultural diversity. Descriptive reporting is encouraged for students whose verbal abilities in English are limited.

REFERENCES

Manual

Gridley, B.E., & McIntosh, D.E. (1992). Test review. In D.J. Keyser & R.C. Sweetland (Eds.), *Test critiques* Vol. IX (167-183). Austin, TX: ProEd.

Keith, T.Z. (1990). Confirmation and hierarchical confirmatory analysis of the Differential Ability Scales. *Journal of Psychoeducational Assessment*, 8, 391-405.

McGhee, R. (1993). Fluid and crystallized intelligence: Confirmatory factor analyses of the Differential Abilities Scale, Detroit Tests of Learning Aptitude-3, and Woodcock-Johnson Psycho-Educational Battery-Revised. *Journal of Psychoeducational Assessment Monograph Series*, 20-38.

Platt, L.O., Kamphaus, R.W., Keltgen, J., & Gilliland, F. (1991). An overview and review of the Differential Ability Scales: Initial and current research findings. *Critiques of School Psychological Materials*, *29*, 271-277.

PUBLISHER'S ADDRESS	PUBLISHER'S TELEPHONE
555 Academic Court San Antonio, TX 78758-6897	800-228-0752 201-567-5124

TITLE	*Stanford-Binet Intelligence Scale - Fourth Edition (S-B IV)*	**AGE**	*2 to adult*
		TIME	*45-90 mins.*
		TYPE	*NRT*

AUTHOR *R.C. Thorndike, E.P. Hagen, & J.H. Satler*
PUBLISHER *Riverside Publishing Company*
COPYRIGHT DATE *1986* **SCORES** *Standard scores*

PURPOSE

The S-B IV is intended to assess the overall level of cognitive development as well as the pattern of this development.

DESCRIPTION

The S-B IV, an individually administered test, consists of 15 tests, each tapping a different cognitive skill. Appropriate age levels for each test have been determined. The four broad categories of the S-B IV are: Verbal Reasoning, Abstract/Visual Reasoning, Quantitative Reasoning, and Short-Term Memory. Scores can be generated for any of the four areas and individual subtests.

This edition of the S-B is a major revision of the 1900 test and, like earlier editions, it requires establishing basal and ceiling levels. Many of the same items are included. However, the age-scale format has been replaced by tests, each using a single item type. The evaluation of general reasoning ability or *g* has been retained, but it is now labeled Standard Age Score (SAS) rather than IQ. Three other second-level scores are available: crystallized abilities, fluid-analytic abilities, and short-term memory. Although the change to a point scale was seen as a major revision, Anastasi (1989) commented that continuity has been maintained through the retention of many item types and adaptive testing procedures.

The advantages of the S-B IV are: broader coverage of cognitive skills and information processing abilities, more flexibility in test administration, higher reliability and test-score precision, and a more detailed diagnostic assessment of cognitive and information processing skills.

A complete S-B IV consists of eight to thirteen subtests, depending upon chronological age and entry level, as determined by performance on the first subtest. This test is administered by a psychologist only.

NORMS

Norms were developed by testing 5013 students who were representative of the nation according to 1980 US census data. Deviations from these data were corrected by differential weighting techniques. The design for standardization was based on geographic region, community size, ethnic group, age, and gender. Standardization took place in 1985, more than ten years ago.

Assessment professionals should note that norms more than ten years old are considered out-of-date and not particularly useful for reporting scores without additional corroborating information.

RELIABILITY

Kuder-Richardson Formula 20 procedures were used to determine internal consistency of the S-B IV. Median reliabilities ranged from .73 to .92 for preschoolers and the school age population. The more reliable score is the Composite Standard Age Score. Therefore, it is recommended that this score be the primary source of information for making decisions. The next most reliable scores are the four area SASs.

Subtest scores do not generally meet accepted criteria for reliability and using scores for individual decision making. SEMs are available.

The sample for test-retest reliability included 112 students over a 2- to 8-month interval. For the Composite Score, test-retest reliability was estimated as .91 for 5-year-olds and .90 for 3-year-olds.

VALIDITY

Concurrent validity was strongly supported in studies correlating the S-B IV and the WPPSI, WISC-R, WAIS-R, and K-ABC.

Construct validity was established through factor analysis which confirmed the existence of the general ability factor, *g*, although the factor structure beyond *g* varied with age (Manual, 1986; Laurent, Swerdlik, & Ryburn, 1992). Confirmatory factor analytical studies generally supported the factor structure (Keith, Cool, Novak, & White, 1989; Keith, Cool, Novak, White, & Pottebaum, 1988; Owsby & Carmin, 1988).

SPECIAL ALERTS/COMMENTS

The S-B IV provides an overall indication of intellectual abilities which may be useful in determining intellectual/cognitive functioning in conjunction with other information, given the datedness of the standardization. The test may be useful for differentiating between students who have specific learning disabilities and those who have mental retardation through a comparison of individual or combinations of cognitive ability areas. However, scatter between subtests and/or areas must be significant; i.e., standard score differences must be greater than one standard deviation.

The SB-IV should only be administered, scored, and interpreted by a psychologist (Sabatino, 1989-1990). Even so, these are not easy tasks. Spruill (1987) indicated that practice was needed for administration expecially in the transition from subtest to subtest.

The SB-IV offers more than just an IQ because it provides much information about language comprehension and use (Sabatino, 1989-1990), which is useful for classroom planning. School psychologists have found that the S-B IV yields information useful for developing an IEP and for relating test results to classroom instruction.

It has been recommended that a subtest should never be dropped from estimating a total score unless the score is zero. However, Cronbach (1989) noted that it is extremely important "not to take seriously the numerical value of a low score (774)". An abbreviated battery is unacceptable for determining a level of cognitive/intellectual functioning in an initial referral; for this purpose a battery of eight to thirteen subtests must be administered to an individual student.

The S-B IV is not recommended for use with linguistically and culturally diverse students nor LEP students. Scores should not be reported for bilingual nor LEP students.

REFERENCES

Manual

Anastasi, A. (1989). Test review. In J.C. Conoley & J.J. Kramer (Eds.), *The tenth mental measurements yearbook* (771-773). Lincoln: NE: University of Nebraska.

Cronbach, L.J. (1989). Test review. In J.C. Conoley & J.J. Kramer (Eds.), *The tenth mental measurements yearbook* (773-775). Lincoln: NE: University of Nebraska.

Keith, T.Z., Cool, V.A., Novak, C.G., & White, L.J. (1989). *Hierarchical confirmatory analyses of the Stanford-Binet Fourth Edition: Testing the theory-test match.* New Orleans: AERA, paper presented at annual meeting.

Keith, T.Z., Cool, V.A., Novak, C.G., White, J.J., & Pottebaum, S.M. (1988). Confirmatory factor analysis of the Stanford-Binet Fourth Edition: Testing the theory-test match. *Journal of School Psychology, 26,* 253-274.

Laurent, J., Swerdlik, M., & Ryburn, M. (1992). Review of validity research on the Stanford-Binet Intelligence Scale: Fourth Edition. *Psychological Assessment, 4,* 102-112.

Owsby, R.I., & Carmin, C.E. (1988). Confirmatory factor analysis of Stanford-Binet Fourth Edition. *Journal of Psychoeducational Assessment, 6,* 331-340.

Sabatino, D.A. (1989-1990). Stanford-Binet Intelligence Scale Fourth Edition (SB-IV). *Diagnostique, 15,* 183-198.

Spruill, J. (1987). Test review. In D.J. Keyser & R.C. Sweetland (Eds.), *Test critiques* Volume VI (544-559). Kansas City, MO: Test Corporation of America.

PUBLISHER'S ADDRESS	PUBLISHER'S TELEPHONE
425 Spring Lake Drive Itasca, IL 60143-2079	800-323-9540 914-238-5094

TITLE	*Wechsler Intelligence Scale for Children - Third Edition (WISC-III)*	**AGE**	*6-16.11 yrs.*
		TIME	*50 - 90 mins.*
		TYPE	*NRT*

AUTHOR *D. Wechsler*
PUBLISHER *Psychological Corporation*
COPYRIGHT DATE *1991* **SCORES** *Standard scores, %iles, IQs*

PURPOSE

The WISC-III is intended to assess intellectual ability.

DESCRIPTION

The WISC-III is an individually administered test of intellectual ability which is a revision of the 1974 edition, the WISC-R. The principal goal of this revision was to update the norms. However, some changes in content were also made. "Items with clinical or provocative emotional content were consistently eliminated" (Kaufman, 1993, 347).

The WISC-III generates three composite scores -- Verbal Intelligence Quotient (VIQ), Performance Intelligence Quotient (PIQ), and Full Scale Intelligence Quotient (FSIQ) -- corresponding to verbal abilities, non-verbal performance abilities and overall abilities, respectively. Additionally, four indices of functioning are optionally computed to estimate verbal comprehension, perceptual organizational abilities, freedom from distractibility, and processing speed; these factors were obtained through factor analysis and are new in this edition.

The instrument consists of ten required subtests and three optional ones. Five required subtests plus one optional subtest compose the verbal portion of the WISC-III. These are Information, Arithmetic, Similarities, Comprehension, Vocabulary, and Digit Span; the latter is optional. Performance subtests include Picture Completion, Picture Arrangement, Block Design, Object Assembly, and Coding plus two optional subtests, Mazes, and Symbol Search.

All twelve subtests from the WISC-R were retained in the WISC-III. Symbol Search was added which can be substituted for Coding. Although all the subtests were retained in this new edition, all of them were modified or expanded. Modifications consisted primarily of enlarging pictures and adding color to visual stimuli. New items were added to some scales so that floors and ceiling could be extended. The order of subtest administration was also changed so that the first subtest to be administered is a performance task.

Some subtests require the use of a stop watch. Administration instructions, including starting and discontinuance procedures, must be followed exactly as indicated in the Manual to enable the reporting of scores. In comparison to the WISC-R, this new edtion is easier to administer and score (Braden, 1995). The WISC-III is administered by a psychologist only.

The standard score scale has remained the same with a mean of 100 and a standard deviation of 15.

NORMS

Stratified random sampling techniques were used to obtain a normative sample of 2200 students, 100 boys and 100 girls in each of 11 age groups, which closely parallels demographic information reported by the 1988 US Census Bureau. Factors for stratification included age, gender, race/ethnicity, geographic region, urban/rural classification, SES, and parent educational level. Students from both public and private schools

were included in the sample. Certified special education students as well as general education students were tested. However, no students who were deaf or blind were included in the sample.

Assessment professionals should note that the norms are more than ten years old and may be considered out-of-date and not particularly useful for reporting scores without additional corroborating information.

RELIABILITY

Internal consistency reliabilities, reported for all subtests except Coding and Symbol Search, were conducted using the split-half procedure, corrected by the Spearman-Brown formula. They ranged from .69 to .87 for individual subtests across ages which confirms the advice of many researchers who discourage the use of subtest scores in "profile analysis" (Glutting, McDermott, & Konold, 1997; Kamphaus, 1993; Kamphaus, Petoskey, & Morgan, 1997; Kaufman, 1997; Sattler, 1992). Use of subtest scores for individual decision making is inappropriate given the low reliabilities of subtests. Assessment professionals should note that this advisement was also presented for the earlier WISC-R (McDermott, Fantuzzo, & Glutting, 1990).

Internal consistency reliabilities, on the other hand, for the three composite scores ranged from .90 to .97, and the four factors from .80 to 95. These scores are more appropriate for individual decision making. SEMs averaged from 1.08 to 1.67 for subtests across ages, about four scale score points for the composites and five for the index scales.

Test-retest reliabilities calculated from scores obtained over a three-week period for three age groups were in the similar range of internal consistency reliabilities for both subtests and composites. Long term stability of global IQ and factor index scores was supported in a study (Canivez & Watkins, 1998) of more than 600 students over a six-month to six-year time span; fewer than half the subtests resulted in significant differences between first and second testing.

Nicholson (1992) reported that the reliability of the WISC-III is excellent.

VALIDITY

The Manual presents construct validity for the WISC-III based on the WISC-R. However, additional studies were conducted which supports the validity of this edition. Median correlations with measures of achievement and school grades ranged from upper .30's to low .70's. Correlations with the WPPSI-R and WAIS-R ranged from .70 to .90.

Both exploratory and confirmatory factor analyses were conducted which supported the two factor model of the Verbal and Performance Scales. Studies conducted with special populations indicated support for the factor structure for students with specific learning disabilities (Slate, 1997) and black special education students (Kush & Watkins, 1997). Other studies supported the two- and four-factor structure (Konold, Kush, & Canivez, 1997). However, the proposed four-factor model, which was also supported in the original data, is a contradiction to the two-factor model according to some researchers (Kamphaus, Benson, Hutchinson, & Platt, 1994). Using standardization sample data, they found that for six- and nine-year-olds there were little differences between the three- and four-factor models, but none of the models fit very well for other age groups. The factor, Freedom from Distractibility, was criticized by some researchers (Little, 1991; Gussin & Javorsky, 1995) as invalid because of low variance and the lack of discrimination between youth with and without ADHD who were in an acute-care psychiatric hospital. Others (Alfonso, Zgodny, Berdugo, & Gorman, 1998) also found lack of support for four factors for special populations.

The WISC-III index scores provided better prediction of WIAT achievement than the FSIQ (Konold, 1998). However, Carroll (1993) noted that hierarchical factor analyses did not support the composition, stability, or uniqueness of the indices. On the other hand, Braden (1995) wrote favorably about the development of

indices, but indicated that the number, relaibility, composition, and interpretation of subfactors are not, and may never be, fully resolved (1099)".

Bias detection procedures were undertaken during the test development process through ongoing review of differential item functioning and replacement or modification of flawed items and the employment of a bias review panel which reviewed item content and presentation.

SPECIAL ALERTS/COMMENTS

The WISC-III is one of the most widely used and accepted instruments to estimate cognitive or intellectual abilities among school age populations. This status reflects the excellence in psychometrics, test development, and the Manual (Kaufman, 1993, Sandoval, 1995). However, the norms are out-of-date and the test should be restandardized. Given the datedness of the norms, decisions of performance level should be based on additional assessment and educational information.

Special procedures are included in the Manual for assessing students with disabilities. Hishinuma (1995) noted that the WISC-III requires no reading except for some items at the end of the Arithmetic subtest and no written expression which will permit assessment of students with language disabilities. However, assessment professionals should be cautious in interpreting results of all subtests with students who are blind, deaf, hard of hearing, or who have language impairments. Assesment professionals also noted the general lack of information about selecting subtests for administration, applying modifications, if any, during test administration, and using suggested guidelines in the Manual for interpreting and reporting results for these students. No standardized instructions are included for assessing students who are deaf (Spragins, Biennerhassett, & Mullen, 1994).

Other comments made by NYC psychologists include the nonuniformity of scale scores, difficulty in interpreting scores when a substitute supplementary subtest was used for a standard subtest, difficulty in scoring some subtests, and practice effects for some items, especially those in Performance subtests. The limited range of scores at the lowest levels is still viewed as a major disadvantage. The elimination of emotional items was not seen unanimously as a positive change since the removal of disturbing material was viewed as a loss of clinical information which also does a disservice to the Wechsler construct of clinical value of a test over its resulting score (Kaufman, 1993).

The WISC-III performance subtests are recommended for use with students who are deaf as part of a "multi-factored evaluation" (Spragins, Biennerhassett, & Mullen, 1994, 3). However, Maller (1994, 2) found that deaf students are at a "distinct disadvantage" when they are given the WISC-III. NYC assessment professionals tend to emphasize subtests yielding a VIQ for the blind and performance subtests yielding a PIQ for the deaf and hard of hearing.

The WISC-III is a C level examination.

Cautious use of the instrument with culturally and linguistically diverse students is recommended. Kaufman (1993) suggested the use of the Performance subtests and Similarities and Digit Span for the assessment of students of different cultural and linguistic backgrounds. Scores should not be reported for these students or for students who are deaf, hard of hearing, or who have language impairments. For these students, qualitative reporting is recommended.

REFERENCES

Manual

Alfonso, V.C., Zgodny, A.R., Berdugo, H., & Gorman, B.S. (1998). *WISC-III confirmatory factor analysis of individuals referred for academic difficulties.* San Francisco: paper presented at APA annual convention.

Braden, J.P. (1995). Test review. In J.C. Conoley & F.C. Impara (Eds.), (1098-1103). *The twelfth mental measurements yearbook.* Lincoln, NE: University of Nebraska.

Canivez, G.L., & Watkins, M.W. (1998). Long term stability of the Wechsler Intelligence Scale for Children-Third Edition. *Psychological Assessment, 10,* 285-291.

Carroll, J.B. (1993). What abilities are measured by the WISC-III? In B.A. Bracken (Ed.), *Monograph series advances advances in psychoeducational assessment: Wechsler Intelligence Scale for Children: Third Edition; Journal of Psychoeducational Assesssment* (131-143). Brandon, VT: Clinical Psychology Publishing Co.

Glutting, J.J., McDermott, P.A., & Konold, T.R. (1997). Ontology, structure and diagnostic benefits of a normative subtest taxonomy from the WISC-III standardization sample. In D.P. Flanagan, J.L. Genshaft, & P.L. Harrison (Eds.), *Contemporary intellectual assessment: theories, tests, and issues* (349-372). NY: Guilford.

Gussin, B., & Javorsky, J. (1995). The utility of the WISC-III freedom from distractibility in the diagnosis of youth with attention deficit hyperactivity disorder in a psychiatric sample. *Diagnostique, 21,* 29-42.

Hishinuma, E.S. (1995). WISC-III accommodations: The need for practitioner guidelines. *Forum, 28,* 130-135.

Kamphaus, R.W. (1993). *Clinical assessment of children's intelligence.* Boston: Allyn & Bacon.

Kamphaus, R.W., Benson, J., Hutchinson, S., & Platt, L.O. (1994). Identification of factor models for the WISC-III. *Educational and Psychological Measurement, 54,* 174-186.

Kamphaus, R.W., Petoskey, M.D., & Morgan, A.W. (1997). A history of intelligence test interpretation. In D.P. Flanagan, J.L. Genshaft, & P.L. Harrison (Eds.), (32-47). *Contemporary intellectual assessment: theories, tests, and issues.* NY: Guilford.

Kaufman, A.S. (1993). King WISC the Third assumes the throne. *Journal of School Psychology, 31,* 345-354.

Kaufman, A.S. (1994). *Intelligence testing with the WISC-III.* NY: Wiley

Konold, T.R. (1998). *Evaluating discrepancy analyses with the WISC-III and WIAT.* San Francisco: paper presented at APA annual convention.

Konold, T.R., Kush, J.C., & Canivez, G.L. (1997). Factor replication of the WISC-III in three independent samples of children receiving special education. *Journal of Psychoeducational Assessment, 15,* 123-137.

Kush, J.C., & Watkins, M.W. (1997). Construct validity of the WISC-III verbal and performance factors for black special education students. *Assessment, 4,* 297-304.

Little, S.G. (1991). Is the WISC-III factor structure valid? *Communique,* October.

Maller, S.J. (1994). *Item bias in the WISC-III with deaf children.* New Orleans: paper presented at AERA annual convention.

McDermott, P.A., Fantuzzo, JW., & Glutting, J.J. (1990). When asked say no to subtest analysis: A critique on Wechsler theory and practice. *Journal of Psychometric Assessment, 8,* 290-302.

Nicholson, C. (Winter 1992). New tests. *CEDS Communique, 19,* (1), 7.

Sandoval, J. (1995) Test review. In J.C. Conoley & F.C. Impara (Eds.), (1103-1104). *The twelfth mental measurements yearbook.* Lincoln, NE: University of Nebraska.

Sattler, J.M. (1992). *Assessment of children: WISC-III and WPPSI-R supplement.* San Diego, CA: Author.

Slate, J. (1997). Differences in WISC-III scores for boys and girls with specific learning disabilities. *Diagnostique, 22,* 133-145.

Spragins, A.B., Biennerhassett, L., & Mullen, Y. (1994). *Cognitive assessment instruments used with deaf students* (1994 Update). Washington, DC: Gallaudet University.

PUBLISHER'S ADDRESS	PUBLISHER'S TELEPHONE
555 Academic Court San Antonio, TX 78204	800-226-0752 201-567-5124 512-299-1061

TITLE	*Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R)*	**AGE**	*3 to 7-3 yrs.*
		TIME	*50-75 mins.*
		TYPE	*NRT*

AUTHOR *D. Wechsler*
PUBLISHER *Psychological Corporation*
COPYRIGHT DATE *1989* **SCORES** *Scale scores, %iles*

PURPOSE

The WPPSI-R is intended to measure the intellectual ability of young children in school.

DESCRIPTION

The WPPSI-R consists of six verbal tests and six performance tests which are administered individually.

Verbal	Performance
Information	Object Assembly
Vocabulary	Geometric Design
Arithmetic	Block Design
Similarities	Mazes
Comprehension	Picture Completion
Sentences	Animal Pegs

Five of the verbal tests are used to compute a Verbal IQ (VIQ), which is based on spoken responses. Five of the six performance tests are used to compute a Performance IQ (PIQ), which is based on the perceptual motor responses of pointing, placing, and drawing. The VIQ and the PIQ are summed to yield the Full Scale IQ (FSIQ).

The verbal and performance tests are alternated in administration.

NORMS

The normative sample consisted of 1700 children, 100 boys and 100 girls in eight half-yearly age groups ranging from three to seven. The sample was stratified according to the 1986 Census survey. Comparison of the demographics of the standardization sample and the 1986 Census survey demonstrates that the sample was representative of the US population. The norms are beyond the five to seven year period considered recent.

RELIABILITY

Split-half reliabilities for PIQ, VIQ, and FSIQ were all above .90 except for the VIQ at age 3 (which was .70). For the tests, the lowest reliability was .63 for the Object Assembly. With the exception of Mazes (.77) and Geometric Design (.79), the remaining split-half reliabilities were above .80. The test-retest reliabilities were .88 (PIQ), .90 (VIQ), and .91 (FSIQ), which is supportive of using scores in decision making for individuals.

VALIDITY

The Manual contains a review of studies with the WPPSI to indicate construct, concurrent, and predictive validity. Construct validity for the WPPSI-R was examined with intercorrelations among the tests and factor analysis. In general, the results of both show evidence of convergent and discriminant validity and support the two-factor approach to intelligence.

Concurrent validity information was presented by means of comparisons with the WISC-R, the SB-IV, the McCarthy, and the K-ABC. The WPPSI-R performance in four identified samples (gifted, mentally deficient, learning disabled, and speech and language impaired) also provided evidence of concurrent validity.

SPECIAL ALERTS/COMMENTS

Given the datedness of the standardization sample, the WPPSI-R information is not sufficient for determining intellectual/cognitive functioning. Given the overlap in age range with the WISC-III, the latter instrument is recommended for administration at the upper level of the age range.

The WPPSI-R is a C level test. Test administrators and scorers should be knowledgeable about testing and work under the supervision of a well-trained professional in psychological assessment. The latter are the only individuals who may interpret test scores.

The WPPSI-R is not recommended for students with severe visual or auditory impairments or significant physical impairments or any situation where the student would be disadvantaged if the test were administered in the standardized manner. Although means for accommodating the impairment might be devised, there is no empirical support for the validity or reliability of scores obtained in nonstandardized administrations.

Given the lack of predictive validity studies for the WPPSI-R, scores should be interpreted cautiously.

The WPPSI-R is not recommended for use with linguistically and culturally diverse children and LEP children.

Scores should not be reported for any students given the datedness of the norms.

REFERENCES

Manual

PUBLISHER'S ADDRESS

555 Academic Court
San Antonio, TX 78204-2498

PUBLISHER'S TELEPHONE

(800) 228-0752
(512) 299-1061
(203) 567-5124

TITLE *Woodcock-Johnson Psychoeducational Battery - Revised Tests of Cognitive Ability (WJ-R COG)*

AGE *2-95 yrs.*
TIME *Untimed (20-40 min)*
TYPE *NRT*

AUTHOR *R. W. Woodcock & M.B. Johnson*
PUBLISHER *Riverside Publishing Company*
COPYRIGHT DATE *1989* **SCORES** *Age/grade equivalents, %iles, standard scores, functional levels, NCE*

The focus of this review is on the use of the cognitive tests only with preschool and school age populations.

PURPOSE

The WJ-R COG is intended to measure cognitive functions.

DESCRIPTION

The WJ-R, an expanded and revised edition of the 1977 of the WJ, is theoretically grounded in *Gf-Gc* theory, usually attributed to the Horn-Cattell model of intelligence. It consists of 21 subtests, all administered individually. The tests measure seven factors: long-term retrieval, short-term memory, processing speed, auditory processing, visual processing, comprehension-knowledge, and fluid reasoning.

Only seven tests are part of the Standard Battery, recommended for use with all students. These are: Memory for Names, Memory for Sentences, Visual Matching, Incomplete Words, Visual Closure, Picture Vocabulary, and Analysis-Synthesis. The remaining tests are in the Supplemental Battery and recommended for use as needed. They are: Visual-Auditory Learning, Memory for Words, Cross Out, Sound Blending, Picture Recognition, Oral Vocabulary, Concept Formation, Delayed Recall - Memory for Names, Delayed Recall-Visual-Auditory Learning, Numbers Reversed, Sound Patterns, Spatial relations, Listening Comprehension, and Verbal Analogies.

NORMS

Normative data were based on 705 preschoolers and 3245 school age subjects in kindergarten to 12th grade. A stratified random sampling procedure was used. Data were collected between September 1986 and August 1988 from more than 100 geographically diverse US communities. The standardization sample did not include students with severe disabilities unless they were enrolled at least part-time in mainstream education, nor did it include those with less than one year of experience in an English-speaking environment. No preschoolers or school-age subjects were tested in NYS.

RELIABILITY

Reliability estimates of Standard Battery subtests for the preschool sample, based on 2- and 4-year-olds varied from a low of ..83 to a high of .94 and for the school age sample, based on 5-. 6-, 9-, 13-, and 18-year-olds, from .60 to .93. Similar results were reported for the Supplemental Battery subtests,

Broad Cognitive Ability reliabilities were reported as above .90 for the same ages.

VALIDITY

Content validity was based on item validity studies and expert opinion. The test authors noted that the free-response format parallels real-life cognitive performance demands.

The authors cited evidence of construct validity by (1) higher correlations within the seven cognitive factors, (2) the low intercorrelations among factors, and (3) the increase in scores by age

In terms of concurrent validity, for preschoolers cognitive tests were correlated with popular measures of intelligence. Coefficients were generally in the .50s and .60s. Similar results were found for the school age population with correlations with other instruments.

SPECIAL ALERTS/COMMENTS

The WJ-R COG provides information about cognitive functions which is insufficient for making performance level decisions in this area, primarily because of the high verbal loading of the items. However, the datedness of the norms also precludes using the WJ-R COG scores for individual decision making.

Use of the WJ-R COG is based upon the APA's *Standards for Educational and Psychological Testing* (Riverside Publishing. 1997) which demands use of instruments like the WJ-R COG by assessment professionals who have completed formal, graduate-level practicum-type courses in the assessment of individually administered tests of cognitive ability. In NYC only psychologists are permitted to administer, score, and interpret these examinations.

McGhee and Buckhalt (undated, 149) indicated that users of the WJ-R COG "can feel confident that the norms, reliability, and validity characteristics are comparable to, or better than, those of other measures of intellectual ability."

Use of the WJ-R COG has not been demonstrated with students who are deaf or hard of hearing; in fact NYC assessment professionals reported heavy emphasis on auditory processing in many subtests. Cautious use with these populations is strongly advised for these students as well as students with language impairment.

The WJ-R COG should be used cautiously with students of diverse cultural and linguistic backgrounds.

Scores should not be reported for any students. The WJ-R COG should be used as a qualitative and descriptive tool only.

REFERENCES

Manuals

McGhee, R.L., & Buckhalt, J.A. Test review. *Journal of Psychoeducational Assessment:* WJ-R Monograph. 136-150.

Riverside Publishers. (1997). *Test User Qualifications for the WJ-R Tests of Cognitive Ability.* Itasca, IL: Riverside, Fax Communication.

Woodcock, R.W. (1990). Theoretical foundations f the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment. 8,* 231-258.

PUBLISHER'S ADDRESS	PUBLISHER'S TELEPHONE
425 Spring Lake Drive Itasca, IL 60143-2079	800-323-9540 914-747-9883

TITLE *Vineland Social-Emotional Early Childhood Scales (SEEC)* **AGE** *Birth to 5-11 mos.*
TIME *20-30 mins.*
TYPE *NRT,CRT*

AUTHOR *S. S. Sparrow, D. A. Balla, & D. V. Cicchetti*
PUBLISHER *American Guidance Service*
COPYRIGHT DATE *1998* **SCORES** *%iles, NCE, Age equivalents, stanines*

PURPOSE

The SEEC is intended to assess social and emotional functioning of young children.

DESCRIPTION

The SEEC was developed from the Socialization Domain of the Vineland Adaptive Behavior Scales (VABS) developed in the 1980's. The original scale consisted of 131 items; this scale appropriate for young children only consists of 123 items. A Social-Emotional composite as well as three scale scores, Interpersonal Relationships, Play and Leisure Time, and Coping Skills are reported in a score profile. Coping Skills items are not administered for children under one year of age. Ratings are based upon the presence or absence of the behavior relevant to the scale items. Both norm referenced and criterion referenced information are available on all forms.

Instructions for determining a starting point (basal) for the SEEC are provided in the Manual. There is no ceiling; all items are administered and scored thereafter.

NORMS

Despite the recent publication date, standardization was conducted between September 1981 and May 1982 at 35 sites in 24 states. The sample was stratified by age, sex, ethnic group, community size, geographic region, and parental education level and corresponded to 1980 US Census figures. SEEC norms are based on 1200 children from birth through five years 11 months. Norms consist of data for 200 children in six age groups.

Assessment professionals should be aware that the norms are seriously out of date.

RELIABILITY

Internal consistency for the composite score ranged between .89 and .96. Scale scores were considerably lower. Median estimates ranged from .80 to .87. Test-retest reliability resulted in exceedingly low reliability coefficients; neither scale nor composite coefficients exceeded .79 in any of the studies that were conducted. SEMs are available.

Although reliability for tests used with young children tend to be low, use of the SEEC scales for individual decision making should be done with great caution. Based solely on reliability, individual decisions using the composite score may generally be made.

VALIDITY

Construct validity was developed through an indication of the relationship of scores to age for the standardization sample. Content validity was developed through the test author's definition of adaptive behavior. No validity studies were reported for SEEC scales and composite.

SPECIAL ALERTS/COMMENTS

The SEEC is a re-packaging of older material which is intended to provide a structured interview procedure for collecting information on a child's socialization which may be helpful in obtaining descriptive information.

Caution should be observed in making decisions about individuals because of inadequate reliability estimates and limitations of the standardization sample which tended to underrepresent low SES children (Reynolds, 1987). Additionally, the datedness of the standardization is a limiting factor. The information is sufficient for developing objectives on the IEP in the area of socialization skills. Even with other information on this domain, SEEC results are insufficient for making performance level decisions.

Following NYC policy, age equivalents should not be used or reported.

The SEEC may be appropriate for some children from culturally and linguistically diverse backgrounds, but attention must be given to the unique characteristics of specific cultures. The SEEC is not a bilingual test. Norms do not exist for bilingual populations. It should be noted that the onset of specific developmental milestones may appear at different ages in different cultures.

Scores may not be reported or used to determine performance level.

REFERENCES

Manual

Reynolds, C.R. (1987). Critiques of school psychological materials. *Journal of School Psychology, 25*, 97-100.

PUBLISHER'S ADDRESS

Publishers Building
Circle Pines, MN 55014-1796

PUBLISHER'S TELEPHONE

800-328-2560
516-623-0127