CHAPTER 4

ASSESSMENT OF CHILD INTELLIGENCE

Kristee A. Beres
Alan S. Kaufman
Mitchel D. Perlman

INTRODUCTION

Over the centuries, many definitions of intelligence have been postulated attempting to explain this elusive construct. This chapter provides a context in which to understand children's intelligence by including a chronology of historical landmarks, by expounding on popular intelligence measures, and by looking at future trends in intelligence testing. In the first part, a brief history of mental measurement is provided. The second part is partitioned into two subsections and describes tests currently available to measure children's intelligence (preschool through adolescence). The first subsection provides detail on five major intelligence tests for children: Wechsler Intelligence Scale for Children-Third Edition (WISC-III); Kaufman Assessment Battery for Children (K-ABC); Stanford-Binet: Fourth Edition (SB-IV); Kaufman Adolescent and Adult Intelligence Test (KAIT); and Woodcock-Johnson Psycho-Educational Battery-Revised: Tests of Cognitive Ability (WJ-R). The second subsection describes other popular tests of mental measurement. The final part of this chapter provides a sample case report that combines the WISC-III, WJ-R, and other tests in the assessment of a male adolescent with academic and emotional difficulties.

A CHRONOLOGY

Classifying and categorizing individuals is by no means a novel concept. From the beginning, so we are told, Adam was classified as being "man" and Eve was classified as being "woman." The fascinating but evasive concept of "intelligence," while called many things, is a concept that has been with us throughout time and has been used both positively and negatively to set mankind apart from beasts and to differentiate within the broad category of mankind itself. Measuring intelligence is a complex and historically sensitive issue that has often been misused.

Measuring intelligence emerged out of both theoretical interest and societal need. Theoretical interest relates to the desire to understand individual differences. Societal need involves the utilization of this understanding to solve practical problems. The evolution of the measurement of intelligence, therefore, did not emerge in a vacuum, but rather, by the interplay in the development of several paradigms: psychology, sociology, psychometrics, and law. The brief historical review that follows highlights landmark events in those areas and relates them to the birth and development of intelligence.

It is difficult to pinpoint when intelligence testing began, but certainly it was conceived prior to
the 1800s. Aristotle, for example, attempted to understand how people behaved by dividing mental functions into cognitive (cybernetic) and dynamic (orectic) categories. Cybernetic functions are thought processes; orectic functions are emotional and moral processes (Das, Kirby, & Jarman, 1979). Since the time of Aristotle, whenever personality is conceptualized, attempts are still made to keep separate these two functions, but the difficulty in doing so has been well recognized (Das et al., 1979; Perlman, 1986; Shapiro, 1965). Others, such as Firtzherbert in 1510, Huarte in 1575, Swinbourne in 1610, and Thomasius in 1692 gave credence to testing one’s cognition. They proposed various definitions of cognition and gathered information on human mentality (Sattler, 1988).

The 1800s ushered in several important advances in intelligence testing. In fact, interest in cognition and in the measurement of cognition in the 19th century was part of the scientific movement that brought psychology into being as a separate and respected discipline. Esquirol (1828) focused on distinguishing between mental retardation and emotional disturbance. It was he who coined terms such as imbecile and idiot to describe diverse levels of mental deficiency. He pointed out that idiots never developed their intellectual capacities, whereas mentally deranged persons lost abilities they once possessed. After studying different methods of measuring intelligence, he concluded that language usage was the most dependable criterion, a philosophy prevailing in most intelligence measures today.

Seguin’s (1907) philosophy was quite different. He stressed sensory discrimination and motor control as aspects of intelligence. The Seguin Form Board, which requires rapid placement of variously shaped blocks into their correct holes, is an application of his theory. Many of the procedures he developed were adopted or modified by later developers of performance and nonverbal tasks. It was during this time that intelligence testing and education had their first formal courtship, for Seguin convinced the authorities of the desirability of educating the “idiots” and “imbeciles.” He is credited for beginning the first school for the feebleminded and for being the author of the first standard book dealing with educating and treating them (Pintner, 1949). Also, his methods provided the inspiration for Maria Montessori’s approach to education.

Galton’s approach was similar to Seguin’s in that he also stressed discrimination and motor control. In accordance with his commitment to the notion that intelligence is displayed through the uses of the senses—sensory discrimination and sensory motor coordination—he believed that those with the highest IQ should also have the best discriminatory abilities. Therefore, he developed tasks such as weight discrimination, reaction time, strength of squeeze, and visual discrimination. Galton is credited both with establishing the first comprehensive individual intelligence test and with influencing two basic notions of intelligence: the idea that intelligence is a unitary construct (which eventually led others to postulate the notion of general intelligence, or the “g” factor), and that individual differences in intelligence are largely genetically determined (possibly influenced by the theory of his cousin, Charles Darwin) (Cronbach, 1970; Das, Kirby, & Jarman, 1979; Pintner, 1949). Perhaps Galton’s greatest contributions to the field of intelligence testing were two crucial psychometric concepts that he originated: regression to the mean and correlation. His concepts allowed for studying intelligence over time, as well as for studying relationships between intelligence scores of parents, children, etc. (a concept for which, on the basis of Galton’s work, Pearson developed the product-moment correlation and other related formulas).

James McKeen Cattell (1888) (as cited in Pintner, 1949), an assistant in Galton’s laboratory, brought Galton’s concepts to the United States. He shared his mentor’s philosophy that intelligence is best measured by sensory tasks, but expanded his use of “mental tests” (a term coined first by Cattell in the literature) to include standardized administration procedures. He pleaded for standardized procedures, and urged the necessity for the establishment of norms. Cattell’s valuable contribution to psychology was that he took the assessment of mental ability out of the field of abstract philosophy and showed that mental ability could be studied experimentally and practically. Under Cattell’s direction, the Pearson correlation technique seems to have been used for the first time for comparison of test with test, and tests with college grades.

By the late 1800s diverse notions of intelligence were conjectured, standardized procedures and norms were urged, and interest in classification had been implemented. Societal need provided the final impetus which led to the development of the individually administered Binet-Simon Scale in
1905. With the specific appointment by the French minister of public instruction to study the education of retarded children, the notion to separate mentally retarded and normal children in the Paris public schools arose. Binet, assisted by Theophile Simon and Victor Henri, rejected Galton’s notions of what made up intelligence and proposed that tasks utilizing higher mental processes (memory, comprehension, imagination, etc.) would be more effective measures. Binet did, however, retain Galton’s idea of general intelligence (“g”), which is reflected in his battery. This 1905 scale might be considered the first practical intelligence test.

The Binet has gone through a number of modifications and revisions throughout the years including the eventual introduction of the term “intelligence quotient” (IQ) in Terman’s 1916 version, the Stanford-Binet. This ratio IQ was computed by dividing mental age by chronological age, and multiplying by 100. While these single IQ scores have become a popular means of classifying individuals, it is a clear departure from Binet’s notion of intelligence as “a shifting complex of inter-related functions” (Tuddenham, 1962, p. 490). In fact, some doubt whether Binet would have accepted the concept of a single IQ score even with Terman’s elaborate standardization (Wolf, 1969). In 1986, Thorndike, Hagen, and Sattler developed a completely modified version of the Stanford-Binet, the SB-IV. The test incorporates Wechsler’s subtest format, and departs so much from the previous test that one wonders whether it merits the same name.

An important note is that this first major intelligence test battery, the Binet, arose to classify individuals. Classification has been fundamental to the history of mental assessment. It is no wonder that this philosophy continues today with such fervor, despite earnest attempts to move beyond single IQ scores in a desire to individualize profile analysis.

Like Binet, David Wechsler included the concept of global intelligence in his Wechsler-Bellevue Scale (published in 1939). Instead of having one global score, his battery included three separate IQ scores, a Verbal IQ, a Performance IQ, and a Full Scale IQ. The Full Scale IQ, for Wechsler, is an index of general mental ability (“g”).

While the formats from the original Stanford-Binet and the Wechsler Scales differ considerably, the subtests themselves do not. Wechsler’s tasks weren’t novel concepts at all, but rather, were borrowed from other tests of cognitive abilities. In many ways, Wechsler combined the philosophies of Esquirol (1828) and Seguin (1866/1907), and the psychometrics of Cattell and Terman. As equal components of intelligence, the Verbal Scale roughly capitalizes upon a person’s language abilities, (which expresses Esquirol’s philosophy), while the Performance Scale roughly capitalizes upon a person’s nonverbal and motoric abilities (as in Seguin’s view). Wechsler’s main ideas for the verbal tasks were the Stanford-Binet and the Army Group Examination Alpha. Ideas (and often specific items) for the performance tasks came primarily from the Army Group Examination Beta and the Army Individual Performance Scale Examination. The Army Alpha and Army Beta tests, published by Yerkes in 1917, were group administered intelligence tests developed to assess United States military recruits. The Wechsler Scales have gone through a number of revisions, but the basic test has remained structurally intact.

Although the Stanford-Binet and Wechsler scales were powerful tools to measure cognitive ability, theories of intelligence continued to be introduced and refined. In 1936, Piaget published Origins of Intelligence. He conceived of intelligence as a form of biological adaptation of the individual to the environment. Just as living organisms adapt to their environments biologically, individuals adapt to their environment through cognitive growth. Cognitive stages, therefore, emerge as a function of psychological structures reorganizing and/or developing out of organismic-environmental interactions (Piaget, 1950). Piaget’s model of intelligence is developmental and hierarchical in that he believes individuals pass through four predetermined, invariant stages of cognition, each more complex than the preceding one: sensorimotor (birth–2 years), preoperational thought (2 years–7 years), concrete operations (7 years–11 years), formal operations (11 years–adult).

With the advancement of psychometrics, factor analytic theories of intelligence emerged espousing either a general-factor theory (“g”) or a multiple-factor theory. Each method can be reduced to the other by either accepting the unrotated first factor, or by rotating the factors by various methods. Within this domain, J. P. Guilford (1959) developed a complex multifactor theory. His three-dimensional Structure of Intellect model (Guilford, 1967; Guilford & Hoepfner, 1971) posited five different operations, four types of content, and six products resulting in 120 possible factors ($5 \times 4 \times 6$). An even larger number of factors is possible.
based on Guilford’s (1988) modification of his model, in which he subdivided both on operation (Memory) and a content area (Figural) into two parts.

Also following the factor-analytic model were Raymond B. Cattell and John Horn (Cattell, 1963; Horn & Cattell, 1966, 1967), who postulated a structural model that separates fluid intelligence from crystallized intelligence. Fluid intelligence traditionally involves relatively culture-fair novel tasks and taps problem-solving skills and the ability to learn. Crystallized intelligence refers to acquired skill: knowledge and judgments that have been systematically taught or learned via acculturation. The latter type of intelligence is highly influenced by formal and informal education and often reflects cultural assimilation. Tasks measuring fluid ability often involve more concentration and problem solving than crystallized tasks, which tend to tap retrieval and application of general knowledge.

Another theoretical approach conceptualizing intelligence is an information-processing model focusing on the strategies individuals use to complete tasks successfully. Within this approach is the neuropsychological processing model which originated with the neurophysiological observations of Alexander Luria (1966a, 1966b, 1973, 1980) and Roger Sperry (1968), the psychoeducational research of J. P. Das (1973; Das et al., 1975, 1979; Naglieri & Das, 1988, 1990), and the psychometric research of A. S. and N. L. Kaufman (1983c). This model possesses several strengths relative to previous models in that it (a) provides a unified framework for interpreting a wide range of important individual difference variables; (b) rests on a well-researched theoretical base in clinical neuropsychology and psychobiology; (c) presents a processing, rather than a product-oriented, explanation for behavior; and (d) lends itself readily to clear remedial strategies based on relatively uncomplicated assessment procedures (Das et al., 1979; Kaufman & Kaufman, 1983c; McCallum & Merritt, 1983; Naglieri & Das, 1988, 1990; Perlman, 1986).

This neuropsychological processing model describes two very distinct types of processes that individuals use to organize and process information received in order to solve problems successfully: successive or sequential, analytic-linear processing versus holistic/simultaneous processing, (Levy & Trevarthen, 1976; Luria, 1966a). These processes have been identified by numerous researchers in diverse areas of neuropsychology and cognitive psychology (Perlman, 1986). From Sperry’s cerebral-specialization perspective, these processes represent the problem-solving strategies of the left hemisphere (analytic/sequential) and the right hemisphere (gestalt/holistic). From Luria’s theoretical approach, successive and simultaneous processes reflect the “coding” processes that characterize “Block 2” functions.

Regardless of theoretical model, successive processing refers to the processing of information in a sequential, serial order. The essential nature of this mode of processing is that the system is not totally surveyable at any point in time. Simultaneous processing refers to the synthesis of separate elements into groups. The essential nature of this mode of processing is that any portion of the result is at once surveyable without dependence on its position in the whole. The model assumes that the two modes of processing information are available to the individual. The selection of either or both modes of processing depends on two conditions: (a) the individual’s habitual mode of processing information as determined by social-cultural and genetic factors, and (b) the demands of the task (Das et al., 1975).

Many different theories and models of cognition underlie intelligence tests. However, it appears that recently factor-analytic and neuropsychological models have had a strong impact on test development in the field of intelligence testing. For example, the WJ-R utilizes a factor-analytic model while the K-ABC is based on a neuropsychological model. Both of these models can be translated into a unique way of viewing and assessing intelligence, and, when properly utilized, they have the ability to provide the examiner with a wealth of information about an individual’s cognitive functioning.

Many laws and judicial decisions have addressed the need for the development of nonbiased IQ tests for those having various learning deficiencies and those in minority groups. These laws and opinions underscore some of the controversy surrounding the appropriate use of intelligence tests and place ethical, if not legal, responsibility on clinicians for determining the adequacy and appropriateness of intelligence tests for children. The American Psychological Association clearly addresses this issue in their Ethical Principles of Psychologists and, under Principle 2-Competence, requires clinicians to recognize differences among people (age, sex, socioeco-
nomic, and ethnic backgrounds) and to understand test research regarding the validity and the limitations of their assessment tools (American Psychological Association, 1990).

**CURRENT MEASURES**

Intelligence tests are administered for a variety of reasons including *identification* (of mental retardation, learning disabilities, other cognitive disorders, giftedness), *placement* (gifted and other specialized programs), and as a cognitive adjunct to a *clinical evaluation*. The Wechsler Scales, Kaufman Scales, Stanford-Binet, and Woodcock-Johnson battery, are probably the most commonly used and most widely accepted individual intelligence measures. Administration of one of these more traditional measures is recommended for the assessment of intelligence when a child has the necessary physical capacities to respond to test questions, when the child meets age requirements of the test, and when there are no time restraints. When verbal responses cannot be elicited from a child, when sensory or motor impairments or both place limits on a child’s performance, or when time is at a premium, other measures become necessary. A review of these tests follows with a summary of other general cognitive measures and tests designed for special populations (infants and pre-schoolers, people with mental retardation, hearing and language impairment, visual impairment, orthopedic impairment, cultural minorities). The list of measures reviewed is by no means exhaustive, but represents the ones that are commonly used in the field today.

**Five Major Intelligence Scales**

*Wechsler Intelligence Scale for Children-Third Edition (WISC-III)*

*Theory.* Wechsler (1974) puts forth the definition that “intelligence is the overall capacity of an individual to understand and cope with the world around him” (p. 5). His tests, however, were not predicated on this definition. Tasks developed were not designed from well-researched concepts exemplifying his definition. In fact, as previously noted, virtually all of his tasks were adapted from other existing tests.

Like the Binet, Wechsler’s definition of intelligence also ascribes to the conception of intelligence as an overall global entity. He believed that intelligence cannot be tested directly, but can only be inferred from how an individual thinks, talks, moves, and reacts to different stimuli. Therefore, Wechsler did not give credence to one task above another, but believed that this global entity called intelligence could be ferreted out by probing a person with as many different kinds of mental tasks as one can conjure up. Wechsler did not believe in a cognitive hierarchy for his tasks, and he did not believe that each task was equally effective. He felt that each task was necessary for the fuller appraisal of intelligence.

**STANDARDIZATION AND PROPERTIES OF THE SCALE**

The WISC-III was standardized on 2,200 children ranging in age from 6 through 16 years. The children were divided into eleven age groups, one group for each year from 6 through 16 years of age. The median age for each age group was the sixth month (e.g., 7 years, 6 months). The standardization procedures followed the 1980 U.S. Census data and the manual provides information by age, gender, race or ethnicity, geographic region, and parent education. “Overall, the standardization of the WISC-III is immaculate...a better-standardized intelligence test does not exist (Kaufman, 1994, p.351).

The WISC-III yields three IQ scores, a Verbal Scale IQ, a Performance Scale IQ, and a Full Scale IQ. All three are standard scores (mean of 100 and standard deviation of 15) obtained by comparing an individual’s score with those earned by the representative sample of age peers. The WISC-III also yields four factor indexes, Verbal Comprehension, Perceptual Organization, Freedom from Distractibility and Processing Speed. The first two factors are in the cognitive domain, whereas the distractibility dimension is in the behavioral or affective domain. “The fourth factor seems to bridge the two domains; “processing” implies cognition, but “speed” has behavioral as well as cognitive components” (Kaufman, 1994, pp. 104, 105). The Verbal Comprehension Index measures abilities related to verbal conceptualization, knowledge, reasoning, and the ability to express ideas in words. The Freedom from Distractibility Index measures number ability and sequential process-
Table 4.1. Summary of Seven Steps for Interpreting WISC-III Profiles

**Step 1. Interpret the Full Scale IQ**
Convert it to an ability level and percentile rank and band it with error, preferable a 90% confidence interval (about ± 5 points).

**Step 2. Determine if the Verbal-Performance (V-P) IQ Discrepancy Is Statistically Significant**
Overall values for V-P discrepancies are 11 points at the .05 level and 15 points at the .01 level. For most testing purposes, the .05 level is adequate.

**Step 3. Determine if the V-P IQ Discrepancy Is Interpretable—Or if the VC and PO Factor Indexes Should Be Interpreted Instead**
Ask four questions about the Verbal and Performance Scales

**Verbal Scale**
1. Is there a significant difference (p < .05) between the child’s standard scores in VC versus FD?
**Size Needed for Significant (VC-FD) = 13+ points**
2. Is there abnormal scatter (highest minus lowest scaled score) among the five Verbal subtests used to compute V-IQ?
**Size Needed for Abnormal Verbal Scatter = 7+ points**

**Performance Scale**
3. Is there a significant difference (p < .05) between the child’s standard scores on PO versus PS?
**Size Needed for Significant (PO-PS) =15+ points**
4. Is there abnormal scatter (highest minus lowest scaled score) among the five Performance subtests used to compute P-IQ?
**Size Needed for Abnormal Performance Scatter =9+ points**

If all answers are no, the V-P IQ discrepancy is interpretable. If the answer to one or more questions is yes, the V-P IQ discrepancy may not be interpretable. Examine the VC-PO discrepancy. Overall values for VC-PO discrepancies are 12 points at the .05 level and 16 points at the .01 level.

**Step 4. Determine if the V-P IQ Discrepancy (or VC-PO Discrepancy) is Abnormally Large**
Differences of at least 19 points are unusually large for both the V-P and VC-PO discrepancies. Enter the table with the IQ or indexes, whichever was identified by the questions and answers in Step 3.
If neither set of scores was found to be interpretable in Step 3, they may be interpreted anyway if the magnitude of the discrepancy is unusually large (19+ points).

**Step 5. Interpret the Meaning of the Global Verbal and Nonverbal Dimensions and the Meaning of the Small Factors**
Study the information and procedures presented in Chapter 4 (verbal/nonverbal) and Chapter 5 (FD and PS factors). Chapter 5 provides the following rules regarding when the FD and PS factors have too much scatter to permit meaningful interpretation of their respective indexes:
1. Do not interpret the FD index if the Arithmetic and Digit Span scaled scores differ by 4 or more points.
2. Do not interpret the PO index if the Symbol Search and Coding scaled scores differ by 4 or more points.

**Step 6. Interpret Significant Strengths and Weaknesses in the WISC-III Subtest Profile**
If the V-P IQ discrepancy is less than 19 points, use the child’s mean of all WISC-III subtests administered as the child’s midpoint.
If the V-P IQ discrepancy is 19 or more points, use the child’s mean of all Verbal subtests as the midpoint for determining strengths and weaknesses on Verbal subtests, and use the Performance mean for determining significant deviations on Performance subtests.
Using either the specific values in Table 3.3 of *Intelligent Testing with the WISC-III* (Kaufman, 1994), rounded to the nearest whole number, or the following summary information of determining significant deviations:

<table>
<thead>
<tr>
<th>Deviation</th>
<th>Subtests</th>
</tr>
</thead>
<tbody>
<tr>
<td>± 3 points</td>
<td>Information, Similarities, Arithmetic, Vocabulary</td>
</tr>
<tr>
<td>± 4 points</td>
<td>Comprehension, Digit Span, Picture Completion, Picture Arrangement, Block Design, Object Assembly, Symbol Search</td>
</tr>
<tr>
<td>± 5 points</td>
<td>Coding</td>
</tr>
</tbody>
</table>

(continued)
Step 7. Generate Hypotheses about the Fluctuations in the WISC-III Subtest Profile

Consult Chapter 6 in Intelligent Testing with the WISC-III, (Kaufman, 1994) as it deals with the systematic reorganization of subtest profiles to generate hypotheses about strengths and weaknesses.

Table 4.1. (Continued)

<table>
<thead>
<tr>
<th>Step 7. Generate Hypotheses about the Fluctuations in the WISC-III Subtest Profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consult Chapter 6 in Intelligent Testing with the WISC-III, (Kaufman, 1994) as it deals with the systematic reorganization of subtest profiles to generate hypotheses about strengths and weaknesses.</td>
</tr>
</tbody>
</table>

**Note:**
- VC = Verbal Comprehension; PO = Perceptual Organization; FD = Freedom from Distractibility; PS = Processing Speed.

Within the WISC-III there are 10 mandatory and 3 supplementary subtests, all of which span the age range of 6 to 16 years. The Verbal Scale’s five mandatory subtests include: Information, Similarities, Arithmetic, Vocabulary, and Comprehension. The supplementary subtest on the Verbal Scale is Digit Span. Digit Span is not calculated into the Verbal IQ unless it has been substituted for another Verbal subtest because one of those subtests has been spoiled (Kamphaus, 1993; Wechsler, 1991).

The five mandatory Performance Scale’s subtests include Picture Completion, Picture Arrangement, Block Design, Object Assembly, and Coding. The two supplementary subtests on the Performance Scale are Mazes and Symbol Search. The Mazes subtest may be substituted for any Performance Scale subtest; however, Symbol Search may only be substituted for the Coding subtest (Kamphaus, 1993; Wechsler, 1991).

“Symbol Search is an excellent task that should have been included among the five regular Performance subtests instead of Coding. Mazes is an awful task that should have been dropped completely from the WISC-III” (Kaufman, 1994, p.58). He goes further to say that “there’s no rational reason for the publisher to have rigidly clung to Coding as a regular part of the WISC-III when the new Symbol Search task is clearly a better choice for psychometric reasons” (Kaufman, 1994, p. 59). Therefore, for all general purposes, Kaufman (1994) strongly recommends that Symbol Search be routinely substituted for Coding as part of the regular battery, and to use Symbol Search to compute the Performance IQ and Full Scale IQ. The manual does not tell one to do this, but neither does it prohibit it.

Reliability of each subtest except Coding and Symbol Search was estimated by the split-half method. Stability coefficients were used as reliability estimates for the Coding and Symbol Search subtests because of their speeded nature. Across the age groups, the average reliability coefficients are: Information (.84), Similarities (.81), Arithmetic (.78), Vocabulary (.87), Comprehension (.77), Digit Span (.85), Picture Completion (.77), Coding (.79), Picture Arrangement (.76), Block Design (.87), Object Assembly (.69), Symbol Search (.76), and Mazes (.70). The average reliability, across the age groups, for the IQs and Indexes are: .95 for the Verbal IQ, .91 for the Performance IQ, .96 for the Full Scale IQ, .94 for the Verbal Comprehension Index, .90 for the Perceptual Organization Index, .87 for the Freedom from Distractibility Index, and .85 for the Processing Speed Index (WISC-III Interpretive Manual, 1991).

Analyzing the WISC-III Data. To obtain the most information from the WISC-III, one should be more than familiar with each of the subtests individually as well as with the potential information that those subtests can provide when integrated or combined. The WISC-III is maximally useful when tasks are grouped and regrouped to uncover a child’s strong and weak areas of functioning, so long as these hypothesized assets and deficits are verified by multiple sources of information.

As indicated previously, the WISC-III provides examiners with a set of four Factor Indexes in addition to the set of three IQs. The front page of the WISC-III record form lists the seven standard scores in a box on the top right. The record form is quite uniform and laid out nicely; however, it is difficult to know just what to do with all of those scores. Kaufman (1994) has developed The Seven Steps which offer a unique and systematic method of WISC-III interpretation that allows the clinician
to organize and integrate the test results in a stepwise, easy-to-use approach. The Seven Steps provide an empirical framework for profile attack while organizing the profile information into hierarchies. Table 4.1 provides an overview of the seven interpretive steps for WISC-III profiles.

**Critique.** Professionals in the field of intelligence testing have described the third edition of the Wechsler Intelligence Scale for Children in a number of different ways. "The WISC-III reports continuity, the status quo, and only the smallest step in the evolution of the assessment of intelligence. Despite more than 50 years of advancement in theories of intelligence, the Wechsler philosophy of intelligence (not actually a formal theory), written in 1939, remains the guiding principle of the WISC-III" (Shaw, Swerdlik, & Laurent, 1993, p. 151). One of the principal goals for developing the WISC-III stated in the manual was merely to update the norms, which is "hardly a revision at all" (Sternberg, 1993). If one has chosen to use the WISC-III because he or she is looking for a test of new constructs in intelligence, or merely a new test, one should look elsewhere (Sternberg, 1993). In contrast to these fairly negative evaluations, Kaufman (1994) reports that the WISC-III is a substantial revision of the WISC-R and that the changes that have been made are considerable and well done. "The normative sample is exemplary, and the entire psychometric approach to test development, validation, and interpretation reflects sophisticated, state-of-the-art knowledge and competence" (Kaufman, 1994). For Kaufman, the WISC-III is not without its flaws but his overall review of the test is quite positive. Although the WISC-III has clearly had mixed reviews, it is one of the most frequently used tests in the field of children's intelligence testing.

*Kaufman Assessment Battery for Children (K-ABC)*

The K-ABC is a battery of tests measuring intelligence and achievement of normal and exceptional children ages 2½ through 12½ years. It yields four scales: the Sequential Processing Scale, the Simultaneous Processing Scale, the Mental Processing Composite (Sequential and Simultaneous) Scale, and the Achievement Scale. The K-ABC is becoming a frequently used test in intelligence and achievement assessment that is used by both clinical and school psychologists (Kamphaus, Beres, Kaufman, & Kaufman, 1995). In a nationwide survey of school psychologists conducted in 1987 by Obringer (1988), respondents were asked to rank the following instruments in order of their usage: Wechsler's scales, the K-ABC, and both the old and new Stanford-Binet. The Wechsler scales earned a mean rank of 2.69, followed closely by the K-ABC with a mean of 2.55, the L-M version of the Binet 1.98 and the Stanford-Binet: Fourth Edition 1.26. Bracken (1985) also found similar results of the K-ABC's increasing popularity. Bracken surveyed school psychologists and found that for ages 5 to 11 years the WISC-R was endorsed by 82 percent, the K-ABC by 57 percent, and the Binet IV by 39 percent of the practitioners. These results suggest that clinicians working with children should have some familiarity with the K-ABC (Kamphaus et al., 1995).

The K-ABC has been the subject of great controversy from the outset, as evident in the strongly pro and con articles written for a special issue of the *Journal of Special Education* devoted to the K-ABC (Miller & Reynolds, 1984). Many of the controversies, especially those regarding the validity of the K-ABC theory, will likely endure unresolved for some time (Kamphaus et al., 1995). Fortunately, the apparent controversy linked to the K-ABC has resulted in numerous research studies and papers that provide more insight into the K-ABC and its strengths and weaknesses.

*Theory.* The K-ABC intelligence scales are based on a theoretical framework of sequential and simultaneous information-processing, which relates to *how* children solve problems rather than *what* type of problems they must solve (e.g., verbal or nonverbal), which is in stark contrast to Wechsler’s theoretical framework of the assessment of "g", a conception of intelligence as an overall global entity. As a result, Wechsler used the Verbal and Performance scales as a means to an end. That end is the assessment of general intelligence. In comparison, the Kaufmans emphasize the individual importance of the Sequential and Simultaneous Scales in interpretation, rather than the overall Mental Processing Composite (MPC) score (Kamphaus et al., 1995).

The sequential and simultaneous framework for the K-ABC stems from an updated version of a
Table 4.2. Representation of the Standardization Sample by Educational Placement (N = 2,000)

<table>
<thead>
<tr>
<th>Educational Placement</th>
<th>K-ABC Standardization Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
</tr>
<tr>
<td>Regular Classroom</td>
<td>1,862</td>
</tr>
<tr>
<td>Speech Impaired</td>
<td>28</td>
</tr>
<tr>
<td>Learning Disabled</td>
<td>23</td>
</tr>
<tr>
<td>Mentally Retarded</td>
<td>37</td>
</tr>
<tr>
<td>Emotionally Disturbed</td>
<td>5</td>
</tr>
<tr>
<td>Other b</td>
<td>15</td>
</tr>
<tr>
<td>Gifted and Talented</td>
<td>30</td>
</tr>
<tr>
<td>Total K-ABC Sample</td>
<td>2,000</td>
</tr>
</tbody>
</table>


Includes other health impaired, orthopedically handicapped, and hard of hearing.


A variety of theories (Kamphaus, 1993). The foundation lies in a wealth of research in clinical and experimental neuropsychology and cognitive psychology. The sequential and simultaneous theory was primarily developed from two lines of theory: the information-processing approach of Luria (e.g., Luria, 1966a), and the cerebral-specialization work of Sperry (1968, 1974), Bogen (1975), Kinsbourne (1978), and Wada, Clarke, and Hamm (1975).

In reference to the K-ABC, simultaneous processing refers to the mental ability to integrate information all at once to solve a problem correctly. Simultaneous processing frequently involves spatial, analogic, or organizational abilities (Kaufman & Kaufman, 1983c; Kamphaus & Reynolds, 1987). There is often a visual aspect to the problem and visual imagery used to solve it. A prototypical example of a simultaneous subtest is the Triangles subtest on the K-ABC, which is similar to Wechsler’s Block Design. To solve both of these subtests, children must be able to see the whole picture in their mind and then integrate the individual pieces to create the whole.

In comparison, sequential processing emphasizes the ability to place or arrange stimuli in sequential or serial order. The stimuli are all linearly or temporally related to one another, creating a form of serial interdependence within the stimulus (Kaufman & Kaufman, 1983c). The K-ABC subtests assess the child’s sequential processing abilities in a variety of modes. For example, Hand Movements involves visual input and a motor response, Number Recall involves auditory input with a verbal response, and Word Order involves auditory input and visual response. These different modes of input and output allow the examiner to assess the child’s sequential abilities in a variety of ways. The sequential subtests also provide information on the child’s short-term memory and attentional abilities.

According to Kamphaus et al. (1995), one of the controversial aspects of the K-ABC was the fact that it took the equivalent of Wechsler’s Verbal Scale and redefined it as “achievement”. The Kaufmans’ analogs of tests such as Information (Faces & Places), Vocabulary (Riddles and Expressive Vocabulary), and Arithmetic (Arithmetic) are included on the K-ABC as achievement tests. The Kaufmans viewed the above tests as diverse tasks that are united by the demands they place on children to extract and assimilate information from their cultural and school environment. The K-ABC is predicated on the distinction between problem solving and knowledge of facts. The former set of skills is interpreted as intelligence; the latter is defined as achievement. This definition presents a break from other intelligence tests, where a person’s acquired factual information and applied skills frequently influence greatly the obtained IQ (Kaufman & Kaufman, 1983c).

Standardization and Properties of the Scale.

Stratification of the K-ABC standardization sample was excellent and closely matched the 1980 U.S. Census data in age, gender, geographic region, community size, socioeconomic status, race and ethnic group, and parental occupation and education. Additionally, unlike most other intelligence measures for children, stratification variables also
Reliability and validity data are impressive. A test-retest reliability study was conducted with 246 children after a 2- to 4-week interval (mean interval = 17 days). The coefficients for the Mental Processing Composite were .83 for ages 2 years, 6 months through 4 years, 11 months; .88 for ages 5 years, 0 months through 8 years, 11 months; and .93 for ages 9 years, 0 months to 12 years, 5 months. Test-retest reliabilities for the Achievement scale composite for the same age groups were .95, .95, and .97 respectively (Kamphaus et al., 1995). The test-retest reliability research reveals that there is a clear developmental trend, with coefficients for the preschool ages being smaller than those for the school-age range. This trend is consistent with the known variability over time that characterizes preschool children's standardized test performance in general (Kamphaus & Reynolds, 1987).

Split-half reliability coefficients for the K-ABC global scales range from 0.86 to 0.93 (mean = 0.90) for preschool children, and from 0.89 to 0.97 (mean = 0.93) for children aged 5 to 12 ½ (Kamphaus et al., 1995).

There has been a considerable amount of research done on the validity of the K-ABC. The K-ABC Interpretive Manual (Kaufman & Kaufman, 1983c) includes the results of 43 such studies. Construct validity was established by looking at five separate topics: developmental changes, internal consistency, factor analysis (principal factor, principal components, and confirmatory), convergent and discriminant analysis, and correlations with other tests. Factor analysis of the Mental Processing Scales offered clear empirical support for the existence of two, and only two, factors at each age level, and for the placement of each preschool and school-age subtest on its respective scale. Analyses of the combined processing and achievement subtests also offered good construct validation of the K-ABC's three-scale structure (Kaufman & Kamphaus, 1984).

Although the K-ABC and the WISC-III differ from one another in a number of ways, there is strong evidence that the two measures correlate substantially (Kamphaus & Reynolds, 1987). In a study of 182 children enrolled in regular classrooms, the Mental Processing Composite (MPC) correlated 0.70 with WISC-R Full-Scale IQ (FSIQ), thus, sharing a 49 percent overlap in variance (Kamphaus et al., 1995; Kaufman & Kaufman, 1983c). There have also been numerous correlational studies conducted with handicapped and exceptional populations that may be found in the Interpretive Manual.

Critique. Although the K-ABC has been the subject of past controversy, it appears that it has held its own and is used often by professionals. The K-ABC is well designed with easels and manuals that are easy to use. The information in the manuals is presented in a straightforward, clear fashion, making use and interpretation of the K-ABC relatively easy (Merz, 1985). There has been a considerable amount of research done on the validity of the K-ABC and the authors have done a thorough job of presenting much of that information in the manual. The reporting of the reliability and validity data in the manual is complete and understandable. However, there is not enough information presented on the content validity of the test. The various tasks on the subtests on the K-ABC are based on clinical, neuropsychological or other research-based validity; however, a much clearer explication of the rationale behind some of the novel subtests would have been quite helpful (Merz, 1985).

The K-ABC measures intelligence from a strong theoretical and research basis, evident in the quality of investigation in the amount of research data presented in the manual (Merz, 1985). The K-ABC was designed to measure the intelligence and achievement of children 2 ½ to 12 ½ years old and the research done to date suggests that in fact the test does just that. The Nonverbal Scale significantly contributes to the effort to address the diverse needs of minority groups and language-handicapped children. Overall, it appears that the authors of the K-ABC have met the goals listed in the interpretative manual and that this battery is a valuable assessment tool (Merz, 1985).

In a number of studies, Keith and his colleagues (Keith, 1985; Keith & Dunbar, 1984) have called the K-ABC processing model into question by applying Wechsler-like content labels to the K-ABC scales. Keith (1985) used labels such as "nonverbal/reasoning" (Simultaneous), "achievement/verbal reasoning" (Achievement), and "verbal memory" (Sequential) for the K-ABC factors, making the scales similar to the tradition of psychological assessment. "The issue of what to call the K-ABC factors remains debated but unresolved" (Kamphaus, 1993).
**Table 4.3.** Representation of the Stanford-Binet, Fourth Edition

<table>
<thead>
<tr>
<th>By Parental Occupation</th>
<th>SAMPLE PERCENT</th>
<th>U.S. POPULATION PERCENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Managerial/Professional</td>
<td>45.9</td>
<td>21.8</td>
</tr>
<tr>
<td>Technical Sales</td>
<td>26.2</td>
<td>29.7</td>
</tr>
<tr>
<td>Service Occupations</td>
<td>9.7</td>
<td>13.1</td>
</tr>
<tr>
<td>Farming/Forestry</td>
<td>3.2</td>
<td>2.9</td>
</tr>
<tr>
<td>Precision Production</td>
<td>6.7</td>
<td>13.0</td>
</tr>
<tr>
<td>Operators, Fabricators, Other</td>
<td>8.3</td>
<td>19.5</td>
</tr>
<tr>
<td>Total</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>By Parental Education</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>College Graduate or Beyond</td>
<td>43.7</td>
<td>19.0</td>
</tr>
<tr>
<td>1 to 3 Years of College</td>
<td>18.2</td>
<td>15.3</td>
</tr>
<tr>
<td>High School Graduate</td>
<td>27.5</td>
<td>36.5</td>
</tr>
<tr>
<td>Less Than High School</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graduate</td>
<td>10.6</td>
<td>29.2</td>
</tr>
<tr>
<td>Total</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>


*Theory.* Like its predecessor, the Fourth Edition (SB-IV) is based on the principal of a general ability factor, "g," rather than on a connection of separate functions. The Fourth Edition has maintained, yet to a much lesser degree, its adaptive testing-format. No examinee takes all the items on the scale, nor do all examinees of the same chronological age respond to the same tasks. Like its predecessor, the scale provides a continuous appraisal of cognitive development from ages two through adult.

One of the criticisms of the previous version is that it tended to underestimate the intelligence of examinees whose strongest abilities did not lie in verbal skills (or overestimate the intelligences of those whose verbal skills excelled). Therefore, consideration when developing the SB-IV was to give equal credence to several areas of cognitive functioning. The authors set out to appraise verbal reasoning, quantitative reasoning, abstract/visual reasoning, and short-term memory (in addition to a composite score representing "g").

This model is based on a three-level hierarchical model of the structure of cognitive abilities. A general reasoning factor is at the top level ("g"). The next level consists of three broad factors: crystallized abilities, fluid analytic abilities, and short-term memory. The third level consists of more specific factors: verbal reasoning, quantitative reasoning, and abstract/visual reasoning.

The selection of these four areas of cognitive abilities came from the authors' research and clinical experience of the kinds of cognitive abilities that correlate with school progress. This foundational emphasis on academic cognition continues the philosophy of the original Binet, which did not extend to measuring adult intelligence as did later versions, including the SB-IV. One wonders whether the same emphasis should be used when measuring adult intelligence. While subtests change (with considerable overlap) for various age groups and while selection reportedly has been subjected to rigorous research, there is considerable dispute whether children and adults utilize the same intellectual processes. After all, any task can be developed and normed for a variety of ages, but does that mean that each age group is calling upon the same processes to accomplish this task?

The SB-IV contains previous tasks combining old with new items and some completely new tasks. In general, test items were accepted if (a) they proved to be acceptable measurements of the construct; (b) they could be reliably administered and scored; (c) they were relatively free of ethnic or gender bias; and (d) they functioned adequately over a wide range of age groups (again, not making philosophical distinctions between intelligence of children and adults).
Standardization and Properties of the Scale. Standardization procedures followed 1980 U.S. Census data. There appears to be an accurate sample representation from geographic region, size of community, ethnic group, and gender. The standardization falls short, however, in terms of age, parental occupation, and parental education. The age representation extends from 2 years of age, 0 months to 23 years, 11 months. The concentration of the sample is on children 4 to 9 years old (41%). Not only were adults 24 years and older not represented, but also representation beyond age 17 years, 11 months was negligible (4%).

In order to assess characteristics of socioeconomic status (SES), information regarding parental occupation and parental education was obtained. A review of Table 4.3 demonstrates that children whose parents came from managerial or professional occupations and/or who were college graduates and beyond were grossly overrepresented in the sample. In other words, the norms are based on a large percentage of individuals from upper-socioeconomic classes. In order to adjust for this discrepancy, an after-the-fact weighting procedure was applied, which makes the norming sample suspect. Unquestionably, SES has been shown time and again to be the single most important stratification variable regarding its relationship to IQ (Kaufman, 1990a, Chapter 6; Kaufman & Doppelt, 1976).

According to McCallum (1990), there is a considerable amount of evidence for the general construct validity of the SB-IV. For example, the difficulty level of items within the various subtests is developmentally determined. In other words, age and cognitive maturity are highly correlated with success on items. Therefore, older children are more likely to succeed on the items than younger children. Additionally, the SB-IV measures intelligence in ways that are similar to older, established tests of intelligence. Correlation coefficients between the SB-IV global scores from the Wechsler scales, the Stanford-Binet (Form L-M), and the K-ABC range from .50 to .85 (McCallum, 1990).

Research also shows that the individual subtests of the SB-IV had impressive high to substantial loadings on “g” (.51-.79). Unfortunately, the four factors were given weak support by the confirmatory procedure. Additionally, exploratory factor analysis gave even less justification for the four Binet Scales; only one or two factors were identified by Reynolds, Kamphaus, and Rosenthal (1988) for 16 of the 17 age groups studied. Clearly, the factor analytic structure does not conform to the theoretical framework used to construct the test. Therefore, once again one is left with the composite score as the only clearly valid representation of a child’s cognitive abilities.

Correlational studies, using non-exceptional children, between the SB-IV and the Stanford-Binet (Form L-M), WISC-R, Wechsler Adult Intelligence Scale-Revised (WAIS-R), Wechsler Preschool and Primary Scale of Intelligence (WPPSI), K-ABC have ranged from .80 to .91 (comparing full-scale composites). Correlational studies using exceptional children (gifted, learning impaired, mentally retarded) produced generally lower correlations, probably because of restricted variability in the test scores. These data and data from similar validity investigations are presented more extensively in the Technical Manual for the SB-IV (Thorndike, Hagen, & Sattler, 1986). Hodapp (1993) conducted a correlational study between the SB-IV and the PPVT-R with a group of 42 children ranging in age from 3 to 6 years. Correlations of .54, .60, and .50 were computed for Standard Age Scores on the SB-IV Composite, Vocabulary, and Absurdities with the PPVT-R standard score equivalent. The seven other SB-IV subtests showed correlations ranging from .25 to .38.

There appears to be a considerable amount of diversity in the conclusions drawn from the research on the validity and usefulness of the SB-IV. However, in general, the evaluations of the SB-IV tend to be rather negative, suggesting that its use in the field may be limited. The irresponsibly gathered normative data, and other difficulties with the SB-IV have led at least one reviewer to recommend that the battery be laid to rest (Reynolds, 1987); “To the SB-IV, Requiescat in pace” (p. 141).

Critique. The SB-IV was developed in an attempt to increase the popularity of the test as well as address some of the negative reviews that had plagued the previous edition. The test authors attempted to make the Fourth Edition significantly different from the previous L-M Edition; however, it does not appear that they succeeded in doing so. Canter (1990) describes the “rebirth” of the Stanford-Binet as giving way to “confusion and even dismay as the primary consumers of intelligence tests learned that the new edition offered a more complicated route to the same destination.” Another reviewer describes the SB-IV as “in most
assesment of child intelligence

respects, a completely new version of a very old test" (Spruill, 1987). It appears that the Fourth Edition of the Stanford-Binet has been a disappointment for most professionals in the field of intelligence testing.

One of the major problems with the Fourth Edition is the fact that it went into publication too soon. As a result, the test was published without accompanying technical data to allow the user to evaluate the appropriateness and technical adequacy of the instrument. This made it difficult for the examiner to know if the test was appropriate for his or her client, not to mention that it is a violation of Standard 5.1 in the Standards for Educational and Psychological Testing (Spruill, 1987). Furthermore, there were errors in the norms tables in the first printing of the administration manual.

A common criticism of the SB-IV and previous versions is that there had been inadequate standardization. For example, in the Fourth Edition the standardization sample contained a larger percentage of high-socioeconomic-status subjects than in the population at large, as demonstrated previously. It is not clear whether or not the weighting procedure that was used to correct for sample bias was adequate (Spruill, 1987). Also, the test was designed to be used with individuals from age 2 to "adult"; however, there are no normative data for "adults" over the age of 23. This may also be misleading to an examiner.

Although there appears to be a number of flaws with the SB-IV, the test is still used and it is not without its strengths. The administration of some of the subtests allow the examiner a little flexibility, and young children seem to find the items challenging and fun. Despite its shortcomings, SB-IV continues to be a very good assessment of cognitive skills related to academic progress (Spruill, 1987). It also includes several excellent, well-constructed tasks that offer valuable supplementary information when they are administered as Weschler supplements (Kaufman, 1990a, 1994).

Theory. The Horn-Cattell theory forms the foundation of the KAIT and defines the constructs presumed to be measured by the separate IQs; however, other theories guided the test development process, specifically the construction of the subtests. Tasks were developed from the models of Piaget's formal operations (Inhelder & Piaget, 1958; Piaget, 1972) and Luria's (1973, 1980) planning ability in an attempt to include high-level decision making on more developmentally advanced tasks. Luria's notion of planning ability involves decision making, evaluation of hypotheses, and flexibility, and "represents the highest levels of development of the mammalian brain" (Golden, 1981, p.285).

Piaget's formal operations depicts a hypotheti-
cal-deductive abstract reasoning system that has as its featured capabilities the generation and evaluation of hypotheses and the testing of propositions. The prefrontal areas of the brain associated with planning ability mature at about ages 11 to 12 years (Golden, 1981), the same ages that characterize the onset of formal operational thought (Piaget, 1972). The convergence of the Luria and Piaget theories regarding the ability to deal with abstractions is striking; this convergence provided the rationale for having age 11 as the lower bound of the KAIT, and for attempting to measure decision making and abstract thinking with virtually every task on the KAIT (Kaufman & Kaufman, 1993).

Within the KAIT framework (Kaufman & Kauf-
man, 1993), Crystallized intelligence "measures the acquisition of facts on problem solving ability using stimuli that are dependent on formal schooling, cultural experiences, and verbal conceptual development" (p.7). Fluid intelligence "measures a person's adaptability and flexibility when faced with new problems, using both verbal and nonverbal stimuli" (Kaufman & Kaufman, 1993, p. 7). It is important to note that this Crystallized-Fluid construct split is not the same as Wechsler's (1974, 1981, 1991) verbal-nonverbal split. More specifically, the KAIT Fluid subtests stress reasoning rather than visual-spatial ability, include verbal comprehension or expression as key aspects of some tasks, and minimize the role played by visual-motor speed for correct responding.

The Core Battery of the KAIT is composed of three Crystallized and three Fluid subtests, and these six subtests are used to compute the IQs. The Expanded Battery also includes two supplementary subtests and two measures of delayed recall that evaluate the individual's ability to retain infor-
Table 4.4. Correlations of the Three KAIT IQ with Standard Scores and IQs yielded by Other Major Intelligence Tests

<table>
<thead>
<tr>
<th>INTELLIGENCE TEST</th>
<th>AGE RANGE</th>
<th>CRYSTALLIZED</th>
<th>FLUID</th>
<th>COMPOSITE</th>
</tr>
</thead>
<tbody>
<tr>
<td>WAIS-R Verbal IQ (N=343)</td>
<td>16–83</td>
<td>0.78</td>
<td>0.62</td>
<td>0.76</td>
</tr>
<tr>
<td>WAIS-R Performance IQ (N=343)</td>
<td>16–83</td>
<td>0.72</td>
<td>0.72</td>
<td>0.77</td>
</tr>
<tr>
<td>WAIS-R Full Scale IQ (N=343)</td>
<td>16–83</td>
<td>0.86</td>
<td>0.73</td>
<td>0.85</td>
</tr>
<tr>
<td>WISC-R Verbal IQ (N=118)</td>
<td>11–16</td>
<td>0.79</td>
<td>0.74</td>
<td>0.83</td>
</tr>
<tr>
<td>WISC-R Performance IQ (N=118)</td>
<td>11–16</td>
<td>0.67</td>
<td>0.67</td>
<td>0.72</td>
</tr>
<tr>
<td>WISC-R Full Scale IQ (N=118)</td>
<td>11–16</td>
<td>0.78</td>
<td>0.75</td>
<td>0.82</td>
</tr>
<tr>
<td>K-ABC Mental Processing Composite (N=124)</td>
<td>11–12</td>
<td>0.57</td>
<td>0.62</td>
<td>0.66</td>
</tr>
<tr>
<td>K-ABC Achievement (N=124)</td>
<td>11–12</td>
<td>0.81</td>
<td>0.64</td>
<td>0.82</td>
</tr>
<tr>
<td>SB-V Test Composite (N=79)</td>
<td>11–42</td>
<td>0.81</td>
<td>0.84</td>
<td>0.87</td>
</tr>
</tbody>
</table>

Note: Data in this table are based on data reported in the KAIT Manual (Kaufman & Kaufman, 1993, Tables 8.15–8.19 and 8.22–8.23. Data for the WAIS-R are averages of values reported separately for four age groups between (a) 16 and 19 years and (b) 50 and 83 years.

The delayed-recall subtests are administered, without prior warning, about 25 and 45 minutes after the administration of the original, related subtests. The two-delayed recall subtests provide good measure of an ability that Horn (1985, 1989) calls TSR (Long-Term Storage and Retrieval). TSR “involves the storage of information and the fluency of retrieving it later through association” (Woodcock, 1990, p. 234).

The Mental Status subtest is comprised of ten simple questions that assess attention and orientation to the world. Most normal adolescents and adults pass at least nine of the ten items, but the task has special use with retarded and neurologically impaired populations. The Mental Status subtest may be used as a screener to determine if the KAIT can be validly administered to an individual. Standardization and Properties of the Scale. The KAIT normative sample, composed of 2,000 adolescents and adults between the ages of 11 and 94 years, was stratified on the variables of gender, racial/ethnic group, geographic region, and socioeconomic status (Kaufman & Kaufman, 1993).

Mean split-half reliability coefficients for the total normative sample were .95 for Crystallized IQ, .95 for Fluid IQ, and .97 for Composite IQ (Kaufman & Kaufman, 1993). Mean test-retest reliability coefficients, based on 153 identified normal individuals in three age groups (11–19 years of age, 20–54 years of age, 55–85+ years of age), retested after a one-month interval, were .94 for Crystallized IQ, .87 for Fluid IQ, and .94 for Composite IQ (Kaufman & Kaufman, 1993). Mean subtest split-half reliabilities of the four Crystallized subtests ranged from .89 to .92 (median = .90). Mean values for the four Fluid subtests ranged from .79 to .93 (median = .88) (Kaufman & Kaufman, 1993). Median test-retest reliabilities for the eight subtests, based on the 153 people indicated previously, ranged from .72 to .95 (median = .78). Rebus Delayed Recall had an average split-half reliability of .91 and Auditory Delayed Recall had an average value of .71; their respective stability coefficients were .80 and .63 (Kaufman & Kaufman, 1993).
Factor analysis, both exploratory and confirmatory, gave strong construct validity support for the Fluid and Crystallized Scales, and for the placement of each subtest on its designated scale. Crystallized IQs correlated .72 with Fluid IQs for the total standardization sample of 2,000 (Kaufman & Kaufman, 1993). Table 4.4 provides the correlations of the three KAIT IQs with standard scores and IQs yielded by other major intelligence tests. The data found in this table are taken from the KAIT Technical Manual (1993). The values shown in Table 4.4 support the construct and criterion-related validity of the three KAIT IQs.

The KAIT benefits from an integration of theories that unite developmental (Piaget), neuropsychological (Luria), and experimental-cognitive (Horn-Cattell) models of intellectual functioning. The theories work well together and do not compete with one another. Together, the theories give the KAIT a solid theoretical foundation that facilitate test interpretation across the broad 11 to 94 year age-range on which the battery was normed.

The KAIT and WISC-R were administered to 118 individuals ages 11 to 16 years, and the KAIT and WAIS-R were administered to 338 individuals ages 16 to 83 years; these data were factor analyzed in two separate joint analyses. A number of analyses were conducted to determine what factors each of the tests have that are unique and what factors they share. "The most crucial finding from these analyses is that the Wechsler Performance subtests and the KAIT Fluid subtests seem to measure markedly different constructs" (Kaufman & Kaufman, 1993, p. 93). According to Horn, there are important differences between Performance IQ and fluid intelligence, noting that Performance IQ "involves visualization to a very considerable extent" (Horn & Hofer, 1992, p. 72). The following conclusions from the joint factor analyses of KAIT and Wechsler subtests were drawn:

1. Three factors define the joint matrices of the KAIT and the Wechsler scales: Crystallized/Verbal, Fluid, and Perceptual Organization.
2. The constructs underlying the KAIT Fluid and the Wechsler Performance Scales are distinctly different. The Fluid and Perceptual Organization factors correlate about as highly with each other as they do with the Crystallized/Verbal factor.
3. The constructs underlying the KAIT Crystallized the Wechsler Verbal scales seem virtually identical; all component subtests load substantially on the Crystallized/Verbal factor.
4. The KAIT Crystallized and Fluid subtests load consistently on the factors underlying their respective scales. The Wechsler subtests, however, sometimes do not load highly on the factor underlying the scale to which they belong (Kaufman & Kaufman, 1993).

Critique. The KAIT represents a reconceptualization of the measurement of intelligence that is more consistent with current theories of intellectual development (Brown, 1994). The fluid-crystallized dichotomy, the theory underlying the KAIT, is based on the original Horn-Cattell theory of intelligence, thus offering a firm and well-researched theoretical framework (Flanagan, Alfonso, & Flanagan, 1994). The fluid-crystalized dichotomy enhances the richness of the clinical interpretations that can be drawn from this instrument (Brown, 1994). The test materials are well constructed and attractive, and the manual is well organized and helpful (Dumont & Hagberg, 1994; Flanagan et al., 1994). Furthermore, the test materials are easy to use and stimulating to examinees (Flanagan et al., 1994).

"The KAIT has been standardized by state-of-the-art measurement techniques" (Brown, 1994). The psychometric properties of the KAIT regarding standardization and reliability are excellent and the construct validity evidence that is reported in the manual provides a good foundation for its theoretical underpinnings (Flanagan et al., 1994).

The theoretical assumption that formal operations are reached by early adolescence limits the application of the KAIT with certain adolescent and adult populations (Brown, 1994). If an individual has not achieved formal operations, many of the subtests will be too difficult for them and perhaps frustrating and overwhelming. Examiners should be aware of this when working with such individuals in order to maintain rapport. The KAIT can be a great assessment tool when working with high-functioning, intelligent individuals; however, it can be difficult to use with borderline individuals and some elderly clients. Elderly clients' scores on some of the subtests may be negatively impacted by poor reading, poor hearing, and poor memory (Dumont & Hagberg, 1994).

Flanagan and colleagues (1994) report that the inclusion of only three subtests per scale may limit or interfere with the calculation of IQs if a subtest
is spoiled. The usefulness of the Expanded Battery and Mental Status subtest of clinical populations is questionable given the reliability and validity data presented in the manual, suggesting that interpretations be made with caution (Flanagan et al., 1994).

Although there clearly are some limitations in the use of the KAIT with some populations, overall, the test appears to be well thought out and validated (Dumont & Hagberg, 1994). The KAIT represents an advancement in the field of intellectual assessment with its ability to measure fluid and crystallized intelligence from a theoretical perspective and, at the same time, maintain a solid psychometric quality (Flanagan et al., 1994).

### Woodcock-Johnson Psycho-Educational Battery—Revised: Tests of Cognitive Ability (WJ-R)

The WJ-R is one of the most comprehensive test batteries available for the clinical assessment of children and adolescents (Kamphaus, 1993). The WJ-R is a battery of tests for individuals from 2 to 90+ years of age, and is composed of two sections, Cognitive and Achievement. The focus of this discussion is the Cognitive portion of the WJ-R battery.


The four subtests that measure Long-Term Retrieval (Memory for Names, Visual-Auditory Learning, Delayed Recall/Memory for Names, Delayed Recall/Visual-Auditory Learning), require the subject to retrieve information stored minutes or a couple of days earlier. In contrast, the subtests that measure Short-Term Memory (Memory for Sentences, Memory for Words, Numbers Reversed) require the subject to store information and retrieve it immediately or within a few seconds. The two Processing Speed subtests (Visual Matching, Cross Out) assess the subject's ability to work quickly, particularly under pressure, to maintain focused attention.

Within the Auditory-Processing domain, three subtests (Incomplete Words, Sound Blending, Sound Patterns) assess the subject's ability to fluently perceive patterns among auditory stimuli. The three Visual-Processing subtests (Visual Closure, Picture Recognition, Spatial Relations) assess the subject's ability to fluently manipulate stimuli that are within the visual domain.

Picture Vocabulary, Oral Vocabulary, Listening Comprehension, and Verbal Analogies are the four subtests that are linked to the Comprehension-Knowledge factor, also known as crystallized intelligence within Horn's theoretical model. These subtests require the subject to demonstrate the breadth and depth of his or her knowledge of a culture. Analysis-Synthesis, Concept Formation, Spatial Relations, and Verbal Analogies (which also loads on the Comprehension-Knowledge factor) assess the subject's Fluid Reasoning. Finally, from the Achievement portion of the WJ-R, both the Calculation and Applied Problems subtests assess the individual's Quantitative Ability.

The cognitive battery consists of 21 subtests, 7 of which comprise the standard battery; the remaining 14 are part of the supplemental battery. There are two composite scores, Broad Cognitive Ability and Early Development (for preschoolers), which are both comparable to an overall IQ score. The individual subtest scores, as well as the composite scores, have a mean of 100 and a standard deviation of 15.

Computer software is available for scoring the WJ-R and is essential if one is to obtain all of the information that the WJ-R is capable of providing. The WJ-R provides the examiner with percentile ranks, grade-based scores, age-based scores and the Relative Mastery Index (RMI). The RMI is unique and similar to a ratio with the second part of the ratio set at a value of 90. The denominator of the ratio means that children in the norm sample can perform the intellectual task with 90 percent accuracy. The numerator of the ratio refers to that child or adolescent's proficiency on that subtest (Kamphaus, 1993). For example, if a child obtains an RMI of 60/90, it would mean that the child's proficiency on the subtest is at a 60 percent level whereas the typical child of his or her age (or grade) mastered the material at a 90 percent level of accuracy.
The entire battery is quite lengthy and therefore can be timely to administer. The Standard Battery takes approximately 40 minutes to administer; however, all the clinician will obtain from it is, essentially, a measure of “g”. In order to obtain all of the information that the WJ-R is capable of providing, a clinician should administer most of the subtests in both the Cognitive and Achievement batteries. Administration of a thorough cognitive and achievement assessment using the WJ-R would take approximately 3½ to 5 hours depending on the subject’s age, abilities, and speed. However, individual subtests may be administered to test specific hypothesis without administering the entire battery.

**Standardization and Properties of the Scale.** The WJ-R was normed on a representative sample of 6,359 individuals selected to provide a cross-section of the U.S. population from 2 to 90+ years of age (Woodcock & Mather, 1989). The sample included 705 preschool children, 3,245 students in grades K through 12, 916 college or university students, and 1,493 individuals aged 14 to 90+ years who were not enrolled in school. Stratification variables included gender, geographic region, community size, and race. However, Kaufman (1990a) reports that although representation on important background variables was adequate, it was not excellent and therefore necessitated the use of a weighting procedure.

The internal consistency estimates for the standard battery are relatively high. The median coefficients are above .80 for five of the seven subtests. The Broad Cognitive Ability composite score based on seven standard battery subtests yields a median internal consistency coefficient of .94, and the Broad Cognitive Ability Early Development scale yields a coefficient of .96 at ages 2 and 4 years (Kamphaus, 1993).

The Woodcock-Johnson Psycho-Educational Battery-Revised: Examiner’s Manual reports that “Items included in the various tests were selected using item validity studies as well as expert opinion” (Woodcock & Mather, 1989, p.7). Kamphaus (1993) states that the manual should have included more information on the results of the experts’ judgements or some information on the methods and results of the studies that were used to assess validity.

It is clear that the WJ-R Cognitive battery is quite comprehensive, providing the clinician with a wealth of information. The standardization sample is large, the factor loadings reveal generally strong factor-analytic support for the construct validity for the battery for adolescents and adults, and the reliability coefficients are excellent (Kaufman, 1990a).

**Critique.** The WJ-R Cognitive battery was developed based on Horn’s expansion of the Cattell-Horn Fluid-Crystallized model of intelligence. This theoretical rationale allows for further empirical analysis of both the WJ-R and the theory (Webster, 1994). The standardization of the battery appears to be sound and the various age groups are adequately represented. According to Webster (1994), the Cognitive battery is quite thorough, and when administered in its entirety, can provide the examiner with a wealth of information about an individual’s intellectual functioning and abilities. The test materials and manuals are easy to use and well designed. The administration is fairly simple; however, scoring the test, especially when the Achievement battery is administered as well, can be quite a lengthy process. The scoring can be done by hand but is done more efficiently with the computer-scoring program. The computer-scoring program is easy to use and provides the examiner with the individual’s raw scores, standard scores, percentile ranks, and age and grade equivalents for each subtest (Webster, 1994).

Kaufman (1986) reviewed the 1977 version of the Woodcock-Johnson (WJ) battery and concluded that it “is a mixture of extremes, possessing some outstanding qualities, yet hampered by glaring liabilities.” He went further to add that the WJ represents a monumental and creative effort by its authors and he encourages examiners to take the time to master the test. Cummings (1985) agreed that the WJ is a “significant addition” to the available psychometric instruments. According to Kaufman (1990a), these comments apply as well to the WJ-R, although he expressed concern about interpreting many scales, each composed of few subtests. The WJ-R Cognitive battery is a well-standardized test developed on an interesting theory of intelligence. However, the test is not without shortcomings. Webster (1994) raises issues with the specific psychometric procedures used in developing test items. Data are lacking that show the efficacy of the WJ-R to predict, from a time-based perspective, actual functional levels of academic achievement and identify children at-risk-for-failure early in the educational process (Webster, 1994).
Other General Cognitive Measures

In addition to the major intelligence tests previously discussed, there are a number of other cognitive measures that are frequently used to assess the intelligence of both children and adolescents. These measures were developed based on a number of different theories and each of them offers a unique way of assessing the individual. This section of the text will provide general information on five cognitive tests for children: Peabody Picture Vocabulary Test-Revised (PPVT-R), Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R), Detroit Tests of Learning Aptitude (DTLA-3), Matrix Analogies Test, and Differential Abilities Scales (DAS). The tests that were chosen for this section are by no means exhaustive. In fact, there are a number of tests that have not been discussed. For example, the Kaufman Brief Intelligence Test (K-BIT) is integrated in the case report at the end of this chapter; however, it was not presented in the group of cognitive tests that were chosen to be discussed in this chapter.

Peabody Picture Vocabulary Test-Revised (PPVT-R)

This brief test provides an approximate estimate of intelligence by measuring receptive vocabulary and replaces the original Peabody Picture Vocabulary Test (PPVT) published in 1959 (Dunn & Dunn, 1981). The 1981 version retains many of its predecessor’s best features: it consists of two equivalent forms, allows for a verbal or nonverbal response, and is untimed. The examinee is shown plates with four pictures on each and is to point to the picture that best illustrates the meaning of the stimulus word spoken by the examiner. The PPVT-R is appropriate for individuals aged 2½ years through adult who can hear the stimulus word, see the drawings, and respond in some manner.

While the original PPVT was normed on a large but restricted sample, the PPVT-R norms were based on a nationwide data-gathering effort which, for children, was representative of the 1970 U.S. Census data with regard to sex, age, geographic region, occupational background, race and ethnic background, and urban-rural distributions. Because only 828 adults (ages 19 through 40 years) in contrast to 4,200 children, were included in the standardization, the manual suggests careful interpretation of scores for individuals above 18 years, 11 months old. Minority groups were included in the normative sample and are also included on the test plates. Sex- and ethnic-stereotyping, a problem with the original PPVT, has been virtually eliminated. The pictorial stimuli were redrawn to reflect a more appropriate racial, ethnic, and gender representation. Following the trend of other new or revised tests, the PPVT-R adopted conversion of raw scores to either percentile ranks, age equivalents, or standard score equivalents (mean = 100; standard deviation = 15).

The test manual reports moderate internal consistency (.61 to .88) and alternate form reliability estimates (.71 to .91) for the standardization sample. The degree of equivalence of the two forms was established for a subsample of 642 children. Coefficients of equivalence ranged from .73 to .91 (median = .82). Correlations of the PPVT-R with other intelligence composites typically range from .40 to .60 (Dunn & Dunn, 1986; Kaufman & Kaufman, 1983c; McCallum, 1985). These modest concurrent validity estimates suggest limited shared variance. Therefore, the PPVT-R should not be interpreted as equivalent to intelligence test scores.

Critique. As a test of hearing vocabulary, the PPVT-R is one of the most widely used instruments of its kind (Umberger, 1985). The PPVT-R is an easy-to-use test of receptive language, providing content that is current and that contains appropriate racial, ethnic, and gender representation. The national representative standardization for the educationally critical age range (2½ to 19 years) responds to requirements set by P.L. 94-242 (Wiig, 1985) and psychometric characteristics of this latest revision appear adequate to excellent (McCallum, 1985). It allows for flexibility in administration which lends itself to applicability to a number of exceptional populations (Umberger, 1985).

As a brief test, the PPVT-R is reliable, but it is not as reliable as one would expect for a test composed of 175 items. The reliability is greatly hindered by the element of chance that the test has within its design. Each item on the PPVT-R is a four-option multiple-choice question, meaning that guessers will be correct one out of four times. Another shortcoming of the PPVT-R is that its norms stop at 40 and that the adult norms are inferior to its superb norms for children and adolescents (Kaufman, 1990). Kaufman (1990a) also
cites a number of strengths of the PPVT-R that should be mentioned. For example, the PPVT-R’s reliability and validity data for adolescents and adults are generally good, and the revisions were done thoroughly and with extreme care.

The test materials are well designed, making the test easy to administer and score. The test can be administered and scored quickly, allowing quick interpretation. Facilitation of interpretation has also been improved by providing the ability to convert raw scores to percentile ranks, age equivalents, and standard-score equivalents. In addition to being a useful assessment tool, the PPVT-R has a wide application as a research tool or as one test in a battery of tests on language competence (Umberger, 1985).

Wechsler Preschool and Primary Scale of Intelligence—Revised (WPPSI-R)

The WPPSI-R is an intelligence test for children aged 3 years, 0 months through 7 years, 3 months. The original version of the WPPSI was developed in 1967 for ages 4 to 6 ½ years, and the WPPSI-R was revised in 1989. Several changes were made to the revised version of the WPPSI-R. The norms were updated, the appeal of the content to young children was improved, and the age range was expanded.

The WPPSI-R is based on the same Wechsler-Bellevue theory of intelligence, emphasizing intelligence as a global capacity but having Verbal and Performance scales as two methods of assessing this global capacity (Kamphaus, 1993). The Verbal scale subtests include: Information, Comprehension, Arithmetic, Vocabulary, Similarities, and Sentences. The Performance scale subtests include: Object Assembly, Block Design, Mazes, Picture Completion, and Animal Pegs. Both the Sentences subtest and the Animal-Pegs subtest are supplemental tests and may be used in place of other subtests when deemed necessary.

Like the K-ABC and the Differential Abilities Scales (DAS), the WPPSI-R allows the examiner to “help” or “teach” the client on early items on the subtests to assure that the child understands what is expected of him or her. Providing this extra help is essential when working with reticent preschoolers (Kamphaus, 1993). Subtest scores have a mean of 10 and a standard deviation of 3. The overall Verbal, Performance, and Full Scale IQs have a mean of 100 and a standard deviation of 15. The examiner manual also provides interpretive tables that allow the examiner to determine individual strengths and weaknesses as well as the statistical significance and clinical rarity of Verbal and Performance score differences.

The WPPSI-R was standardized on 1,700 children from ages three years through seven years, 3 months. The standardization procedures followed the 1986 U.S. Census Bureau estimates. Stratification variables included gender, race, geographic region, parental occupation, and parental education.

The WPPSI-R appears to be a highly reliable measure. The internal consistency coefficients across age groups, for the Verbal, Performance, and Full Scale IQs, are .95, .92, and .96 respectively. The reliability coefficients for the individual subtests vary considerably, from an average internal consistency coefficient of .86 for Similarities to an average of .63 for Object Assembly. With a group of 175 children from the standardization sample, a test-retest investigation was conducted. The investigation yielded coefficients in the high .80s and low .90s. The test-retest coefficient for the Full Scale IQ is .91 (Kamphaus, 1993).

The WPPSI-R manual provides some information on validity; however, it provides no information on the predictive validity of the test. Various studies have shown that concurrent validity between the WPPSI-R and other tests is adequate. The correlation between the WPPSI and the WPPSI-R Full Scale IQs was reported at .87, and the correlation between WPPSI-R and WISC-III Performance, Verbal, and Full Scale IQs for a sample of 188 children was .73, .85, and .85 respectively. The correlations between he WPPSI-R and other well known cognitive measures is, on average, much lower. The WPPSI-R Full Scale IQ correlated .55 with the K-ABC Mental Processing Composite (Kamphaus, 1993) and .77 with the SB-IV. In general, the validity coefficients provide strong evidence for the construct validity of the WPPSI-R (Kamphaus, 1993).

Critique. The WPPSI-R is a thorough revision of the 1967 WPPSI and is for an expanded age range. It has new colorful materials, and item-types for very young children, as well as a new icebreaker subtest (Object Assembly) and a comprehensive manual (Kaufman, 1990a). The revision of the test has resulted in an instrument that is more attractive, and engaging, and has materials that are easier
to use (Buckhalt, 1991; Delugach, 1991). The normative sample is large, provides recent norms, and is representative of the 1986 U.S. Census data (Delugach, 1991; Kaufman, 1990a). The split-half reliability of the IQs and most subtests are exceptional, the factor analytic results for all age groups are excellent, and the concurrent validity of the battery is well supported by several excellent correlational studies (Delugach, 1991; Kaufman, 1990a). The manual provides a number of validity studies, factor-analytic results, research overviews, and state-of-the-art interpretive tables, which provide the examiner with a wealth of information.

“The WPPSI-R is standing on a rock-solid psychometric foundation” (Kaufman, 1990a).

In spite of its reported strengths, the WPPSI-R has numerous flaws. The WPPSI-R has an insufficient floor at the lowest age levels, which limits the test’s ability to diagnose intellectual deficiency in young preschoolers (Delugach, 1991). The directions on some of the Performance subtests are not suitable for young children because they are not developmentally appropriate, and the heavy emphasis on response speed on some nonverbal test is inappropriate for young children who have not yet internalized the importance of working very quickly (Kaufman, 1990a). However, Delugach (1991) reports that if the directions are too difficult, the test provides procedures to ensure that the child understands the demands of the task.

The WPPSI-R is a useful assessment tool, but, like all others, it possesses certain weaknesses that limit its usefulness (Delugach, 1991). Examiners should be aware of the WPPSI-R’s inherent strengths and weaknesses and keep them in mind during administration, scoring, and interpretation. The WPPSI-R may provide the examiner with useful information; however, "it does little to advance our basic understanding of the development and differentiation of intelligence or our understanding of the nature of individual differences in intelligence" (Buckhalt, 1991).

Detroit Tests of Learning Aptitude (DTLA-3)

Harry J. Baker and Bernice Leland recognized that the study of intra-individual strengths and weaknesses could be enhanced by the availability of a test battery composed of sort subtests that measured different abilities and that were standardized on the same population. In order to properly assess these intra-individual strengths and weaknesses, Baker and Leland developed the Detroit Tests of Learning Aptitude (DTLA) in 1935. The DTLA was comprised of 19 subtests and was appropriate for use with individuals between the ages of 4 and 19 years. A number of abilities could be assessed by the DTLA, including reasoning, verbal skills, time and space relationships, number, attention, and motor abilities.

Baker and Leland’s original DTLA was used until it was revised in 1985 by Donald Hammill. The DTLA-2 was designed by Hammill to be used for individuals aged 6 years through 17 years 11 months. The DTLA-2 included 11 subtests and 9 composites. The reviews of the DTLA-2 include both positive and negative evaluations. One of the primary criticisms was that there was not enough information provided on how the standardization sample was selected. An attempt was made to rectify the shortcoming of the DTLA-2 and incorporate many of the suggestions from the original reviews into the DTLA-3 (Hammill, 1991).

The DTLA-3, developed by Hammill in 1991, was designed to measure different, but interrelated, mental abilities for individuals ages 6 years through 17 years, 11 months. It is a battery of 11 subtests and has 16 composites that measure both general intelligence and discrete ability areas. Hammill and Bryant (1991) report that the DTLA-3 was greatly influenced by Spearman’s two-factor theory (1927). This theory of “aptitude” consisted of a general factor “g” that is present in all intellectual pursuits, and specific factors that vary from task to task (McGhee, 1993).

The 11 subtests are used to form the 16 composite scores. The subtests are grouped into different combinations according to various hypothetical constructs that exist in current theories of intelligence and information-processing (McGhee, 1993). In general, the composite scores estimate general mental ability; however, they all do so in a somewhat different manner. The General Mental Ability Composite is formed by combining the standard scores of all 11 subtests, and thus, has been referred to as the best estimate of “g”. The Optimal Level Composite is composed of the four largest standard scores that the individual earns. This individualized score is often referred to as the best estimate of a person’s overall “potential.” The Domain Composites may be divided into three areas; Linguistic, Attentional, and Motoric. Furthermore, there is a Verbal and Nonverbal Composite in the Linguistic domain, an Attention-Enhanced and Attention-Reduced Composite in
the Attentional Domain, and a Motor Enhanced and a Motor-Reduced composite in the Motoric Domain. Finally, there are the Theoretical Composites of the DTLA-3 on which the battery’s subtests are constructed. The major theories upon which the subtests were developed include Horn and Cattell’s (1966) fluid and crystallized intelligences, Das’s (1973) simultaneous and successive processes, Jensen’s (1980) associative and cognitive levels, and Wechsler’s (1974, 1981, 1989) verbal and performance scales.

The DTLA-3 yields five types of scores: raw scores, subtest standard scores, composite quotients, percentiles, and age equivalents. Standard scores for the individual subtests have a mean of 10 and a standard deviation of 3 and the Composite Quotients have a mean of 100 and a standard deviation of 15. The individual subtest reliabilities range from .77 to .94 (median=.87) and the averaged alphas for the composites range from .89 to .96 (median=.94). To assess the DTLA-3’s stability over time, the test-retest method was used with a sample of 34 children residing in Austin, Texas. The children, ages 6 through 16 years, were tested twice, with a two-week period between testings (Hammill, 1991). The results of this test-retest analysis indicate that individual subtest reliabilities range from .75 to .96 (median=.86) and composite reliabilities range from .81 to .96 (median=.89).

Critique. The DTLA-3 was designed to measure both general intelligence and discrete ability for children ages 6 years to 17 years 11 months. The DTLA-3 is not grounded in one specific theory but rather can be linked to a number of different theorists and their views on intelligence and achievement. This “eclectic” theorizing has resulted in the DTLA-3’s numerous subtests, composites, and various combinations of the two that yield potentially important information about an individual’s abilities.

Reliability and validity studies are encouraging but are based on specific and limited samples (VanLeirsburg, 1994). Additional research in this area would be beneficial. Furthermore, test-retest reliability data were collapsed across age levels, which makes it impossible to determine the stability of scores of the various age levels (Schmidt, 1994). The standardization sample was representative of the U.S. population but more information on socioeconomic level is needed (Schmidt, 1994). Also, there is no normative data reported for subjects with handicapping conditions and sample stratification for age was not equalized (VanLeirsburg, 1994).

The testing manual suggests that individual testing may vary but that on average it takes 50 minutes to 2 hours to administer. Scoring and interpretation of the results is easy, yet it can be quite time-consuming without the aid of the accompanying computer program (VanLeirsburg, 1994). Despite apparent shortcomings, the DTLA-3 should be useful for eligibility or placement purposes and for research (Schmidt, 1994).

The Matrix Analogies Test

The Matrix Analogies Test (Naglieri, 1985) is composed of a set of figural matrices that can be used as a measure of general intelligence. There is little language involvement; therefore, the test is particularly well suited to assessing the intelligence of individuals with hearing impairments and language disabilities as well as the intelligence of children whose first language is not English.

The age range of the Matrix Analogies Test is from 5 through 18 years. There are two forms of the test: a group-administered form and the expanded form. The group-administered or short form consists of 34 multiple-choice items and the expanded form, which is individually administered, consists of 64 multiple-choice items. Raw scores on the expanded form can be converted into standard scores with a mean of 100 and a standard deviation of 15. These scores can then be converted into age equivalents and percentile ranks. On the short form of the test, the raw scores can be converted into percentiles, stanines, and age equivalents (Kamphaus, 1993).

The expanded form of the Matrix Analogies Test was standardized on a sample of 5,718 children in the early 1980s. The stratification of the sample matched U.S. Census statistics by race, sex, age, geographic region, community size, and socioeconomic status.

Internal consistency reliability of the expanded form ranges from .88 to .95. However, test-retest reliability of the total test score over a one-month interval was lower (.77). The validity of the expanded form has been evaluated primarily via correlations of other tests. Correlations with the WISC-R Full Scale IQ were .41 for a sample of 82 nonhandicapped children; .43 for Native American children, and .68 for hearing-impaired children (Kamphaus, 1993). According to Naglieri and
Prewett (1990), the trend across studies on the expanded form is for individuals to score about ten points lower on the Matrix Analogies test than on the Performance score of the WISC-R.

**Critique.** The Matrix Analogies Test—Expanded Form and the Matrix Analogies Test—Short Form can be useful tools in assessing the intelligence of children with communication or motor problems as well as the intelligence of children whose first language is not English. The test is user-friendly and easy to score and administer (Robinson, 1987). The Matrix Analogies Test is considerably more modern than many of its predecessors and its norming sample is more recent, larger, and more psychometrically sophisticated. The expanded form is a useful single-screener of intelligence for clinical or research use, based upon the mental processing of figural matrices by children (Kamphaus, 1993), while the short form may serve as a useful screening device (Robinson, 1987).

**Differential Abilities Scales (DAS)**

The DAS was developed by Elliott (1990a) and is an individually administered battery of cognitive and achievement tests for use with individuals aged 2½ through 17 years. The DAS Cognitive Battery has a preschool level and a school-age level. The preschool core consists of the following cognitive core subtests: Verbal Comprehension, Naming Vocabulary, Picture Similarities, Pattern Construction, Copying, and Early Number Concepts. The school-age cognitive core subtests include: Word Definitions, Similarities, Matrices, Sequential and Quantitative Reasoning, Recall of Designs, and Pattern Construction. The school-age level also includes reading-, mathematics-, and spelling-achievement tests that are referred to as “screeners.” The same sample of subjects was used to develop the norms for the Cognitive and Achievement Batteries; therefore, intra- and inter-comparisons of the two domains is possible.

The DAS is not based on a specific theory of intelligence. Instead, the test’s structure is based on tradition and statistical analysis. Nonetheless, the test is not theory-free, and, in fact, is based in part on “g” and the view of intelligence as hierarchical in nature (McGhee, 1993). Elliott (1990b) described his approach to the development of the DAS as “eclectic” and cited researchers such as Cattell, Horn, Das, Jensen, Thurstone, Vernon, and Spearman. Indeed, there are some clear-cut relationships between several DAS scales and theoretical constructs. For example, Horn’s (1985, 1989) concepts of fluid and crystallized intelligence are measured quite well by the Nonverbal Reasoning and Verbal Ability scales, respectively. Elliott emphasizes Thurstone’s ideas that the emphasis on intellectual assessment should be on the assessment and interpretation of distinct abilities (Kamphaus, 1993). Therefore, subtests were constructed to emphasize their unique variance, which should translate into unique abilities.

The cognitive portion of the DAS consists of core and diagnostic subtests designed to assess intelligence. The achievement portion measures skills in the areas of word reading, spelling, and mathematics. The core subtests are averaged to obtain the General Conceptual Ability (GCA) score and, depending on the age of the individual, additional composite scores, referred to as Cluster scores, are calculated (McGhee, 1993).

The individual Cognitive subtests have a mean of 50 and a standard deviation of 10. The GCA scores, Cluster scores, and Achievement scores, have a mean of 100 and a standard deviation of 15. Percentile ranks, age equivalents, and score comparisons are also available in the examiner’s manual. Score comparisons provide a profile analysis and allow the examiner to ascertain information regarding aptitude-achievement discrepancies.

The norm sample of the DAS closely approximated U.S. census statistics estimated from 1986 to 1988, with the sample stratified by English-proficient, noninstitutionalized children from four U.S. geographic regions. At the preschool level, 175 children were included in each six-month age sample for ages 2 years, 6 months to 4 years, 11 months, with 200 children included at the 5-years to 5-years-11-months age-range. Children were divided equally at each age level for gender (Irvin, 1992). Exceptional children were also included in the standardization sample. Sex, race, geographic region, community size, and enrollment (for ages 2–5 years through 5–11 years) in an educational program were controlled. Socioeconomic status was estimated using the average education level of the parents living with the child (Kamphaus, 1993).

The average internal-consistency estimates for the clusters at the school-age level are .88 for Verbal Ability, .90 for Nonverbal Reasoning Ability, and .92 for Spatial Ability. Internal consistency reliabilities of the subtests are also relatively
Table 4.5. Case Report Test Results

Wechsler Intelligence Scale for Children—Third Edition (WISC—III)

**IQs**

Verbal Scale 93 ± 5 (32nd percentile)
Performance Scale 84 ± 5 (14th percentile)
Full Scale 88 ± 4 (21st percentile)

**Factor Index Scores**

Verbal Comprehension 95 ± 5 (37th percentile)
Perceptual Organization 85 ± 6 (16th percentile)
Freedom from Distractibility 101 ± 8 (53rd percentile)
Processing Speed 91 ± 7 (27th percentile)

**Subtest Scaled Scores**

<table>
<thead>
<tr>
<th>Subtest Scaled Scores</th>
<th>SCALED SCORE</th>
<th>PERCENTILE RANK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information</td>
<td>8</td>
<td>25</td>
</tr>
<tr>
<td>Similarities</td>
<td>11-S</td>
<td>63</td>
</tr>
<tr>
<td>Arithmetic</td>
<td>8</td>
<td>25</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>8</td>
<td>25</td>
</tr>
<tr>
<td>Comprehension</td>
<td>9</td>
<td>37</td>
</tr>
<tr>
<td>(Digit Span)</td>
<td>12-S</td>
<td>75</td>
</tr>
<tr>
<td>Picture Completion</td>
<td>3-W</td>
<td>1</td>
</tr>
<tr>
<td>Coding</td>
<td>9</td>
<td>37</td>
</tr>
<tr>
<td>Picture Arrangement</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Block Design</td>
<td>9</td>
<td>37</td>
</tr>
<tr>
<td>Object Assembly</td>
<td>12-S</td>
<td>75</td>
</tr>
<tr>
<td>(Symbol Search)</td>
<td>7</td>
<td>16</td>
</tr>
</tbody>
</table>

Kaufman Brief Intelligence Test (K-B±T)

Composite IQ 93 ± 6 (32nd percentile)

| Vocabulary            | 89           | 23              |
| Matrices              | 98           | 45              |

Woodcock—Johnson Psycho Educational Battery-Revised: Tests of Cognitive Ability (WJ-R): Selected Subtests

<table>
<thead>
<tr>
<th>SUBTEST/CLUSTER</th>
<th>STANDARD SCORE (± SEM)</th>
<th>PERCENTILE RANK</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Memory for Names</td>
<td>108 ± 3</td>
<td>69</td>
</tr>
<tr>
<td>8. Visual-Auditory Learning</td>
<td>97 ± 5</td>
<td>41</td>
</tr>
<tr>
<td>Long-Term Retrieval (Tests 1 &amp; 8)</td>
<td>103 ± 4</td>
<td>59</td>
</tr>
<tr>
<td>4. Incomplete Words</td>
<td>90 ± 3</td>
<td>26</td>
</tr>
<tr>
<td>11. Sound Blending</td>
<td>101 ± 4</td>
<td>54</td>
</tr>
<tr>
<td>Auditory Processing (Tests 4 &amp; 11)</td>
<td>96 ± 7</td>
<td>40</td>
</tr>
<tr>
<td>5. Visual Closure</td>
<td>109 ± 6</td>
<td>73</td>
</tr>
<tr>
<td>12. Picture Recognition</td>
<td>91 ± 5</td>
<td>27</td>
</tr>
<tr>
<td>Visual Processing (Tests 5 &amp; 12)</td>
<td>98 ± 7</td>
<td>45</td>
</tr>
<tr>
<td>7. Analysis-Synthesis</td>
<td>103 ± 4</td>
<td>59</td>
</tr>
<tr>
<td>14. Concept Formation</td>
<td>112 ± 4</td>
<td>78</td>
</tr>
<tr>
<td>Fluid Reasoning (Tests 7 &amp; 14)</td>
<td>108 ± 4</td>
<td>71</td>
</tr>
<tr>
<td>Delayed Recall Memory for Names</td>
<td>69 ± 7</td>
<td>02</td>
</tr>
<tr>
<td>Delayed Recall Visual-Auditory Learning</td>
<td>72 ± 7</td>
<td>03</td>
</tr>
</tbody>
</table>

(continued)
Table 4.5. (Continued)

Kaufman Test of Educational Achievement-Comprehensive Form (K-TEA)

<table>
<thead>
<tr>
<th></th>
<th>SCALED SCORE</th>
<th>PERCENTILE RANK</th>
<th>GRADE EQUIVALENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mathematics Applications</td>
<td>115</td>
<td>84</td>
<td>12.3</td>
</tr>
<tr>
<td>Reading Decoding</td>
<td>101</td>
<td>53</td>
<td>8.7</td>
</tr>
<tr>
<td>Spelling</td>
<td>92</td>
<td>30</td>
<td>6.4</td>
</tr>
<tr>
<td>Reading Comprehension</td>
<td>104</td>
<td>61</td>
<td>8.8</td>
</tr>
<tr>
<td>Mathematics Computation</td>
<td>107</td>
<td>68</td>
<td>9.3</td>
</tr>
</tbody>
</table>

strong, with only a few exceptions. The mean reliability coefficient for Recall of Objects, for example, is only .71, and for Recognition of Pictures, only .73 (Kamphaus, 1993).

Correlational research has shown good evidence of concurrent validity for the DAS (Kamphaus, 1993). With a sample of 27 children aged 7 to 14 years, the WISC-III Full Scale IQ correlated very highly with the DAS GCA score (.92), and the WISC-III Verbal IQ score correlated highly with the DAS Verbal Ability score (.87). The WISC-III Performance IQ correlated .78 with Nonverbal Reasoning and .82 with Spatial Ability. Additionally, the DAS Speed of Information Processing subtest score correlated .67 with the WISC-III Processing Speed Index score. The SB-IV Composite IQ also yielded strong correlations with the DAS GCA score, .88 for 9- and 10-year-olds and .85 for a sample of gifted children. The K-ABC Mental Processing Composite correlated with the DAS GCA yielded a correlation of .75 for 5- to 7-year-olds (Kamphaus, 1993).

Critique. In general, the professional reviews of the DAS seem to be quite positive. Sandoval (1992) believes that the DAS is one of the least obviously biased tests available today. The test development and the test results have resulted in a relatively culturally fair measure. However, its use with linguistically different children needs to be explored further (Sandoval, 1992). Sandoval does not provide clear evidence for support of his statement that the DAS is one of the least biased tests available. In fact, when one considers that the test is comprised of a number of verbal subtests it becomes very unclear how Sandoval could make such a claim. In fact, according to Bain (1991), the DAS appears to be useful in assessing both white and black students with learning problems; however, caution is recommended in using the DAS to predict achievement for Hispanic students because there is evidence that the test over-predicts achievement for this group based on group achievement results.

ILLUSTRATIVE CASE REPORT

An illustrative case report follows for Jared, an adolescent of 13 ½ years of age, who was referred for psychoeducational assessment because of school problems as well as emotional and behavior difficulties. The WISC-III, Kaufman Brief Intelligence Test (K-BIT), selected subtests from the Woodcock-Johnson Tests of Cognitive Abilities-Revised (WI-R), Kaufman Test of Educational Achievement: Comprehensive Form (K-TEA), Rorschach Inkblot Test, Incomplete Sentence Test, Bender Gestalt Test, House-Tree-Person Test, and Kinetic Family Drawing Test were administered to Jared. The Child Behavior Checklist (CBCL) was also administered to Jared and his parents. (The client’s name and pertinent identifying information have been changed to ensure anonymity.)

Referral and Background Information

Jared, a 13 ½-year-old male, was referred for evaluation by his parents, Mr. and Mrs. P. and Dr. Z., his psychiatrist. Jared’s parents and Dr. Z would like to gain insight into Jared’s current and previous level of cognitive and emotional functioning. Two months prior to this evaluation, on two separate occasions, Jared was placed in a psychiatric hospital for out-of-control behavior. After the hospitalizations, Jared was expelled from school for multiple suspensions and having a
Jared lives at home with his sister and both parents. He is the oldest of two children; his younger sister is 11 years old. Jared’s birth history and early developmental history are unremarkable. He reached all developmental milestones within a normal time frame. It is reported that Jared has always had difficulty sleeping, even as an infant.

Jared attended preschool one to two mornings a week from age 2 to 5 years. His parents report that Jared cried when he first began school but that it only took him a short time to adjust. In the second grade, Jared was identified as gifted and was placed in the gifted program at his school. He reportedly did well in school and did not have any difficulties until the 4th grade when he had two different teachers. Mr. and Mrs. P. described that year at school, where the teachers alternated instruction, as problematic for Jared. Although Jared did well during his first few years in elementary school, his parents report that he has always had difficulty paying attention and sitting still for school work. His parents also related that as school became more difficult and advanced, Jared seemed to have more problems. Specifically, he had great difficulty with reading and organizational skills.

School records, as well as parental and self-report, suggest that Jared both does well in, and enjoys, math. His parents also describe him as having a “great memory.”

Jared’s Middle School records indicate that he was quite disruptive in class. His grades prior to being expelled were primarily Ds and Fs. In comparison, Jared had approximately a B+ average, two years earlier, in the sixth grade. It appears that his grades did not begin to drop until the seventh grade, despite a reported history of difficulty with reading, focusing, and paying attention. When not in school, Jared enjoys skiing, surfing, watching television, and listening to music. He currently plays in a basketball league and he states that he really enjoys playing and that he looks forward to the games.

Appearance and Behavioral Characteristics

Jared is a handsome adolescent with big green eyes and short straight blonde hair. He was well groomed and dressed casually in stylish clothing. Jared appeared his stated age and his overall presentation was consistent with an independent adolescent. He made little-to-no eye contact, his posture was poor, and he did not converse easily with the examiners. Jared appeared to feel relatively uncomfortable during most of the testing and even though he complied with all requests, it was apparent that Jared retained his sense of privacy and minimal social involvement with the examiner. Jared provided one-word answers whenever possible and never initiated conversation. Jared participated in a considerable amount of testing and there were only two times when he seemed to feel a little more at ease, let his guard down somewhat, and conversed readily with the examiner. The first time that this occurred was during the third appointment when Jared described his relationship with his mother. While discussing their relationship, Jared sat up straight in his chair, made direct eye contact, and spoke with emotion. He stated that his relationship with his mother was strained and that he felt that she continually tries to interfere when he is talking to his father. During this discussion, Jared was able to express himself clearly and he communicated his feelings and thoughts in an age-appropriate manner. The second time that Jared opened up was during the home visit, where Jared was described as pleasant and able to interact well with the home-visit staff member.

Jared arrived for his appointments on time and was cooperative. He spoke softly and did not enunciate well, making it difficult to understand what he was saying at times. Although Jared seemed to be somewhat uninterested, he appeared to be trying his best. He was a little anxious, as evidenced by his excessive psychomotor activity and self-stimulating behavior. For example, Jared continually engaged in the following behaviors: touching his face, rubbing his arm, playing with his fingers, cracking his knuckles, cracking his neck, and tapping his fingers on the underside of his chair. These behaviors did not seem to distract him from what he was doing but rather seemed to soothe him emotionally and/or reduce his anxiety.

While solving problems and answering test questions, Jared spoke in a flat, monotone voice. The tone of his speech made it sound as if he
was bored and not trying; however, his scores and other behavioral observations suggest that he was, in fact, exerting full effort. When Jared felt that he did not know how to do something, or that he did not know the correct answer, he would become very frustrated. For example, he would get a look of disgust on his face or sometimes softly hit his fists on the table. In general, Jared did not respond well to feedback and encouragement from the examiner. In fact, he seemed to not recognize or care that he had received feedback.

Overall, Jared’s affect was blunted, his responses were given in monotone and were brief. Although his affect was flat and he was not very communicative, he worked hard and was compliant. Jared was attentive and was not easily distracted. He appeared to be trying his best; however, when he felt that he was not doing well, he was very hard on himself. For example, he would say “I should have known that one” or he would roll his eyes and sigh as if he were exasperated with himself.

**Tests Administered**

- Wechsler Intelligence Scale for Children—Third Edition (WISC-III)
- Kaufman Brief Intelligence Test (K-BIT)
- Kaufman Test of Educational Achievement: Comprehensive Form (K-TEA)
- Rorschach Inkblot Test
- Incomplete Sentence Test
- Bender Gestalt Visual-Motor Test
- House-Tree-Person Test
- Kinetic Family Drawing Test
- Child Behavior Checklist (CBCL): Administered separately to Mr. and Mrs. P.; Jared responded to the Self-Report version.

**Test Results (See Table 4.5) and Interpretation**

**Cognitive Functioning**

Jared scored in the average range of intelligence on the Wechsler Intelligence Scale for Children-Third Edition (WISC-III), earning a Verbal IQ of 93±5, a Performance IQ of 84±5, and a Full Scale IQ of 88±4. The 9-point discrepancy between his Verbal and Performance IQs is not statistically significant, and indicates that he performs about equally well whether solving verbal problems and expressing his ideas orally or solving nonverbal items via the manipulation of concrete materials.

His overall performance on the WISC-III indicates that he is functioning at a little below average when compared to other children of his approximate age.

Jared was also administered the Kaufman Brief Intelligence Test (K-BIT). The K-BIT consists of two subtests, Vocabulary and Matrices, both of which have been shown to provide basic information about an individual’s cognitive functioning. More specifically, the Vocabulary subtest measures verbal ability and crystallized knowledge while the Matrices subtest measures nonverbal ability and fluid reasoning. Jared’s overall K-BIT Composite IQ was 93±6, 32nd percentile, (average), which is consistent with his WISC-III Composite IQ of 88.

Test results also indicate that Jared has a fluid reasoning strength, relative to his crystallized knowledge. On the K-BIT Vocabulary subtest (crystallized knowledge), Jared earned a standard score of 89±7, 23rd percentile, which is a little below average. In comparison, on the Matrices subtest (fluid reasoning), he earned a standard score of 98±7, 45th percentile, which is average. His fluid-reasoning strength is also evident on his performance on the WJ-R as well as the WISC-III. On the WJ-R, Jared earned his highest score on the fluid-reasoning cluster with a standard score of 108±4, 71st percentile. On the WISC-III, consistent with this fluid strength, Jared scored in the 75th percentile on Object Assembly and in the 63rd percentile.

On the WISC-III, Jared earned his lowest overall scores on two subtests on the Performance Scale: Picture Completion and Picture Arrangement. Picture Completion required him to look at a drawing of an object and/or individual that was somehow
incomplete, and to determine what important part of the picture was missing. Picture Arrangement, on the other hand, required him to look at a series of cards that when placed in the correct order, tell a sequential story about an event or situation. Both Picture Completion and Picture Arrangement involve drawings of people, places, and things and tend to involve human relationships. This type of content appears to be problematic for Jared and causes him discomfort. On these two subtests, it appears that Jared had difficulty solving the problems because he became emotionally overwhelmed with the material, and as a result he had great difficulty organizing his perceptions efficiently and accurately. Jared’s performance on some of the subtests further substantiate this hypothesis. On the WJ-R, Jared earned a standard score of 109±6 (73rd percentile) on Visual Closure, a task that measures his ability to identify a drawing or picture that was altered in some way. For example, the picture may have been distorted, have missing lines or areas, or have a superimposed pattern. In addition to some type of distortion, the pictures or drawings are partially covered up by horizontal lines, making the pictures and drawings appear more distant and abstract. As a result of the abstraction, Jared was able to distance himself and process more effectively. In comparison, Jared earned a standard score of 91±5 (27th percentile) on Picture Recognition, a task that measures the ability to recognize a subset of previously presented pictures within a field of distracting pictures. On this task the pictures are concrete and straightforward, which leads Jared to distort them and perceive them less accurately, perhaps because he viewed them as threatening. Importantly, when reasoning tasks do not involve threatening content, Jared displays quite good ability. He earned a standard score of 108 (70th percentile) on the WJ-R Fluid Reasoning Scale, and performed at a similar level on a WISC-III puzzle-solving test (75th percentile). All of these tasks measure a child’s ability to solve problems that are novel, and not dependent on schooling.

Jared’s average to low average scores earned on the WISC-III, K-BIT, and WJ-R are in stark contrast to the scores that he reportedly earned on the WISC-R that was administered to him in 1989. The WISC-R scores that were reported indicated that Jared’s Verbal IQ was 128, his Performance IQ was 133, and his Full Scale IQ was 135. However, these scores were based on that examiner’s questionable use of a scoring system that involved eliminating some of Jared’s scores on several pertinent subtests of the WISC-R and then calculating an estimate based on an incomplete set of subtests. There was no indication why these important subtests were eliminated, resulting in a prorated IQ; therefore, it is difficult, if not impossible, to determine if these results were valid or meaningful in any way.

To assess Jared’s academic achievement abilities, he was administered the Kaufman Test of Educational Achievement-Comprehensive Form (K-TEA). Jared’s achievement abilities ranged from the 30th percentile on Spelling to the 84th percentile on Mathematics Applications. In general, all of Jared’s scores fell within the average range, although his excellent performance on Mathematics Applications was above average. In addition to the individual subtest scores, the K-TEA provides three composite scores of overall academic functioning. On the Reading Composite, Jared earned a standard score of 103±4, 58th percentile, grade equivalent 8.7, (average). On the Mathematics Composite he earned a standard score of 112±5, 79th percentile, grade equivalent 10.5, (above average). His Battery Composite standard score of 104±3, 61st percentile, was equivalent to grade 8.9 (average). These results indicate that Jared’s academic abilities are generally average and that, based on his cognitive abilities, he is working up to his potential.

The most significant difficulty noted in this cognitive evaluation was Jared’s deficient performance on two-delayed recall tasks. He scored in the 2nd and 3rd percentile on two subtests on the WJ-R that measure incidental retention of previously taught material. During the administration of these subtests, Jared appeared to be paying attention and he was very focused. On the WISC-III, Jared scored in the 75th percentile on the Digit-Span subtest which evaluates short-term memory and retrieval. Based on these results, attentional difficulties do not appear to be the cause of Jared’s specific memory difficulties.

The results from the cognitive and achievement portions of this assessment suggest that Jared is a young adolescent of basically average intelligence. He has average academic achievement and inconsistent long-term memory. There is no indication that he has a learning disability or that he has any significant academic weaknesses. Instead, it appears that Jared’s academic difficulties stem from emotional factors that inhibit and interfere with his ability to do his school work. More impor-
tantly, Jared’s current emotional state is, in general, significantly impeding his cognitive functioning.

**Personality Functioning**

Projective personality testing indicates that Jared appears to approach his world in a careless and unsystematic way. He often spends an insufficient amount of time sorting out the important from the unimportant details of a situation, which leads him to make hasty decisions in an attempt to resolve issues. In general, this style of coping is often associated with anxiety and depression. Jared also has a tendency to approach new situations with a negative and oppositional attitude in an attempt to defend himself from environmental influences that he perceives to be potentially harmful.

Personality assessment also indicates that Jared becomes easily overwhelmed. His thoughts and feelings are not organized in a way to permit their controlled use, which tends to stimulate undeliberate and erratic behavior. In an effort to avoid complexity and ambiguity, he often becomes emotionally constricted, which eventually leads to emotional explosions and emotional lability.

Feelings significantly disrupt Jared’s ability to perceive reality accurately. When Jared experiences either his own or other people’s emotions he becomes overwhelmed and often reacts impulsively. In emotional situations, he is unable to think rationally and reflect upon the information he has received. Jared is much more likely to display emotion and action in such coping situations rather than reflect and think about what is occurring. This behavior often results in Jared losing his temper and ultimately getting into trouble for his actions. It also appears that arousal of his emotions significantly interferes with his work output. This hypothesis is supported by Jared’s school records, which indicate that he has been able to do very little in school and that his grades are poor.

Test results also suggest that Jared has an unusually painful and critical introspective orientation. He has poor self-esteem and has great difficulty relating to others. Jared feels extremely uncomfortable around other people and has trouble empathizing with them; therefore, he limits and controls his emotional connections to others. Jared’s tendency to view himself and others in a somewhat negative light is indicative of depression.

Currently, Jared is unwilling to exert himself intellectually. This unwillingness to put energy into cognitive activity appears to stem from both immaturity and oppositionality. Jared’s approach to some of the testing was relatively immature. He revealed a lack of complexity and an immaturity in his thought processes. His approach to the testing indicates that he is very concrete. His concrete cognitive style was present on the cognitive portion of the testing as well. When asked to answer questions about what one should do in a variety of life situations (e.g., “If you saw a person fall on the street in front of you, what should you do?”), Jared gave simple responses that were not well thought-out. Concrete thinking is often associated with younger clients; however, developmentally, Jared should be beyond this stage. Additionally, Jared is unwilling to interact much with the world because of his stereotyped and negative view of it.

**Diagnostic Summary**

Jared appears to be a very troubled young man. He is easily overwhelmed with emotion and has extremely limited coping abilities. Jared’s inability to deal with his feelings for himself, as well as his feelings toward others, is quite taxing on him. Jared does not seem to be able to handle his current situation effectively and he does not have any idea what he should do or to whom he can turn. Feeling backed into a corner and hopeless is difficult for any individual, but is especially trying for a young adolescent. To make matters worse, Jared’s current emotional state is significantly impeding his cognitive abilities. He is unproductive and disruptive at school, and his emotional overload and lability have resulted in both disturbed interpersonal relationships as well as distorted thought processes.

**Recommendations**

The following recommendations have been made to assist Jared and his parents with Jared’s academic, emotional, and behavioral difficulties. This assessment suggests that Jared’s difficulties are the result of a complex set of variables and dynamics and should be addressed from a multimodal approach.
1. This evaluation suggests that Jared is a very disturbed young man who is in a state of crisis. Therefore, it is recommended that immediate and drastic interventions be made as soon as possible. Jared would benefit from placement in a long-term residential treatment center that would be able to provide him with full-time intensive therapy and treatment in a structured and safe environment.

2. Jared may also benefit from participating in individual therapy with a clinician who is able to provide him with a supportive and trusting relationship. Jared needs to work with a therapist who is trained to work with adolescents, depression, and Conduct Disorder.

3. Jared should continue taking his medication. Medication will help Jared keep his emotional lability under control, which will make it easier for him to focus on other aspects of his life.

4. Mr. and Mrs. P. may want to consider participating in family therapy with their younger daughter, in order to deal with the impact that Jared’s behavior has had on the family. Often, younger siblings emulate their older brothers or sisters and it will be important for his sister to recognize and understand that Jared’s behavior is not ideal and that there are more effective ways of behaving and coping with situations.

5. Jared would benefit from carrying a notebook with him for writing down important information that he needs to remember at a later time. It is sometimes difficult to get into the habit of making notes to oneself; however, Jared needs to be encouraged to do so because of his poor memory.

REFERENCES


