Chapter 5

WECHSLER INTELLIGENCE SCALES

The Wechsler intelligence scales are individually administered, composite intelligence tests in a battery format. They assess different areas of intellectual abilities and create a situation in which aspects of personality can be observed. Each of the different versions of the Wechsler intelligence scales provides three different IQ scores: an overall or Full Scale IQ, a Verbal IQ, and a Performance IQ. More specific factor or index scores also can be calculated using various combinations of subtests. The Wechsler intelligence scales are considered to be among the best of all psychological tests because they have sound psychometric properties and produce information relevant to practitioners. As a result, they have become the most frequently used tests in clinical practice (Camara et al., 2000; Watkins et al., 1995).

TESTING OF INTELLIGENCE: PRO AND CON

The testing of intelligence has had a consistent history of misunderstanding, controversy, and occasional misuse (D. Flanagan et al., 1997; Mackintosh, 1998; Weinberg, 1989). Criticisms have ranged from moral indictments against labeling individuals, to cultural bias, and even to accusations of flagrant abuse of test scores. Although valid criticisms can be made against testing intelligence, such procedures also have a number of advantages.

One of the main assets of intelligence tests is their accuracy in predicting future behavior. Initially, Binet was able to achieve a certain degree of predictive success with his scales, and, since that time, test procedures have become progressively more refined and accurate. More recent studies provide ample support that intelligence tests can predict an extremely wide number of variables. In particular, IQ tests are excellent predictors of academic achievement (see R. Gregory, 1999; Mackintosh, 1998; Neisser et al., 1996), occupational performance (J. Hunter & Schmidt, 1996; F. Schmidt & Hunter, 1998; R. Wagner, 1997), and are sensitive to the presence of neuropsychological deficit (Groth-Marnat, 2002; Groth-Marnat, Gallagher, Hale, & Kaplan, 2000; Lezak, 1995; Reitan & Wolfson, 1993). However, certain liabilities are also associated with these successes. First, intelligence tests can be used to classify children into stereotyped categories, which limit their freedom to choose fields of study. Furthermore, IQ tests are quite limited in predicting nontest or nonacademic activity, yet they are sometimes incorrectly used to make these inferences (Snyderman & Rothman, 1987; Sternberg, 1999). It should also be stressed that intelligence tests are measures of a person's present level of functioning and, as such, are best used for making short-term predictions.

Long-term predictions, although attempted frequently, are less accurate because there are many uncontrolled, influencing variables. Similarly, even short-term academic placements made solely on the basis of an IQ score have a high chance of failure because all the variables that may be crucial for success are not and cannot be measured by an intelligence test. It can sometimes be tempting for test users to extend the meaning of test scores beyond their intended scope, especially in relation to the predictions they can realistically be expected to make.

In addition to predicting academic achievement, IQ scores have also been correlated with occupation, ranging from highly trained professionals with mean IQs of 125, to unskilled workers with mean IQs of 87 (Reynolds, Chastain, Kaufman, & McLean, 1987). Correlations between job proficiency and general intelligence have been highest in predicting relatively more complex jobs rather than less demanding occupations. J. Hunter (1986) reported moderately high correlations between general intelligence and success for managers (.53), salespersons (.61), and clerks (.54). For intellectually demanding tasks, nearly half the variance related to performance criteria can be accounted for by general intelligence (F. Schmidt, Ones, & Hunter, 1992). The use of intelligence tests for personnel selection has demonstrated financial efficacy for organizations (F. Schmidt & Hunter, 1998). In addition, the accuracy of using IQ tests can be incrementally increased by combining the results with integrity tests, work samples, and structured interviews (F. Schmidt & Hunter, 1998).

Another important asset of intelligence tests, particularly the WAIS-III and WISC-III, is that they provide valuable information about a person's cognitive strengths and weaknesses. They are standardized procedures whereby a person's performance in various areas can be compared with that of age-related peers. In addition, useful comparisons can be made regarding a person's pattern of strengths and weaknesses. The WAIS-III, WISC-III, and other individually administered tests provide the examiner with a structured interview in which a variety of tasks can be used to observe the unique and personal ways the examinee approaches cognitive tasks. Through a client's interactions with both the examiner and the test materials, an initial impression can be made of the individual's self-esteem, behavioral idiosyncrasies, anxiety, social skills, and motivation, while also obtaining a specific picture of intellectual functioning.

Intelligence tests often provide clinicians, educators, and researchers with baseline measures for use in determining either the degree of change that has occurred in an individual over time or how an individual compares with other persons in a particular area or ability. This may have important implications for evaluating the effectiveness of an educational program or for assessing the changing abilities of a specific student. In cases involving recovery from a head injury or readjustment following neurosurgery, it may be extremely helpful for clinicians to measure and follow the cognitive changes that occur in a patient. Furthermore, IQ assessments may be important in researching and understanding more adequately the effect on cognitive functioning of environmental variables, such as educational programs, family background, and nutrition. Thus, these assessments can provide useful information about cultural, biological, maturational, or treatment-related differences among individuals.

A criticism leveled at intelligence tests is that almost all have an inherent bias toward emphasizing convergent, analytical, and scientific modes of thought. Thus, a person who emphasizes divergent, artistic, and imaginative modes of thought may be at a distinct disadvantage. Some critics have even stressed that the current approach to intelligence testing has become a social mechanism used by people with similar values to pass on educational advantages to children who resemble themselves. Not only might IQ tests tend to place creative individuals at a disadvantage but also they are limited in assessing nonacademically oriented intellectual abilities (Gardner, 1999; Snyderman & Rothman, 1987). Thus, social acumen, success in dealing with people, the ability to handle the concrete realities of the individual's daily world, social fluency, and specific tasks, such as purchasing merchandise, are not measured by any intelligence test (Greenspan & Driscoll, 1997; Sternberg, 1999). More succinctly, people are capable of many more cognitive abilities than can possibly be measured on an intelligence test.

Misunderstanding and potential misuse of intelligence tests frequently occur when scores are treated as measures of innate capacity. The IQ is not a measure of an innate fixed ability, nor is it representative of all problem-solving situations. It is a specific and limited sample, made at a certain point in time, of abilities that are susceptible to change because of a variety of circumstances. It reflects, to a large extent, the richness of an individual's past experiences. Although interpretation guidelines are quite clear in pointing out the limited nature of a test score, there is a tendency to look at test results as absolute facts reflecting permanent characteristics in an individual. People often want a quick, easy, and reductionist method to quantify, understand, and assess cognitive abilities, and the IQ score has become the most widely misused test score to fill this need.

An important limitation of intelligence tests is that, for the most part, they are not concerned with the underlying processes involved in problem solving. They focus on the final product or outcome rather than on the steps involved in reaching the outcome. They look at the "what" rather than the "how" (Embretson, 1986; E. Kaplan et al., 1999; Milberg et al., 1996). Thus, a low score on Arithmetic might result from poor attention, difficulty understanding the examiner because of disturbances in comprehension, or low educational attainment. The extreme example of this "end product" emphasis is the global IQ score. When the examiner looks at the myriad assortment of intellectual abilities as a global ability, the complexity of cognitive functioning may be simplified to the point of being almost useless. The practitioner can apply labels quickly and easily, without attempting to examine the specific strengths and weaknesses that might make precise therapeutic interventions or knowledgeable recommendations possible. Such thinking detracts significantly from the search for a wider, more precise, and more process-oriented understanding of mental abilities.

A further concern about intelligence tests involves their limited usefulness in assessing minority groups with divergent cultural backgrounds. It has been stated that intelligence-test content is biased in favor of European American, middle-class values. Critics stress that minorities tend to be at a disadvantage when taking the tests because of deficiencies in motivation, lack of practice, lack of familiarity with culturally loaded items, and difficulties in establishing rapport. Numerous arguments against using intelligence tests for the assessment and placement of minorities have culminated in legal restrictions on the use of IQ scores. However, traditional defenses of IQ scores suggest that they are less biased than has been accused. For example, the removal of biased items has done little to alter overall test scores, and IQs still provide mostly accurate predictions for many minorities (see Chapter 2 for a further discussion). The issue has certainly not been resolved, but clinicians should continue to be aware of this dilemma, pay attention to subgroup norms, and interpret minority group IQ scores cautiously (see Lopez, 1997). Finally, many people feel that their IQs are deeply personal pieces of information. They would prefer that others, even a psychologist who is expected to observe confidentiality, not be allowed access to this information. This problem is further compounded when IQ scores might be given to several different persons, such as during legal proceedings or personnel selection.

Intelligence tests provide a number of useful and well-respected functions. They can adequately predict short-term scholastic performance; assess an individual's relative strengths and weaknesses; predict occupational achievement; reveal important personality variables; and permit the researcher, educator, or clinician to trace possible changes in an individual or population. However, these assets are helpful only if the limitations of intelligence tests are adequately understood and appropriately taken into consideration. They are limited in predicting certain aspects of occupational success and nonacademic skills, such as creativity, motivational level, social acumen, and success in dealing with people. Furthermore, IQ scores are not measures of an innate, fixed ability, and their use in classifying minority groups has been questioned. Finally, there has been an overemphasis on understanding the end product of cognitive functioning and a relative neglect in appreciating underlying cognitive processes.

HISTORY AND DEVELOPMENT

During the 1930s, Wechsler began studying a number of standardized tests and selected 11 different subtests to form his initial battery. His search for subtests was in part guided by his conception that intelligence is global in nature and represents a part of the greater whole of personality. Several of his subtests were derived from portions of the 1937 revision of the Stanford-Binet (Comprehension, Arithmetic, Digit Span, Similarities, and Vocabulary). The remaining subtests came from the Army Group Examinations (Picture Arrangement), Koh's Block Design (Block Design), Army Alpha (Information, Comprehension), Army Beta (Digit Symbol-Coding), Healy Picture Completion (Picture Completion), and the Pinther-Paterson Test (Object Assembly). These subtests were combined and published in 1939 as the Wechsler-Bellevue Intelligence Scale. The Wechsler-Bellevue had a number of technical deficiencies primarily related to both the reliability of the subtests and the size and representativeness of the normative sample. Thus, it was revised to form the Wechsler Adult Intelligence Scale (WAIS) in 1955, and another revised edition (WAIS-R) was published in 1981. The 1981 revision was based on 1,880 individuals who were generally representative of the 1970 census and categorized into nine different age groups.

The Wechsler Adult Intelligence Scale-III (WAIS-III) became available in August 1997 and was developed to revise the earlier (1981) WAIS-R. The primary reason for the revision was to update the norms. Additional reasons included extending the age range, modifying items, developing a higher IQ "ceiling" and "floor," decreased reliance on timed performance, developing index/factor scores, creating linkages to other measures of cognitive functioning/achievement, and extensive testing of reliability validity. Despite these changes, many of the traditional features of the WAIS-R were maintained, including the six Verbal subtests and the five Performance subtests. This still enables practitioners to calculate the Full Scale, Verbal, and Performance IQs. An added feature of the WAIS-III is the inclusion of three new subtests, which enables the calculation of four index scores. Thus, the WAIS-III is not merely a renormed "facelift," but also enables the clinician to do more with the different test scores. This might involve being able to assess persons with either greater age or IQ ranges as well as linking scores with the Wechsler Memory Scales or calculating both IQ and index/factor scores.

The above additions and arrangement of subtests represent the most obvious changes on the WAIS-III. Although not as obvious, its restandardization also represents a major development. The sample was composed of 2,450 adults between the ages of 16 and 89. Each of the 13 age groups was composed of 200 participants with the exception of the 80 to 84 and 85 to 89 age groups, which contained 150 and 100 participants, respectively. Gender and ethnicity closely corresponded to the 1995 U.S. Census data. This included a slightly greater number of women than men at the higher age levels to represent the greater proportion of females in this group. European Americans, African Americans, and Hispanics were also represented in each age band according to the 1995 Census data. The sample was selected from all geographical regions in the United States and stratified to represent the different educational levels in each age group.

The original Wechsler-Bellevue Scale was developed for adults, but in 1949, Wechsler developed the Wechsler Intelligence Scale for Children (WISC) so that children from the age of 5 years 0 months could be assessed in a similar manner. Easier items, designed for children, were added to the original scales and standardized on 2,200 European American boys and girls selected to be representative of the 1940 census. However, some evidence shows that Wechsler's sample may have been overrepresentative of children in the middle and upper socioeconomic levels. Thus, ethnic minorities and children from lower socioeconomic levels may have been penalized when compared with the normative group. The WISC was revised in 1974 and standardized on a new sample that was more accurately representative of children in the United States. The WISC-III (Wechsler, 1991) was released in 1991 with the major change being the inclusion of four factor/index scores (Verbal Comprehension, Perceptual Organization, Freedom from Distractibility, and Processing Speed). The new Processing Speed factor has involved the inclusion of the new subtest of Symbol Search along with the older Coding subtest. As with the earlier WISC-R, the standardization and reliability are excellent. The scales were standardized on 2,200 children between the ages of 6 and 16 who closely matched the 1988 census. The sample consisted of 100 boys and 100 girls for each of the different age groups. The new materials are colorful, contemporary, and easy to administer (see review by Little, 1992). The WISC-IV is anticipated to become available in 2003/2004.

In 1967, the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) was first published for the assessment of children between the ages of 4 and 6 years 6 months. Just as the WISC is a downward extension of the WAIS, so the WPPSI is generally a downward extension of the WISC in which easier but similar items are used. Although most of the scales are similar in form and content to the WISC, a number of them are unique to the WPPSI. The WPPSI was revised in 1989 (WPPSI-R; Wechsler, 1989) and again in 2002 (WPPSI-III; Psychological Corporation, 2002).

RELIABILITY AND VALIDITY

WAIS-III Reliability and Validity

The reliabilities for the WAIS-III are generally quite high (Psychological Corporation, 1997). Areas of note are that average split-half reliability for the Full Scale IQ (FSIQ) is .98, Verbal IQ (VIQ) is .97, and Performance IQ (PIQ) is .94. The Index reliabilities were similarly quite high with a Verbal Comprehension of .96, Perceptual Organization of .94, Working Memory of .93, and a Processing Speed reliability of .87. The somewhat lower reliability for the Processing Speed Index is primarily because only two subtests (Digit Symbol-Coding and Symbol Search) were used to calculate this index. It should also be noted that, since these two subtests are speeded tests, it was not appropriate to use split-half reliability, and test-retest reliability was calculated instead. Reliabilities for the individual subtests were, as expected, somewhat lower. The highest reliabilities were for Vocabulary (.93) and Information (.91) with the lowest for Object Assembly (.70) and Picture Arrangement (.74). Average subtest test-retest reliabilities over a 2- to 12-week interval ($M = 34.6$ days) were generally comparable, although slightly lower, than the above split-half reliabilities.

While the above test-retest reliabilities indicate a high degree of temporal stability, there is still some degree of improvement on retesting because of practice effects. The Full Scale IQ was found to increase by 4.5 points, the Verbal IQ increased 2.4 points, and the Performance Scale increased a much larger 6.5 points. These increases are not only statistically significant but may have clinical significance when making inferences about the extent to which real improvement/deterioration has occurred for a particular client. This can be crucial when interpreting either specific performance subtests or scores derived from the performance subtests (i.e., Performance IQ, Processing Speed). Thus, a client who has a Performance IQ increase of 6 points on retesting may not really be improving in his or her everyday functions but will merely be demonstrating practice effects. Indeed, a difference of 15 points (for ages 16 to 54) would be required to infer that there has been an actual improvement in abilities (Kaufman & Lichtenberger, 2002). Research with the WAIS-R indicates that these practice effects can occur up to nine months later even among head-injured patients (see p. 148). However, retest gains have also been found to diminish with age (J. Ryan, Paolo, & Brungardt, 1990).

Because extensive validity studies exist for the WAIS-R, one of the most important initial steps in WAIS-III validation was to determine the comparability between the two tests. Comparability would be expected given that the two versions share 70% of their items. As expected, correlations were found to be a quite high .94, .86, and .93 for the Verbal, Performance, and Full Scale IQs, respectively (Psychological Corporation, 1997). This suggests that the WAIS-III measures essentially the same constructs as the WAIS-R. Noteworthy high correlations between the different subtests were .90 for Vocabulary, .83 for Information, and .82 for Digit Span. In contrast, relatively low correlations were found for Picture Completion (.50), Picture Arrangement (.63), and Object Assembly (.69). Correlations between the WAIS-III and WISC-III for a group of 16-year-olds were also quite high (VIQ = .88, PIQ = .78, FSIQ = .88). The index scores were somewhat more variable (Verbal Index = .87, Perceptual Organization Index = .74, Working

Memory = .50, Processing Speed = .79). The low correlation for Working Memory is most likely because the WAIS-III includes the new Letter-Number Sequencing subtest. In contrast, the WISC-III uses only Arithmetic and Digit Span to determine the Working Memory Index. These sets of correlations indicate a mostly high level of correspondence between the WAIS-III and WAIS-R as well as the WAIS-III and WISC-III.

The WAIS-III has also been found to correlate highly with several standard ability measures (Psychological Corporation, 1997). The Standard Progressive Matrices is an untimed, nonverbal test and, as such, the WAIS-III correlations between the Performance IQ and Perceptual Organization Index were moderately high (.79 and .65, respectively). In contrast (and consistent with the construct that the Standard Progressive Matrices is both untimed and nonverbal), the correlation with the Processing Speed Index was low (.25). The correlation between the WAIS-III and Stanford-Binet IV was .88. Further, high to moderate correlations (typically in the high .60s to .70s) were found between the WAIS-III and the Wechsler Individual Achievement Test (Psychological Corporation, 1997). While beyond the scope of this review, correlations have also supported expected associations with measures of attention and concentration, memory, language, fine motor speed/dexterity, spatial processing, and executive functioning (Psychological Corporation, 1997).

Because the Wechsler Memory Scales-III (WMS-III) and WAIS-III have been more closely linked, it is important to evaluate the extent and manner in which they were related (Psychological Corporation, 1997). Correlations between the WAIS-III IQ/Index scores and WMS-III Index scores have generally ranged from .33 to .77 (Psychological Corporation, 1997, p. 124). The VIQ was found to correlate moderately with both the WMS-III Verbal Memory Index (.71) and Visual Memory Index (.73). However, somewhat low correlations were found between the WAIS-III PIQ and WMS-III Visual Immediate (.39) and Visual Delayed (.44) scores. The strongest correlation was between WAIS-III Working Memory and WMS-III Working Memory (.82), which is expected because they share the Digit Span and Letter-Numbering subtests (Psychological Corporation, 1997, p. 93). This pattern of correlations between the WAIS-III and standard tests of intelligence, achievement, and memory provides support for the convergent and divergent validity of the WAIS-III.

Factor analysis of the WAIS-III has supported the presence of *g* in that most subtests correlate with each other, as well as with the FSIQ at least to a moderate extent (Caruso & Cliff, 1999; Psychological Corporation, 1997). Dividing subtests into four Indexes is supported by current theories of intelligence as well as factor analytic procedures (Saklofske, Hildebrand, & Gorsuch, 2000; Wechsler, 1997a) although the fourth Processing Speed factor was found to be relatively weak (Ward, Ryan, & Axelrod, 2000). Despite this, the Processing Speed factor/index has been found to be particularly sensitive to brain dysfunction (K. Hawkins, 1998). In contrast to a clear four-factor solution, Caruso and Cliff (1999) stress that the two most reliable factors were related to crystallized intelligence (*Gc;* composed primarily of Vocabulary and Information) and fluid intelligence (*Gf;* composed primarily of Digit Span and Matrix Reasoning).

A variety of clinical populations has patterns of deficits in learning, cognition, and memory. It would thus be expected that the WAIS-III would be sensitive to these patterns. This was somewhat supported in that the mean WAIS-III IQ and index scores for Alzheimer's disease patients were lower than expected when compared with their age-related peers. Comparisons among the index scores indicated differential cognitive abilities in that the mean Verbal Comprehension Index was relatively higher (93.0) than either the Processing Speed ($M = 79.6$) or Perceptual Organization ($M = 84.8$) Index. However, it would have been expected that the Working Memory Index would have been somewhat lower than the mean of $M = 87.2$ given the considerable memory complaints among this population. A variety of other neurological disorders (Huntington's disease, Parkinson's disease, traumatic brain injury) found somewhat similar patterns to those with Alzheimer's disease in that verbal abilities were relatively spared (relatively higher VIQ and Verbal Comprehension Index) whereas Processing Speed was lowest. This indicates that the WAIS-III is sensitive to the difficulties these patient populations have with rapidly processing and consolidating information.

Whereas the mean IQ scores for clients diagnosed with attention-deficit hyperactivity disorder (ADHD) did not differ from the standardization sample, the mean Working Memory Index scores were 8.3 points lower than their Verbal Comprehension Index scores (Psychological Corporation, 1997). Similarly, subjects diagnosed with learning disabilities were found to have IQ scores within the normal range (Psychological Corporation, 1997). However, pronounced discrepancies on Index scores were found. Mean Verbal Comprehension scores were 7 points higher than Working Memory scores for reading-disabled subjects and 13 points higher for math-disabled subjects. A subgroup (47.7%) of persons with reading disabilities had at least a 15-point higher mean Verbal Comprehension than Working Memory scores. Discrepancies were further reflected in that mean Perceptual Organization scores were 7 points higher than Processing Speed scores for both math and reading-disabled groups. The ACID profile (lower **A**rithmetic, **C**oding, **I**nformation, **D**igit Span) was also found in that 24% of learning disabled subjects expressed a partial ACID profile and 6.5% expressed a pronounced ACID profile. However, the Verbal Comprehension/Working Memory and Perceptual Organization/Processing Speed discrepancies seemed to more strongly reflect the patterns of cognitive strengths and weaknesses than the ACID profile. This data indicates that the WAIS-III accurately reflected the patterns of deficits related to known characteristics of various clinical and psychoeducational groups.

WISC-III Reliability and Validity

The WISC-III has generally excellent reliability. The average WISC-III internal consistency reported by Wechsler (1991) across all 11 age groups was .96 for the Full Scale IQ, .95 for the Verbal Scale, and .91 for the Performance Scale. Internal consistency for the specific subtests was far more variable, ranging from a low for Object Assembly of .69 to a high of .87 for Vocabulary. The average reliabilities for Verbal subtests ranged between .77 and .87 ($Mdn\ r = .83$), while the Performance subtests were somewhat lower, ranging between .69 and .89 ($Mdn\ r = .78$). However, the reliabilities vary somewhat according to different age levels, with the younger subgroups having lower reliabilities than older groups.

Test-retest reliabilities are likewise quite high for the three IQ scores and somewhat lower for the specific subtests. Full Scale IQ reliability for all ages over a 23-day (median) retesting was .94 and the Verbal and Performance Scales were .94 and .87, respectively (Wechsler, 1991). The average increase in scores for retesting over the 23-day interval was 7 to 8 points for the Full Scale IQ, 2 to 3 points for the Verbal IQ, and 11 to 13 points for the Performance IQ. This can mainly be accounted for by practice effects that seem to be particularly pronounced for the Performance Scale. The practical implication is that clinicians should incorporate the meaning of these short-term increases into their interpretations. Specifically, moderate short-term increases in scores of 5 to 10 points should not usually be considered to indicate true improvement in ability. Longer term retesting for the WISC-R over a two-year interval (which is more typical in clinical settings) has shown somewhat more stability with less than an average three-point difference in Full Scale IQ (Haynes & Howard, 1986). This suggests similar long-term test-retest stability for the WISC-III although no longer term studies are currently available. Test-retest reliabilities for the specific subtests ranged from a high of .89 for Vocabulary to a low of .69 for Object Assembly with an overall median of .76.

The standard error of measurement (indicated in IQ points) for the Full Scale IQ was 3.20, Verbal IQ was 3.53, and Performance IQ was 4.54. The standard error of measurement (given in subscale scores) for the Verbal subtests ranged from 1.08 to 1.45, with the narrowest range of error for Vocabulary (1.08) and the widest for Comprehension (1.45). The Performance subtests ranged from 1.11 to 1.67, with the narrowest range for Block Design (1.11) and widest for Object Assembly (1.67) and Mazes (1.64). Further information for incorporating specific standard error of measurement scores into WISC-III (and WAIS-III) interpretations is included in the Interpretation Procedures section.

The underlying factor structure, while still somewhat controversial, has generally supported Wechsler's conceptualization of abilities into a Verbal Comprehension factor that roughly corresponds with the Verbal Scale, and a Perceptual Organizational factor that generally corresponds with the Performance Scale (Kaufman, 1975, 1994; Sherman, Strauss, Slick, & Spellacy, 2000). More importantly, four factors have emerged from the WISC-III comprising Verbal Comprehension, Perceptual Organization, Freedom from Distractibility (Working Memory), and Processing Speed (Grice, Krohn, & Logerquist, 1999). This is comparable to the factors identified on the WAIS-III and also allows an attractive means of interpreting more specific aspects of intelligence than can be found using only IQ scores.

Given the high degree of item overlap, subtest correlations, and IQ score correlations between the WISC-R and WISC-III, much of the extensive validity research on the WISC-R can be generalized to the WISC-III (Dixon & Anderson, 1995). This validity relates primarily to extensive correlations with relevant criterion measures, including other ability tests, school grades, and achievement tests. Selected median correlations reviewed and reported by Sattler (2001) include those for the Stanford-Binet: Fourth Edition (.78), K-ABC (.70), group IQ tests (.66), WRAT (.52 to .59), Peabody Individual Achievement Test (.71), item overlap with the WPPSI-R (Sattler & Atkinson, 1993), and school grades (.39).

ASSETS AND LIMITATIONS

Since their initial publication, the Wechsler intelligence scales have been used in numerous research studies and have become widely used throughout the world. Thus, they are familiar to both researchers and practitioners and also have a long and extensive history of continued evaluation. This enormous research base allows practitioners to make relatively accurate predictions regarding clients. Inconsistencies between an individual's performance and relevant research can also be noted, alerting the practitioner that he or she needs to develop and pursue further hypotheses. Furthermore, the subtests are relatively easy to administer, and the accompanying manuals provide clear instructions, concise tables, and excellent norms.

Norms for both the WAIS-III and WISC-III represent a further clear strength. The size is adequate and, for the most part, has corresponded to the demographics of the U.S. census. Cross-national use has been developed through research on how residents in other countries perform. Oversampling on the WAIS-III was done for African American and Hispanics as well as on a wide range of educational and ability levels to better understand how these groups perform. A further important feature is that the WAIS-III was co-normed with the Wechsler Memory Scale-III (WMS-III) and the Wechsler Individual Achievement Test (WIAT). This means that a high degree of confidence can be placed in comparing scores among these three different tests. Finally, the WAIS-III has extended its age range to include the performance for persons in the 74 to 89 range. This is an important feature given the increases in knowledge related to this age group along with the expanding number of persons over 65. One of the findings, for example, is that the Perceptual Organization and Processing Speed factors do not appear to be separate constructs for the 74 to 89 group.

Perhaps of even more practical importance to the clinician is the clear, precise data obtained regarding the person's cognitive functioning from the IQ, index, and subtest scores. For example, high scores on the Verbal Comprehension Index indicate good verbal abilities and that the person has benefited from formal education. In contrast, a low score on Processing Speed suggests the person would have a difficult time processing information quickly. A clinician can become extremely sensitive to the different nuances and implications of various patterns of scores. Thus, many of these interpretive guidelines, particularly for the IQ and index scores, have substantial theoretical and empirical support.

A final, but extremely important, asset of the Wechsler scales is their ability to aid in assessing personality variables. This can be done by directly observing the individual as he or she interacts with the examiner, studying the content of test item responses, or evaluating information inferred from the individual's pattern of subtest scores. For example, a person scoring low on Digit Span, Arithmetic, and Digit Symbol is likely to be experiencing anxiety, to have an attentional deficit, or a combination of both. On the other hand, it might be hypothesized that a person who scores high on both Comprehension and Picture Arrangement is likely to have good social judgment. Despite attempts to establish descriptions of the manner in which different clinical groups perform on the Wechsler intelligence scales, few clear findings have emerged (Piedmont, Sokolove, & Fleming, 1989a, 1989b). Thus, the Wechsler scales should not be seen as similar to "personality scales" or "clinical scales." Rather, the subject's subtest patterns, behavior surrounding the test, and qualitative responses to the items should be considered as a means of generating hypotheses related to personality. In this context, the Wechsler intelligence scales are noteworthy in the degree to which they can provide personality variables and clinical information.

One significant criticism leveled at the Wechsler scales has been their lack of data supporting their ecological (or everyday) validity (Groth-Marnat & Teal, 2000; Reinecke, Beebe, & Stein, 1999; Sbordone & Long, 1996). This is particularly important as referral questions are increasingly related to a client's everyday levels of functioning (i.e., extent of disability, ability to function independently, everyday aspects of memory). Although the Wechsler scales have been correlated with other measures, including the Stanford-Binet and academic achievement, for the most part, there has been a notable lack of comparisons with behavior external to the scales themselves. This is true despite the belief that many significant areas of a person, such as adaptive behavior, personal competence, or need for achievement, are separate (but related) constructs (Greenspan & Driscoll, 1997; Sternberg et al., 1995). In particular, the meanings associated with subtest scores should be investigated in more depth. For example, Picture Completion has traditionally been considered a measure of a person's ability to distinguish relevant from irrelevant details in his or her environment, yet this assumption has not been adequately tested. Likewise, no studies have been made to determine if high or low Digit Span scores relate to actual day-by-day behaviors, such as recalling telephone numbers, facility with computer programming sequences, or following directions.

An extension of this concern is that a number of authors have criticized what they believe is an overinterpretation of subtest and index scores (Glutting, McDermott, Konold, Snelbaker, & Watkins, 1998; Konold, Glutting, McDermott, Kush, & Watkins, 1999; MacMann & Barnett, 1997). Specifically, they believe that individual subtest reliabilities are too low and not sufficiently specific for interpreting individual profiles. For example, they note that, compared with *g* (as represented by the Full Scale IQ), WISC-III index scores do not account for a sufficient proportion of the variance in predicting achievement. As a result, index interpretation does not demonstrate sufficient incremental increases in prediction. In addition, the ipsative patterns of subtest strengths and weaknesses are not sufficiently stable over time (MacMann & Barnett, 1997). Clinicians might, therefore, be advised to rely on the Full Scale IQ rather than index scores when making academic (and possibly other) predictions. Various authors counter this by emphasizing the importance of hypothesis testing, combining interpretations with external criteria, and noting the conceptual importance of the complexity of intelligence (A. Kaufman, 1994, 1999; A. Kaufman & Lichtenberger, 2000, 2002; Lezak, 1988, 1995; Milberg et al., 1996).

There are several additional limitations to the Wechsler scales. Some critics believe that norms may not be applicable for ethnic minorities or persons from lower socioeconomic backgrounds. In addition, the complexity of scoring, particularly the numerous calculations required for the WAIS-III, is likely to increase the probability of clerical errors by examiners (Slate & Hunnicutt, 1988; Slate, Jones, & Murray, 1991). A further potential difficulty is that when supplementary subtests are substituted for regular subtests, it is unclear how these supplementary subtests will affect the three IQ or index scores. As a result, supplementary subtests should be given only under unusual circumstances, such as when one of the regular subtests has been "spoiled."

A further issue is that there is a certain degree of subjectivity when scoring many of the items on Comprehension, Similarities, and Vocabulary. Thus, a "hard" scorer may develop a somewhat lower score than an "easy" scorer. This is particularly true for

Similarities, Comprehension, and Vocabulary, where scoring criteria are less clear than for other subtests. The Wechsler scales, like other tests of intelligence, are also limited in the scope of what they can measure. They do not assess important factors, such as need for achievement, motivation, creativity, or success in dealing with people.

It should finally be noted that the WAIS-III and WISC-III have continued the traditional measurement of intelligence as represented by the Stanford-Binet scales and the earlier versions of the Wechsler scales. Although their revisions have provided features such as updated norms and index scores (especially the inclusion of Working Memory/Freedom from Distractibility and Processing Speed), the underlying theories and essential construction of these scales have remained relatively unchanged for well over 50 years. This is despite numerous developments in both theory and measurement. These include Luria's PASS (**P**lanning-**A**ttention-**S**uccessive-**S**equencing: Luria, 1980) model, Gardner's independent competencies (Gardner, 1999), various theories on emotional intelligence (Bar-On, 1998; Ciarochi, Chan, & Caputi, 2000), and commonsense problem solving (Sternberg et al., 1995). Thus, one criticism of the Wechsler intelligence scales is that they have not responded to more current views on intelligence (Kaufman & Lichtenberger, 2002; Sternberg, 1999; Sternberg & Kaufman, 1998; Styles, 1999). It remains to be seen whether newer models and assessment tools will have much of an impact on assessing either intelligence or, especially, the frequency to which the Wechsler scales will be used in this process.

MEANING OF IQ SCORES

Because only a weak and vague relation exists between theories of intelligence and the Wechsler intelligence scales, it is important for all persons involved with testing to understand the meaning of IQ scores. Untrained persons are particularly likely to misinterpret IQ scores, which may result in poor decisions or negative attitudes toward either the client or the testing procedure itself. The meaning of IQ scores can be partially clarified by elaborating on some of the more common misinterpretations. IQ is often incorrectly believed to be fixed, unchangeable, and innate. Although there does tend to be considerable stability of IQ scores throughout adulthood ($r = .85$; Schuerger & Witt, 1989) it is possible for changes in IQ to occur, particularly among children (Perkins & Grotzer, 1997). For example, the greatest longitudinal increases in IQs occurred among children who were from homes that provided strong encouragement and avoided severe forms of punishment (McCall, Appelbaum, & Hogarty, 1973). Similarly, Sameroff, Seifer, Baldwin, and Baldwin (1993) found that multiple environmental risk factors (e.g., number of major stressful events, mother's mental health) were able to predict one-third to one-half of IQ variance between the ages of 4 and 13. In addition, education can increase aspects of IQ primarily related to crystallized intelligence even among adults. Thus, IQ can be related to a number of environmental influences. Second, IQ scores are not exact, precise measurements; rather, they are estimates in which there is an expected range of fluctuation between one performance and the next. Finally, tests such as the Wechsler scales measure only a limited range of abilities, and a large number of variables usually considered "intelligent" are beyond the scope of most intelligence tests. No test or battery of tests can ever give a complete picture; they can only assess various areas of functioning. In summary, an IQ is an estimate of a person's current level of functioning as measured by the various tasks required in a test.

An assumption of any global IQ score is that it derives from a wide array of interacting abilities. A subtest such as Information assesses specific areas of a person's range of knowledge and is related to general intelligence. IQ scores are also influenced by achievement orientation, curiosity, culture, and the person's interests. More general prerequisites are that the client must comprehend what has been requested, be motivated to do well, follow directions, provide a response, and understand English. Factors such as persistence and drive are also likely to influence any type of task presented to the person. The tasks included in IQ tests are those, based on judgments by psychometrists, most valued by Western society. In other words, they relate to and are predictive of relevant skills outside the testing situation. It is certainly possible to test a much wider range of areas (as in Guilford's Structure of Intelligence), but either these are not routinely done, or many potential measures may be of little relevance to predicting academic achievement or vocational performance.

Despite the many relevant areas measured by IQ tests, practitioners need to observe some humility when making predictions based on them. Many persons with quite high IQs achieve little or nothing. Having a high IQ is in no way a guarantee of success, but merely means that one important condition has been met. In contrast, persons with relatively low IQs have more severe limitations placed on them. As a result of their relatively narrower range of options, predictions regarding their behavior tend to be more accurate. However, it is possible that persons with average or below average WAIS-III/WISC-III IQs may have high levels of interpersonal, practical, or emotional "intelligence," which may help them compensate for lower levels of formal intelligence.

Regardless of the person's IQ range, clinicians should be clear regarding the likely band of error (standard error of measurement). It is often useful to include the standard error of measurement in a report. For example, the WAIS-III Full Scale IQ has an average standard error of measurement of 2.30 (Psychological Corporation, 1997). Thus, a particular IQ has a 95% chance of being within ±5 IQ points of a person's obtained IQ. The WISC-III has a slightly higher average standard error of measurement of 3.20 for the Full Scale IQ, 3.53 for the Verbal IQ, and 4.54 for the Performance IQ (Wechsler, 1991). Error can also be the result of unforeseen events beyond the context of IQ tests. Although 50% to 75% of the variance of children's academic success is dependent on nonintellectual factors (persistence, personal adjustment, curiosity), most of a typical assessment is spent evaluating IQ. Some of these nonintellectual areas might be quite difficult to assess, and others might even be impossible to account for. For example, a student might unexpectedly develop an excellent relationship with a teacher, which significantly changes his or her attitude toward school, thereby stimulating his or her interest to passionately pursue a specific area. Thus, any meaning attached to an IQ score should acknowledge the possible effects of uncertainty both in the measurement itself and from the wider context of the person's life.

Another important aspect of IQ is the statistical meaning of the different scores. Binet originally conceptualized intelligence as the difference between a person's mental age and his or her chronological age. This was found to be inadequate and has been replaced by the use of the deviation IQ. The assumption behind the deviation IQ is that intelligence falls around a normal distribution (see Figure 5.1). The interpretation of an IQ score, then, is straightforward because it gives the relative position of a person

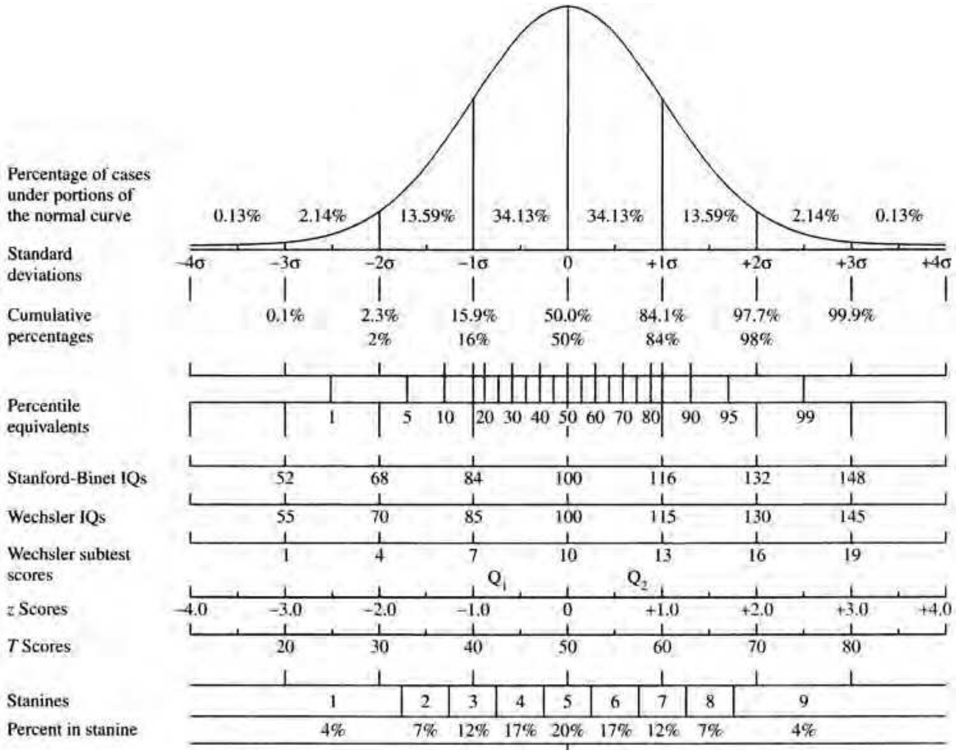

Figure 5.1 Relationship of Wechsler scores to various types of standard measures

compared with his or her age-related peers. The IQ can thus be expressed in deviation units away from the norm. Each of the three Wechsler IQs and four indexes has a mean of 100 and a standard deviation of 15. Scores also can be easily translated into percentile equivalents. For example, an IQ of 120 is 1.33 standard deviations above the mean and places an individual in the ninety-first percentile (see Appendix B on p. 677). Thus, this person's performance is better than 91% of his or her age-related peers. The IQ cutoff for mental retardation is 70, which indicates that such individuals are functioning in the lowest 2% when compared with their age-related peers. Appendix B can be used to convert Wechsler IQ scores ($M = 100$, $SD = 15$) into percentile rankings.

A final consideration is the different classifications of intelligence. Table 5.1 lists commonly used diagnostic labels and compares them with IQ ranges and percentages. These terms are taken from the 1997 WAIS-III manual. Thus, an IQ can be expressed conceptually as an estimate of a person's current level of ability, statistically as a deviation score that can be transformed into percentile equivalents, and diagnostically using common terms for classification.

CAUTIONS AND GUIDELINES IN ADMINISTRATION

The Wechsler manuals generally provide quite clear guidelines for administration and scoring. Despite this clarity, the number of administration and scoring errors on the

Table 5.1 Intelligence classifications

WAIS-III/WISC-III	More Value-Neutral Terms	Corresponding IQ Range
Very superior	Higher extreme above average	130+
Superior	Well above average	120–129
High average	High average	110–119
Average	Average	90–109
Low average	Low average	80–89
Borderline	Well below average	70–79
Extremely low (WAIS-III) or intellectually deficient (WISC-III)	Lower extreme	69 and below

Source: The classification systems of the WAIS-III are from Wechsler, 1997, Table 2.3, and for the WISC-III are from Wechsler, 1991, Table 2.8. Percentile ranks can be determined by consulting Appendix A.

part of trainee and experienced clinicians is far higher than they should be (Alfonso, Johnson, Patinella, & Radar, 1998; Moon, Blakey, Gorsuch, & Fantuzzo, 1991; Moon, Fantuzzo, & Gorsuch, 1986; Slate & Hunnicutt, 1988; Slate et al., 1991). Because the WAIS-III has a far greater number of calculations than the WAIS-R (or WISC-III), the likelihood (even probability) of clerical errors is significantly increased. One way of reducing clerical errors is to use the computer scoring software developed by The Psychological Corporation (i.e., Scoring Assistant for the Wechsler Scales for Adults, WAIS-III/WMS-III Writer, Scoring Assistant for the Wechsler Scales, WISC-III Writer). Even with repeated administration of the Wechsler scales, often examiners end up "practicing their mistakes" rather than correcting them (Slate et al., 1991). The causes of these errors include lack of proper instruction, lack of clarity between academic versus clinical site regarding where training is supposed to occur, carelessness, variations in the quality of the examiner-examinee relationship, and work overload for clinicians (Slate & Hunnicutt, 1988). One approach to reducing errors is awareness regarding the most frequent general categories of errors. These have been investigated by Slate et al. and the most common errors, in order of frequency, are as follows:

1. Failing to record examinee responses, circle scores, or record times (errors of administration).
2. Assigning too many points to an examinee's response (leniency by examiner).
3. Failing to question when required by test manual (poor reading and recalling of information in the manual).
4. Questioning examinee inappropriately (poor reading and/or incorrect integration of the manual).
5. Assigning too few points when required by test manual (examiner too hard).
6. Incorrectly converting raw score to standard score (clerical error).
7. Failing to assign correct points for Performance items (clerical and timing error).
8. Incorrectly calculating raw score for subtest totals (clerical error).
9. Incorrectly calculating chronological age (clerical error).

Whereas the preceding list covers quite general categories, the following list, adapted from Moon et al. (1991), includes a series of recommendations based on the most frequently occurring errors but does so by listing concrete and specific recommendations:

1. Recite digits (on Digit Span) and digits and letters (on Letter Number Sequencing) at the rate of one per second with the pitch of the voice dropping on the last digit/letter of each trial.
2. State during the introduction that each task begins with easy questions and ends with difficult ones. Examiners may also note that not everyone is expected to succeed on all problems.
3. Record responses verbatim on Vocabulary. At times, the examinee provides so much detail that this is not possible, but the essential components should be written down verbatim. This can be facilitated by the use of abbreviations.
4. Properly orient blocks (on Block Design) at examinee's midline.
5. The first time the examinee points out a nonessential part on Picture Completion, the examiner should comment, "Yes, but what is the most important thing missing?"
6. Attempt to elicit the examinee's perception of the testing situation and correct any misconceptions.
7. Check to see if the examinee is comfortable.

Despite clear guidelines in the manual as well as awareness of frequent errors, examiners are still likely to make mistakes. Thus, optimal training guidelines should be incorporated into graduate programs and continuing education. A recommended format is the Mastery Model, which involves the following steps: (a) 1 to 2 hours studying the manual, (b) viewing a videotape of a flawless WAIS-III/WISC-III administration, (c) viewing a videotaped lecture of major pitfalls of administration, (d) successfully detecting errors in a videotaped flawed WAIS-III/WISC-III administration, (e) actually administering the WAIS-III/WISC-IIIs to be evaluated by a rating device such as Sattler's (2001) "Administrative Checklist for the WAIS-III" (pp. 398–405) or "Administrative Checklist for the WISC-III" (pp. 243–248). Such procedures are likely to significantly shorten the length of training time, number of training administrations, and yet significantly increase the level of competence related to Wechsler scale administration and scoring (Moon et al., 1986; Slate et al., 1991).

The WAIS-III manual indicates that the average administration time to determine the Full Scale IQ for the standardization sample was 75 minutes (and 60 minutes for administering all subtests required to calculate only the indexes). In contrast, Ryan, Lopez, and Werth (1998) found that, for a heterogeneous clinical population, the average time was 91 minutes (and 77 minutes for subtests used to determine the indexes). Time estimates for the WISC-III standardization sample were 50 to 70 minutes (and the three supplementary subtests added an additional 10 to 15 minutes; Wechsler, 1991). These times were for administration only and did not include time required for scoring, breaks, or interpretation. The practical implications of this are that clinicians typically need to allocate more time for assessing clinical populations than might be inferred from reading the manual. Clients who fatigue easily may also need to have the

Wechsler intelligence scales administered over two sessions. Finally, clinicians should make realistic appraisals of required times and use these estimates to make sure that they are appropriately compensated.

WAIS-III/WISC-III SUCCESSIVE LEVEL INTERPRETATION PROCEDURE

The following successive-level approach to interpreting Wechsler scores represents an integration and synthesis of the approaches outlined by major resources in the field (A. Kaufman & Lichtenberger, 1999, 2000, 2002; J. H. Kramer, 1993; Naglieri, 1993; Sattler, 2001). This approach provides clinicians with a sequential, five-level format for working with and discussing a person's performance. The underlying purpose for each of these steps should be based on confirming, disconfirming, or altering hypotheses based on patterns of scores combined with relevant background information. The next section of this chapter ("Wechsler Subtests") covers descriptions of the Wechsler subtests, including the more frequently encountered abilities associated with these subtests. This section can serve as a summary and quick reference for clinicians, especially in analyzing test profiles (Levels II and III).

Examiners who are relatively unfamiliar with the Wechsler scales are likely to find the level of detail in the following interpretation procedure and Wechsler subtest sections somewhat daunting because of its complexity. It is thus recommended that they initially read the interpretation procedures to gain familiarity with the material. It might be particularly helpful to review the summary of these procedures in Table 5.2, both before and after reading this section. Table 5.2 can also serve as a useful future quick reference guide when actually working with Wechsler protocols. After perusing the "Interpretation Procedures" section, student examiners should next obtain a completed WAIS-III/WISC-III profile, preferably one they themselves have administered, and then work through the levels of interpretation in a sequential manner. This should add the required level of clarity and integration of the material to begin to work more confidently with future protocols.

The following are principles to keep in mind when working through the interpretation procedures:

- The successive steps begin with the most general aspects of the WAIS-III/WISC-III (Full Scale IQ) and gradually work their way to more specific aspects of the person's performance (indexes, additional groupings, subtest scatter, qualitative responses to individual items, etc.).
- Examiners can interpret the more global measures (Full Scale, Verbal, and Performance IQs) with greater meaning, usefulness, and certainty if there is not a high degree of subtest scatter. With increasing subtest scatter, the purity of the global measures becomes contaminated so that interpretations of these global measures become less meaningful. For example, if the Verbal scales display a pattern in which the Verbal Comprehension and Working Memory/Freedom from Distractibility Indexes are significantly different from each other, it makes more sense to focus on these two indexes rather than the more global Verbal IQ.

Table 5.2 Summary of successive five-level WAIS-III/WISC-III interpretive procedures

Level I. Interpret the Full Scale IQ

Determine percentile rankings and IQ classification (see Appendix B and Table 5.1)

Level II. Interpret Verbal-Performance, Factor Scores, and Additional Groupings

a. Verbal-Performance IQs.
   Interpret if V-P discrepancy is significant (9 points for the WAIS-III and 12 points for the WISC-III).
b. Factor Scores: Verbal Comprehension, Perceptual Organization, Working Memory/Freedom from Distractibility, Processing Speed.
   Interpret if significant discrepancies occur between the mean of the four WAIS-III/WISC-III factor scores.
c. Additional Groupings: Bannatyne's Categories, ACID/SCAD profiles, Horn groupings, Fuld profile (see Appendixes C and D).
   Interpret if significant differences occur between means of groupings and individual grouping/category.

Level III. Interpret Subtest Variability (Profile Analysis)

a. Determine whether subtest fluctuations are significant:
   1. Decide appropriateness of using full scale versus verbal and/or performance subtest means; calculate relevant means.
   2. Calculate the difference scores between subtests and relevant means.
   3. Determine whether the difference between subtest score and scale means is significant (see Appendix E for WISC-III or "Score Conversion" page on WAIS-III Record Form).
   4. Indicate on profile as either a strength or a weakness.
   5. Repeat steps 1–5 for each relevant subtest.
b. Develop hypotheses related to the meaning of subtest fluctuations (Appendix F).
c. Integrate subtest hypotheses with additional information.

Level IV. Analyze Intrasubtest Variability.

Level V. Conduct a Qualitative Analysis.

- The recommended level set to establish significant difference is the .05 level. This is true for differences through all levels of interpretation including Verbal-Performance, indexes, additional groupings, and subtest differences. It was felt that this level of significance is sufficiently rigorous for clinical purposes. If either less stringent ($p = .15$) or more stringent ($p = .01$) levels are desired, relevant tables can be found in Wechsler (1991, 1997a) and Kaufman and Lichtenberger (1999, 2000, 2002). When possible, Bonferroni corrections have been included to correct for the possible statistical error resulting from inflation of significant results because of the number of comparisons.
- To determine whether index/factor scores are significantly (.05 level) discrepant, tables are consulted in the manuals (Table B.1, p. 205 in the *WAIS-III Administration and Scoring Manual* and Table B.1 in the *WISC-III Manual*). Thus, *comparisons are made between the different pairs of IQs/indexes/factors.*
- In contrast to the previous system, subtest fluctuations are based on *comparisons with mean scores.* One strategy is to compare the scaled score of each individual

subtest with the mean for all the subtests administered (and then calculate the difference that the subtests fluctuate from this mean to see if the difference is significant). A slightly different strategy is to compare each individual Verbal subtest with the overall mean for all Verbal subtests and also compare each individual Performance subtest with the overall mean for all Performance subtests. This latter method is appropriate if there is a significant difference between the Verbal and Performance IQs (9 points for the WAIS-III and 12 points for the WISC-III).

- Any interpretations, especially those related to the more specific levels (Levels III, IV, and V), should be considered as tentative hypotheses requiring support from additional sources of information (behavioral observations, school records, etc.). Preferably, each hypothesis should be supported by at least two additional sources. This process of hypothesis generation, confirmation/disconfirmation, and integration with other sources is not merely a statistical procedure but also involves considerable clinical wisdom and judgment.

Level I. The Full Scale IQ

An examinee's Full Scale IQ should be considered first because it provides the basis and context for evaluating other cognitive abilities. It is generally the single most reliable and valid score. The Full Scale IQ gives the person's relative standing in comparison with his or her age-related peers and provides a global estimate of his or her overall mental abilities. It is often useful to transform the Full Scale IQ into a percentile rank (see Appendix B) or intelligence classification (see Table 5.1). This is especially important when relating test results to untrained persons because both percentile rank and intelligence classifications are usually less subject to misinterpretation than IQ scores. Many examiners also prefer to include the standard error of measurement (SEM) as an estimate of the confidence that can be placed in the obtained score. For example, a WAIS-III Full Scale IQ of 110 has a 95% probability of falling between 105 and 115 IQ points. This clarifies that the IQ score is not a precise number but is rather a range with an expected margin of error. One classification, "Borderline," might potentially be misinterpreted, because it might be confused with the *DSM-IV* psychiatric diagnosis of *Borderline Personality.* Examiners might counter this by clarifying in parentheses that the "Borderline" range can also be described as "Well below Average."

Although the Full Scale IQ is the most stable and well-validated aspect of the Wechsler scales, its significance becomes progressively less important as the fluctuations increase between Verbal and Performance IQs, with high fluctuations between the factor scores, or when there is a high degree of subtest scatter. When such fluctuations occur, it is incumbent on the examiner to work in more detail to extract the significance of these relative strengths and weaknesses. The next four successive levels of interpretation provide a sequential method of accomplishing this goal.

Level II. Verbal-Performance IQs, Indexes, and Additional Grouping

The second level of interpretation is to consider Verbal and Performance IQs, index scores, and additional groupings. The .05 level of significance is consistently used to determine if fluctuations are significant. In some cases, procedures and formulas are provided to determine the significance of various fluctuations and to convert scores into the familiar IQ-related standard scores having a mean of 100 and standard deviation of 1 (see summary of formulas in Appendixes C and D on pp. 678 and 679).

Step IIa. Verbal-Performance IQs

The Verbal IQ is an indication of an individual's verbal comprehensive abilities, while the Performance IQ provides an estimate of his or her perceptual organizational abilities. However, clinicians should be aware that a pure test of verbal comprehension or perceptual organization does not exist. A seemingly simple task, such as repeating numbers, involves not only verbal comprehension but also adequate rapport, ability to concentrate, number facility, and adequate short-term memory.

One of the central principles behind interpreting Verbal and Performance IQs is that there needs to be a significant difference between them. If such a difference occurs, an explanation for these differences should be developed. An area of difficulty (and controversy) lies in deciding what should be considered an interpretable difference. On the WAIS-III, a significant difference (at the .05 level) is 9 or more points. On the WISC-III, a 12-point difference is significant at the .05 level. It should still be noted that a full 18% of the WAIS-III and 24% of the WISC-III standardization samples obtained Verbal-Performance differences of 15 points or greater (Wechsler, 1981, 1991, 1997b; Psychological Corporation, 1997). This means that, although differences in the range of 9 to 15 points are statistically significant, they are still fairly common occurrences. The difference, then, may represent merely useful information rather than "lurking pathology." The possibility of pathology is far more likely with a 22-point discrepancy on the WAIS-III (occurred in only 5% of the standardization sample) or a 25-point discrepancy on the WISC-III (occurred in only 5% of the standardization sample).

Interpreting the magnitude of and meaning behind Verbal-Performance differences should always occur in the context of what is known about the person (particularly age and education) as well as his or her condition. For example, persons from higher socioeconomic backgrounds or with higher IQs are likely to have verbal scores significantly higher than their performance scores (R. A. Bornstein, Suga, & Prifitera, 1987; Wechsler, 1997a). In contrast, unskilled workers are more likely to have higher performance scores relative to verbal. If these trends are reversed (e.g., an attorney with higher performance scores or unskilled worker with higher verbal than performance scores), the importance of such a result becomes greater. The major variables influencing, and possible meanings behind, Verbal-Performance score differences are summarized in the sections "Verbal Scales" and "Performance Scales," as well as in the sections on special populations. However, Sattler (2001) has summarized the possible general meanings associated with such differences as relating to cognitive style, patterns of interests, sensory deficits, psychopathology (such as emotional disturbance or brain damage), deficiencies/strengths in information processing, or deficiencies/strengths in ability to work under pressure. After an interpretation has been made, practitioners can eventually work to develop implications and instructional recommendations for high and low scores (for Full Scale, Verbal, and Performance IQs, and Working Memory/Freedom from Distractibility).

Under certain conditions, even statistically significant differences between Verbal and Performance IQs can be considered meaningless. The first condition occurs when the Verbal scale splits into significant differences (10 points or more for the WAIS-III; 13 points or more for the WISC-III) between the Verbal Comprehension and Working Memory/Freedom from Distractibility Indexes. This separation means that the Verbal scales do not cohere into a unitary construct. Thus, it may make more sense to interpret the two index scores separately (see interpretative level IIb). Similarly, if a significant difference (13 points for the WAIS-III; 15 points for the WISC-III) exists between Perceptual Organization and Processing Speed, it means that the Performance IQ is not coherent. Accordingly, the two indexes should be interpreted separately. A further condition that might render the Verbal and/or Performance IQs meaningless occurs when there is a high degree of subtest scatter in general (WAIS-III Verbal subtest range 8+ points, Performance 8+ points; WISC-III Verbal range 7+ points, Performance range 9+ points). This happens because the intent of the three IQs represents a unitary construct in which the person's Full, Verbal, or Performance IQs are general, integrated means of functioning. In contrast, high subtest scatter attacks the unitary, integrated nature of the IQs. It is then the examiner's task to work with the relative high and low combinations of subtests to make sense of the person's intellectual strengths and weaknesses. These steps are outlined in Levels IIb, IIc, and III. However, before continuing to an interpretation of subtest scatter, important clusters of subtests might be found through a consideration of indexes and additional groupings.

Step IIb. Index Scores

The WAIS-III and WISC-III have both been found to have four different index scores. Their interpretation provides an empirically and conceptually based means of understanding more detailed aspects of the person's intellectual functioning (see interpretive summaries in Appendix E). As with the Verbal and Performance Scales, the indexes should be interpreted only if discrepancies between the highest and lowest subtests comprising the indexes are not too large (WAIS-III: Verbal Comprehension, 5 points; Perceptual Organization, 6 points; Working Memory, 6 points; Processing Speed, 4 points. WISC-III: Verbal Comprehension, 7; Perceptual Organization, 8; Freedom from Distractibility, 4; Processing Speed 4).

The Verbal Comprehension Index (WAIS-III: Vocabulary, Similarities, and Information; WISC-III: Information, Similarities, Vocabulary, and Comprehension) is a purer, more refined measure of verbal abilities than the Verbal Scale itself. This has been achieved by excluding Digit Span and Arithmetic (and Letter-Number Sequencing on the WAIS-III), which focus primarily on sequential processing, number ability, attention, and working memory rather than strictly verbal abilities. The material presented to them is in the form of oral questions that they need to answer. As such, an examinee's score on Verbal Comprehension reflects the extent to which they understand the meanings of words, can conceptualize verbal information, the extent of factual knowledge related to verbal material, and their ability to adequately express the material in words.

The Perceptual Organization Index (POI) is likewise a somewhat purer measure of perceptual abilities (WAIS-III: Picture Completion, Block Design, Matrix Reasoning; WISC-III: Picture Completion, Picture Arrangement, Block Design, and Object

Assembly). Perceptual Organization is less a measure of processing speed than the Performance Scale because only one subtest (Block Design) emphasizes speed. An examinee's score reflects the extent to which he or she has good nonverbal, fluid reasoning; can integrate nonverbal material; pays close attention to detail; and accurately responds to the visual-spatial material presented to him or her. Much of this involves using the kind of visual-spatial and visual-motor skills to solve problems that are not taught in formal academic schooling.

Working Memory/Freedom from Distractibility (WAIS-III: Arithmetic, Digit Span, Letter-Numbering Sequencing; WISC-III: Arithmetic and Digit Span) is a more complex and controversial construct and has been extensively studied with children but much less so with adults. It has primarily been related to concentration, attention, and short-term memory and is believed to be lowered by poor number facility, anxiety, difficulty making mental shifts, and poor self-monitoring. Sequencing is also crucial because each of the relevant subtests requires that the respondent place numbers and symbols in their proper order. Wielkiewicz (1990) has suggested that the low concentration, memory, and sequencing reflected on this factor is often because of a poorly functioning executive ability. Specifically, the person experiences difficulty attending to stimuli and simultaneously performing other mental tasks (e.g., listening to spoken digits and storing them while simultaneously reversing them and then repeating them backward). Good performance also requires a high level of motivation. As a result of these diverse functions, a low Working Memory/Freedom from Distractibility factor is also likely to lower performances in other areas, and this should be considered when estimating the person's overall potential.

It is crucial to consider a variety of interpretive possibilities to interpret the Working Memory/Freedom from Distractibility Index. Often behavioral observations can be crucial. A client who frequently asks to have the questions repeated might have a high level of distractibility. Alternatively, a high degree of motor activity or excessive talking might highlight a client's high level of anxiety. If number skills have not been developed, the client might ask to write out the numbers related to the arithmetic problems or count out the numbers with his or her fingers. The importance of cautiously interpreting (and providing additional support) for Freedom from Distractibility is highlighted because Reinecke et al. (1999) were unable to find an association between Freedom from Distractibility and children diagnosed with ADHD.

The Processing Speed Index (PSI; WAIS-III/WISC-III: Symbol Search and Coding) reflects the mental and motor speed with which a person can solve nonverbal problems. Further subtest support for this index can be found if the person also has correspondingly high (or low) performances on the timed nonverbal tests of Object Assembly and Block Design. In addition to mental and motor speed, the Processing Speed factor is also a measure of a person's ability to plan, organize, and develop relevant strategies. Low scores on Processing Speed can also reflect poor motor control. Because speed and concentration require good test-taking attitudes, Processing Speed (as well as Freedom from Distractibility) can also be lowered by poor motivation to perform well. For this reason, these two factors are sometimes referred to as *validity* factors. Whether a lowered performance is the result of poor motivation can often best be assessed by behavioral observations in combination with clarification and consideration of the presenting problem. An overly reflective problem-solving style could also lower the Processing Speed factor because the person would take too much time cautiously considering his or her response to each item.

Both the WAIS-III and WISC-III have made calculating index scores easy; that is, conversion into standard scores ($M = 100$, $SD = 15$) has been incorporated into the normal scoring procedure. However, examiners still need to determine the significant (interpretable) differences between the combinations of index scores. This can be done for the WAIS-III by consulting Table B.1 in the *WAIS-III Administration and Scoring Manual* (Wechsler, 1997a, p. 205) and summarizing the results on the "Discrepancy Comparisons" table on the Discrepancy Analysis Page. The significance of the WISC-III Index score discrepancies can be determined by consulting Table B.1 in the *WISC-III Manual* (Wechsler, 1991, p. 261). There is no table to summarize the WISC-III index discrepancies on the WISC-III Record Form. Interpretations are then based on the comparisons between the different indexes. For example, if a person's WAIS-III Perceptual Organization Index was 15 points higher than his or her Processing Speed Index (POI > PSI), this difference would clearly exceed the .05 level and, according to Table B.2, a POI-PSI difference of 15 points occurred in only 32.4% of the standardization sample. It should be noted that this figure combines both POI > PSI and PSI > POI. Thus, it would be more accurate to half the percentage ($32.4 \div 2 = 16.2\%$) because there is interest only in knowing the frequency which POI > PSI occurs (and not PSI > POI). This results in a rounded-off frequency of 16%.

Index scores can be used for interpreting a person's relative strengths and weaknesses. However, the actual factor-based standard scores should not be presented in the psychological report because readers may confuse them with IQ scores. In addition, including both IQ scores and factor-based standard scores would make the report unnecessarily cluttered with too many numbers. After interpretations have been made, practitioners can then work to develop appropriate instructional recommendations if an educational or rehabilitation plan needs to be developed.

Step IIc. Additional Groupings: Bannatyne's Categories, ACID/SCAD Profiles, Horn Groupings

Four additional factors or groupings can often yield useful interpretations. These are optional and should be calculated when, on initially appraising the pattern of subtest scatter, it seems they might be relevant to investigate more formally. For example, if subtests that are highly loaded on spatial abilities (Picture Completion, Block Design, Matrix Reasoning, Object Assembly) appear significantly higher than sequencing subtests (Digit Span, Arithmetic, Digit Symbol), a formal calculation of Bannatyne's categories will serve to confirm or disconfirm initial impressions related to the subtest profiles. Another reason to calculate the groupings listed in Level IIc occurs when an examiner wishes to see if a person's subtest profile is similar or dissimilar to a person from an actual or suspected client population (e.g., learning disabled, Alzheimer's disease).

Bannatyne's Categories Bannatyne's categories comprise subtest patterns in which *Spatial* abilities are relatively higher than *Verbal Conceptualization,* which is in turn higher than *Sequential* abilities, with *Acquired Knowledge* typically being the lowest (Verbal Conceptualization > Spatial > Sequential > Acquired Knowledge).

These categories were originally developed as a means of detecting and understanding learning-disabled populations (Bannatyne, 1974). However, it has been found that many learning-disabled persons do not necessarily have this pattern, and many non-learning-disabled populations do have the pattern (see Groth-Marnat, 2002 and subsection "Learning Disabilities"). The result has been that Bannatyne's categories have been used to further understand Wechsler scale profiles in general and not for the diagnosis of specific conditions.

The recommended method of interpreting the WAIS-III Bannatyne's factors is to first use the following formulas to transform the subtest groupings comprising each of the factors into the familiar standard scores (Mean of 100 and Standard Deviation of 15). Note that scaled scores must be used.

WAIS-III Verbal Conceptualization: 1.8 (Vocabulary + Similarities + Comprehension) + 46

WAIS-III Spatial: 1.5 (Matrix Reasoning + Block Design + Object Assembly + Picture Completion) + 40

WAIS-III Sequential: 1.6 (Arithmetic + Digit Span + Digit Symbol + Letter-Number Sequencing) + 36

WAIS-III Acquired Knowledge: 1.9 (Information + Arithmetic + Vocabulary) + 43

An appraisal of the standard scores gives a general idea as to whether the classic Bannatyne pattern is present. However, to more precisely determine whether or not the differences are significant (and therefore interpretable), the following additional procedures must be taken. First, find the means of the standard scores and then subtract the mean from each of the standard scores. This indicates the various differences between the mean and the standard scores. To be significant at the .05 level, the following difference scores must be equal to or greater than:

Verbal Conceptualization	7.1
Spatial	7.9
Sequential	7.5
Acquired Knowledge	6.6

Note that the inclusion of the WAIS-III Matrix Reasoning and Letter-Number Sequencing subtests is somewhat speculative at this time but their inclusion is conceptually consistent because these subtests are measuring spatial and sequencing abilities, respectively. Picture Arrangement has been excluded from the categories because it is a maverick subtest that doesn't clearly load on any of the four categories.

The procedure for interpreting Bannatyne's factors on the WISC-III is similar. First, the subtest scores comprising each of the groupings should be transformed into standard scores using the following formulas:

WISC-III Verbal Conceptualization: 1.9 (Vocabulary + Comprehension + Similarities) + 43

WISC-III Spatial: 2.0 (Picture Completion + Block Design + Object Assembly) + 40

WISC-III Sequential: 2.3 (Digit Span + Arithmetic + Coding) + 31

WISC-III Acquired Knowledge: 1.9 (Information + Vocabulary + Arithmetic) + 43

As with the WAIS-III, the mean of the standard scores must then be calculated and the differences between the means must then be determined. To be significant at the .05 level, the following differences must be equal to or greater than the following values:

Spatial	10.0
Verbal Conceptualization	8.5
Sequential	12.0
Acquired Knowledge	8.5

Several interpretive warnings are advisable. Only about 20% to 25% of learning-disabled persons will actually demonstrate the classic sequence of Bannatyne's Verbal Conceptualization > Spatial > Sequential > Acquired Knowledge pattern (see Groth-Marnat, 2002 and A. Kaufman & Lichtenberger, 2002). One noteworthy variation from the Bannatyne profile is that sometimes a bright, highly motivated learning-disabled person with poor sequencing abilities compensates by developing a high level of acquired knowledge. Thus, the Acquired Knowledge category might be outstandingly high even though Sequential abilities might still be quite low. Another less bright and/or less motivated learning-disabled person might experience the disruption of poor sequencing to a greater extent and may have then become correspondingly alienated from academic learning. This would then be reflected in an outstandingly low Acquired Knowledge category. This is consistent with the finding that learning disabilities are a heterogeneous group of disorders with sometimes well-defined subtypes (A. Kaufman & Kaufman, 2002). This means that examiners need to take a flexible approach toward interpreting the relation and implications between the Bannatyne categories.

The ACID/SCALD/SCAD Profiles The ACID, SCALD, and SCAD profiles are similar to those of Bannatyne's categories. Low scores on each of these profiles have been found to occur more frequently among learning-disabled populations (Cordoni, O'Donnell, Ramaniah, Kurtz, & Rosenshein, 1981; A. Kaufman, 1994; A. Kaufman & Lichtenberger, 2002; Wechlser, 1997b). The WAIS-III/WISC-III *ACID* profile comprises **A**rithmetic, **C**oding (Digit Symbol-Coding for the WAIS-III), **I**nformation, and **D**igit Span. Note that three of the subtests (Arithmetic, Coding/Digit Symbol, Digit Span) comprise Bannatyne's Sequential Category and one (Information) is included in Bannatyne's Acquired Knowledge category. As with the Bannatyne categories, an exception to the pattern is that often learning-disabled university students who have academically compensated for their learning disabilities have relatively good performances on Information (Ackerman et al., 1987). A new *SCALD* profile has been proposed for the WAIS-III (A. Kaufman & Lichtenberger, 1999, 2002) and comprises **S**ymbol Search, **C**oding (Digit Symbol-Coding), **A**rithmetic, **L**etter-Number Sequencing, and **D**igit Symbol. This is based on the finding that the Working Memory and Processing Speed indexes were generally lowest among adult learning disabled persons (Psychological Corporation, 1997) and are composed of these subtests.

For the WISC-III, A. Kaufman (1994) recommends deleting Information and instead inserting Symbol Search. The profile then comprises **S**ymbol Search, **C**oding, **A**rithmetic, and **D**igit Span and can then be appropriately renamed the *SCAD* profile. These four subtests are a merging of the WISC-III's Freedom from Distractibility (Arithmetic and Digit Span) and Perceptual Speed (Symbol Search and Coding) factors. To convert the SCAD profile into a standard score, the following formula can be used (A. Kaufman, 1994):

SCAD (WISC-III): 1.7 (SS + C + A DSp) + 32

To vary from the WISC-III Full Scale IQ (standard score for all subtests) at the .05 level, the standard SCAD score must be 9.5 points above/below the Full Scale IQ. However, if there is a significant difference (16 or more points) between the Processing Speed (Symbol Search and Coding) and Freedom from Distractibility (Arithmetic and Digit Span) indexes, the SCAD profile should not be interpreted.

Horn Groupings Horn and Cattell's (1966) fluid versus crystallized intelligence has been used to organize many of the Wechsler intelligence scales (J. Caruso & Cliff, 1999; A. Kaufman, 1994; A. Kaufman & Lichtenberger, 2002; Woodcock, 1990). On the WAIS-III, *Fluid Intelligence (Gf)* primarily includes Digit Span and Matrix Reasoning whereas *Crystallized Intelligence (Gc)* is measured by Vocabulary and Information (J. Caruso & Cliff, 1999). Kaufman and Lichtenberger (1999, 2002) have conceptually reorganized the WAIS-III subtests around the more detailed Horn (1985) groupings of *broad visualization* (*Gv*; Matrix Reasoning, Block Design, Object Assembly, Picture Completion), *broad speediness* (*Gs;* Digit Symbol-Coding, Symbol Search, Object Assembly) and *short-term memory* (Letter-Number Sequencing, Arithmetic, Digit Span). A rough idea of the examinee's relative strengths and weaknesses can be obtained by adding the subtest scaled scores, finding the mean, and comparing the strength and weaknesses. A more precise method is to convert the groupings into standard scores using the following formulas (A. Kaufman & Lichtenberger, 2002):

WAIS-III Fluid Intelligence: 1.1 (MR + BD + OA + S + PA + A) + 34

WAIS-III Crystallized Intelligence: 1.2 (I + V + C + S + PA) + 40

WAIS-III Broad Visualization: 1.5 (MR + BD + OA + PC) 40

WAIS-III Broad Speediness: 1.9 (CD + SS + OA) + 43)

WAIS-III Short-Term Memory: (Same as Working Memory Index)

After these standard scores have been determined, examinees next should find the mean of the five standard scores and calculate the differences that each of the groupings varies from the mean. To be significant at the .05 level (and thus interpretable), the following values should be either equal to or greater than the following:

Fluid Intelligence	7.5
Crystallized Intelligence	7.0
Broad Speediness	10.5
Short-Term Memory	8.3

The Horn groupings on the WISC-III are somewhat different from the WAIS-III in that Fluid Intelligence includes Picture Arrangement, Block Design, Object Assembly, Similarities, and Arithmetic. Crystallized Intelligence includes Information, Similarities, Vocabulary, Comprehension, and Picture Arrangement. Because Picture Arrangement and Similarities include skills related to both crystallized and fluid intelligence, they are included in both groupings. An additional grouping is Achievement, which is a composite of all tests most influenced by academic learning. This grouping includes Information, Similarities, Arithmetic, Vocabulary, Comprehension, and Picture Arrangement. To convert the WISC-III Horn groupings into standard scores ($M = 100$, $SD = 15$), the following formulas can be used (A. Kaufman, 1994):

WISC-III Fluid Intelligence: 1.3 (S + A + PA + BD + OA) + 35

WISC-III Crystallized Intelligence: 1.3 (I + S + V + C + PA) + 35

WISC-III Achievement: 0.85 (I + S + A + V + C + PA) + 49

Comparisons should be made between each of the WISC-III three standard scores to determine whether they are significantly different. To do this, first calculate the mean for the total subtests used in the three Horn groupings by summing the three standard scores and dividing by three. Then calculate the difference that each of the three standard scores varies from the mean. To be significant at the .05 level, the following values must be achieved:

Fluid Intelligence: 8.5 points
Crystallized Intelligence: 9 points
Achievement: 8.5 points

Level III. Interpreting Subtest Variability

The third step is to consider the degree to which the individual subtests deviate from the full scale, verbal, or performance subtest means and to determine the meaning associated with the subtest fluctuations. The outcome should be a description of a person's relative cognitive strengths and weaknesses. A listing and discussion of the meaning of each subtest and the abilities it measures is provided in the next major section of this chapter ("Wechsler Subtests"). Clinicians can refer to this section, as well as to information on how to assess special populations in developing their own hypotheses about important dimensions of intersubtest scatter. Readers may also wish to refer to R. Gregory (1999), A. Kaufman (1990, 1994), A. Kaufman and Lichtenberger (1999, 2000, 2002), Naglieri (1993), and Sattler (2001), who have provided detailed lists of hypotheses and useful tables for various combinations of high and low subtest scores. However, Level III interpretation is necessary only if there is sufficient subtest scatter. If all the subtests are fairly even, it is not necessary to attempt subtest profile interpretation.

Clinicians need to be aware that interpreting subtest variability involves clinical judgment guided by theory, observation, and an integration of the specifics of each case. Because there is little research base to support this process, it should be approached with caution. As a rule, the more subtests that can be combined to make inferences based on their shared abilities, the more support can be found for such an inference. At one extreme would be a series of subtests that combine to make up one of the previously discussed indexes/factors. The opposite would be only one subtest used to make an inference. In general, inferences based on only a single subtest should be treated with the most caution. While these single subtest-based inferences can be viable, it is incumbent on the clinician to obtain as much supporting evidence as possible.

It should also be noted that subtest interpretation has been the source of controversy in that some authors have pointed out that the subtests are not sufficiently reliable, do not have enough subtest specificity, and do not provide sufficient incremental validity beyond what might be accounted for by the Full Scale IQ (Konold et al., 1999; McDermott, Fantuzzo, Glutting, Watkins, & Baggaley, 1992). In part, this relates to empirical concerns, but there are also underlying conceptual differences centered around whether intelligence is mainly accounted for by *g* ("lumpers") as opposed to its being composed of a number of different components ("splitters"). This debate seems to have been present almost as long as conceptions of intelligence have been in existence. One common response to this issue is that subtest interpretation is not merely an empirical activity but also involves a clinical process of hypothesis testing and integrating a variety of sources of data (see A. Kaufman, 1994; A. Kaufman & Lichtenberger, 1999, 2000, 2002; Lezak, 1995; E. Kaplan et al., 1999). Accordingly, the following three steps are recommended in interpreting subtest variability: (a) Determine whether subtest fluctuations are significant, (b) develop hypotheses related to the meaning of any subtest fluctuations, and (c) integrate these hypotheses with additional relevant information regarding the examinee. Clinicians should very clearly *not* attempt to interpret subtests by merely listing the abilities provided in the subtest descriptions. It is hoped that the following guidelines will help ensure that clinicians develop accurate, useful, and well-integrated interpretations of subtests.

Step IIIa. Determine Whether Subtest Fluctuations Are Significant

The first step in profile analysis is to account for the implications of the mostly modest reliabilities associated with the different subtests. This means each clinician needs to seriously consider whether the variability results from reliably measured strengths or weaknesses or is merely from chance error inherent in the subtest. The WAIS-III Record Form conveniently allows examinees to determine whether subtests are significantly different from the mean of all subtests included under the listing for Verbal subtests and from the mean of all subtests included under the listing for Performance subtests (see WAIS-III Score Conversion Page). This may be somewhat confusing because there are some subtests included (Symbol Search, Letter-Number Sequencing, Object Assembly) in calculating the "Verbal" and "Performance" means that are not actually used to calculate the Verbal and Performance IQs. However, they are still listed as verbal or performance subtests. Whether the discrepancies are significant can be determined by calculating the magnitude that each WAIS-III subtest varies from the "Verbal" and "Performance" mean and noting whether it is significant based on Table B.3 in the *WAIS-III Administration and Scoring Manual* (pp. 208–209). The various subtests can then be indicated as either strengths or weaknesses ("Ss" or "Ws"). The WAIS-III also provides procedures for establishing the frequency with which the subtests fluctuate in the standardization population.

One area of potential confusion is that the WAIS-III manual does not provide tables to develop scatter analysis when six Performance Scale subtests are administered. Unfortunately, using six subtests is necessary if examiners would like to obtain both the Performance IQ and Processing Speed index score. There are also no tables available if certain optional subtests are used instead of the standard subtests. LoBello, Thompson, and Venugopala (1998) have provided supplementary tables to help examiners work with these situations.

The WISC-III does not provide a worksheet for determining subtest fluctuations. However, the magnitude (and whether the subtests are strengths or weaknesses) can be determined by following the guidelines detailed in Appendix E (Worksheet for Determining Magnitude of WISC-III Subtest Fluctuations). In most instances, the Worksheet for Determining Magnitude of WISC-III Subtest Fluctuations is sufficient for calculating whether subtests are relative strengths or weaknesses. If an unusual number of subtests have been administered, it might be necessary to consult Table B.3 in the *WISC-III Manual* (p. 263) to determine means and magnitudes of discrepancies. These guidelines will help clinicians determine whether the subtests actually fluctuate from the Verbal, Performance, or Full Scale means to a significant (at the .05 level) extent. It should be noted that a moderately high range (highest minus lowest subtest score) is a common occurrence. The average range on the WISC-III standardization sample was 8.5 points (SD = 2.3) for the Full Scale (13 subtests), 5.5 (SD = 2.1) points for the Verbal Scale (6 subtests), and 7.1 points (SD = 2.4) for the Performance Scale (7 subtests; Wechsler, 1991). Approximately two thirds of the WISC-III standardization sample had subscales that ranged between 7 and 8 points. Thus, clinicians should be cautious about not inferring pathology when such differences might merely indicate preferences for differing cognitive styles.

Often, it is found that there are no significant subtest fluctuations. In these cases, do not proceed to Steps IIIb and IIIc. Instead, focus on interpreting the profile based on information derived from Levels I and II and possibly Levels IV and V if these are relevant.

Step IIIb. Develop Hypotheses Related to the Meaning of Subtest Fluctuations

Just because a subtest or group of subtests has been designated as a relative strength or weakness does not mean that it is clear which of the various functions involved with the subtest is a strength or weakness. For example, Picture Arrangement involves planning (sequencing), visual organization, distinguishing essential from nonessential detail, and comprehending social situations. For one person, scoring high in Picture Arrangement might reflect excellent planning/sequencing abilities, for another it might reflect good social skills, and a third person might be high in both. It is the examiner's responsibility to become actively engaged with the pattern of subtests and any other relevant sources of information to determine which ability or abilities are high and low for the person. Interpreters who merely list the subtest's abilities as they are listed in a book are quite likely to make incorrect and even potentially damaging conclusions about the examinee. This cookbook type of approach should be strongly discouraged.

The underlying principle in uncovering the actual subtest strengths or weaknesses is to initially consider a significantly high or low subtest score in the context of scores on other subtests. If a person has scored high on Picture Arrangement and this might reflect good planning/sequencing, an examiner would expect other subtests that also measure planning/sequencing to be at least within the average range, if not higher. Thus, the examiner might make sure that other sequencing-oriented tasks, primarily Digit Span, Arithmetic, and Letter-Number Sequencing, were also high.

The difficulty with such a procedure is that it requires an in-depth knowledge of each subtest's abilities, familiarity with frequent clusters of subtests, and an overreliance on intuition in terms of noticing and testing different patterns. This is a particularly daunting task for beginning and even experienced clinicians. Thus, a formal step-by-step procedure of comparing and contrasting relative strengths and weaknesses is recommended. This can be accomplished by completing Appendix F ("Guidelines for Hypothesizing Subtest Strengths and Weaknesses" p. 682). These guidelines use the same underlying principle in that consistencies and inconsistencies among patterns of subtests are determined. However, these patterns are investigated in a thorough and systematic pattern. The directions are adapted from A. Kaufman (1990, 1994) and A. Kaufman and Lichtenberger (1999, 2000, 2002) and the listed subtest abilities were adapted from those described by a wide variety of sources, including Bannatyne (1974), Horn (1985), A. Kaufman (1990, 1994), A. Kaufman and Lichtenberger (1999, 2000, 2002), Lezak (1995), and Sattler (2001). After completing Appendix F, the clinician will have arrived at a series of empirically derived and partially tested hypotheses.

An important consideration in this strategy of subtest interpretation is that it should not be a rigid, mechanical process. For example, a client who presents with subjective complaints related to poor sequencing (e.g., difficulty following directions, placing things in the wrong order) may not necessarily have all the expected WAIS-III/WISC-III subtests quite within the statistically interpretable range. However, given the quite clear symptom reports (and possibly behavioral observations), practitioners may still choose to interpret the sequencing-related subtests. In contrast, another client might have most sequencing subtests in the statistically significant range but poor sequencing was neither a symptom complaint, nor were behavioral observations noted that would have been consistent with poor sequencing. As a result, the hypothesis of poor sequencing might be rejected as not applying to the person. The outlined procedure, then, should be used for hypothesis generation in which other factors beyond the mechanical interpretation procedure can confirm or disconfirm these hypotheses.

Step IIIc. Integrate Subtest Hypotheses with Additional Information

Before finally accepting or rejecting the step-by-step empirically derived hypotheses from Steps IIIa and IIIb, examiners should consider additional sources of relevant information. This might include behavioral observations, medical records, school records, teacher's reports, other test data, or qualitative responses that examinees have made to the test items (see Level V). For example, an examiner might be trying to decide whether low scores on Arithmetic and Digit Span reflect poor attention or poor sequencing. If the examinee was observed to have attended well to the tasks but had difficulty following a series of directions, then it suggests sequencing is more likely to be the difficulty. Or, an examiner might be trying to decide whether the examinee prefers a simultaneous or sequential style of processing information. A relevant behavioral observation is careful observation of the way the person worked on Block Design. Did he or she proceed in a step-by-step sequence, trying to match each block with a segment of the picture, or, rather, did he or she try to understand the design as a whole while attempting to complete the task? A final relevant example might be low scores on Arithmetic, Digit Span, Digit Symbol-Coding, and Symbol Search. Each of these subtests requires a high level of motivation. Indeed, they have sometimes been referred to as validity scales because they are likely to be lowered as a result of poor motivation (A. Kaufman, 1994). Rather than work to decipher the examinee's low abilities as reflected in these subtests, the examiner might decide that behavioral observations more accurately suggest the person was not expending a sufficient amount of effort.

A focus on additional sources of information, particularly behavioral observations, also has relevance for determining the significance of subtest fluctuations (Step IIIa) and developing hypotheses (Step IIIb). As was stressed previously, sometimes a subtest fluctuation may not quite achieve formal statistical significance, yet, because of additional information, the practitioner feels justified in giving the score greater clinical importance and considering it for interpretation. Similarly, generating hypotheses by formally putting the data through Step IIIb may not have confirmed a suspected hypothesis. However, if a clinician has additional information that might justify accepting the suspected hypothesis, he or she may be persuaded to accept it although some of the formal procedures have not quite supported it. This highlights an essential underlying philosophy of Wechsler scale and subtest interpretation: It is not solely a statistical and empirical exercise, but, more importantly, it involves the use of clinical skills and judgment.

Level IV. Intrasubtest Variability

A further, potentially important area of analysis involves looking at the patterns of performance within the items of each subtest. These items are arranged in sequences that become progressively more difficult. Thus, a normal and expected pattern would have the examinee pass the initial items and slowly but evenly begin to fail more difficult ones. A more sporadic pattern, in which the examinee misses initial easier items but passes later more difficult ones, may suggest an attentional deficit or specific memory losses, particularly related to retrieval difficulties (E. Kaplan, Fein, Morris, & Delis, 1991; E. Kaplan et al., 1999). If performance is highly sporadic, the reason should be explored further. For example, clients might be consciously faking if they miss every other item, miss extremely easy items, and/or appear much more alert than their obtained IQ. Sporadic performance might also be characteristic of patients with brain damage with diffuse cortical (Mittenberg, Hammeke, & Rao, 1989) or subcortical involvement (Godber, Anderson, & Bell, 2000). An analysis of the intrasubtest scatter can thus provide a type of information different from that obtained by merely looking at the quantitative-scaled scores. It should be noted, however, that research on this is equivocal given that

J. Ryan, Paul, and Arb (1999) were unable to find high subtest scatter on the Information subtest among patients who had documented retrieval difficulties.

Level V. Qualitative Analysis

The final step is to look at the content of responses, especially on Information, Vocabulary, Comprehension, and Similarities. Frequently, the presence of unique, highly personal, or unusual responses can suggest some important dimensions of an individual's intellectual or personality functioning (see Groth-Marnat et al., 2000; E. Kaplan et al., 1991, 1999). For example, some responses may reflect aggressive tendencies, concrete thinking, or unusual associations. A highly aggressive person might provide unusual responses on some of the Vocabulary items, or a person with paranoid personality characteristics might provide rigid, cautious, and legalistic responses. Similarly, impulsivity might be suggested by persons who quickly place incorrect blocks together on Block Design and then do not reflect on whether their designs were correct.

WECHSLER SUBTESTS

To interpret the Wechsler scales adequately, it is essential to understand the various abilities that each subtest measures. This section presents the different abilities involved in each of the 14 WAIS-III and 13 WISC-III subtests, followed by a discussion of their relevant features, including the possible meanings associated with high or low scores. Descriptions of the subtest abilities and data on factor loadings presented for most of the WISC-III subtests are derived from A. Kaufman (1994) and A. Kaufman and Lichtenberger (2000, 2002). Subtest abilities and factor loadings for the WAIS-III are based on research reviewed by A. Kaufman and Lichtenberger (1999, 2002) and Sattler (2001). Some citing of relevant and usually recent sources is also provided.

In keeping with the overall approach of this book, any interpretations suggested in the discussion of the subtests should be considered tentative. They are merely beginning possibilities that must be explored further and placed in a proper context. In addition, no subtest is a pure measurement of any single intellectual ability; rather, each represents a combination of skills. It is important to emphasize that a low or high score on a specific subtest can occur for a variety of reasons, which the examiner must consider in interpreting the overall profile. This section is most helpful only after practitioners are familiar with the subtest stimuli and administration procedure outlined in the WAIS-III and WISC-III manuals.

Verbal Scales

The Wechsler Verbal Scales assess an individual's proficiency in the following areas:

- The ability to work with abstract symbols.
- The amount and degree of benefit a person has received from his or her educational background.
- Verbal memory abilities.
- Verbal fluency.

The WAIS-III and WISC-III Verbal Scales are generally more subject to cultural influences, whereas the Performance Scales are considered to be somewhat more culture free. If an individual does significantly better (9 points or more for the WAIS-III or 12 points or more for the WISC-III) on the Verbal Scales compared with the Performance subtests, this difference may indicate a number of interpretative possibilities, including a relatively high level of education; a tendency toward overachieving; psychomotor slowing because of depression; difficulty working with practical tasks; deficits in performance abilities; poor visual-motor integration; a slow, deliberate, reflective work style that results in relatively lower scores on timed tests (but higher scores on verbal tests); or a quick, impulsive work style resulting in relatively more errors on Performance subtests (A. Kaufman, 1994; A. Kaufman & Lichtenberger, 1999, 2000, 2002; Sattler, 2001). In addition, persons from professional occupations, high educational attainment, and high IQs in general are likely to have quite high Verbal IQs. Also, psychiatric populations (5–6 point V > P discrepancy), persons with Alzheimer's disease, and persons with motor coordination problems tend to have higher verbal scores relative to their performance scores.

Studies with the WAIS-R have typically found that persons with unilateral right hemisphere lesions have, on average, a 9-point higher Verbal than Performance IQ (A. Kaufman, 1994; A. Kaufman & Lichtenberger, 1999, 2001, 2002; Reitan & Wolfson, 1993; Sattler, 2001). It is likely that future research will also find similar patterns with the WAIS-III. However, a V > P (e.g., depressed Performance IQ) should never be *diagnostic of* unilateral right hemisphere brain damage but rather *consistent with* this condition in some cases. It should be stressed that there is a complex interaction with a wide number of variables. A V > P effect is likely to be most pronounced among adult, educated (12+ years), Caucasian males with acute lesions who have strokes, tumors, or other focal lesions toward the posterior (versus anterior/frontal) regions. These variables have been extensively reviewed by Kaufman (1994) and Kaufman and Lichtenberger (2002) and are summarized in the following list:

- *Age.* Whereas the V > P effect has been clearly and consistently found for most adult populations, studies with children have been met with numerous contradictions. This is because there are a greater number of intervening variables for children, and their brains are more symmetrical and characterized by greater plasticity. Thus, neurological inferences related to Verbal-Performance discrepancies should *not* be made for children.
- *Education.* Because persons with higher education (and generally persons with higher IQs) typically score higher on Verbal subtests, a further lowering in performance abilities because of a right hemisphere lesion will serve to exaggerate the V > P discrepancy to an even greater extent. Persons from lower educational backgrounds often have higher Performance IQs relative to their Verbal IQs so that a lowering in their Performance IQ because of a right hemisphere lesion may either not produce the expected V > P effect, or the difference may not be as wide as for persons with higher educational attainment.
- *Race.* European American and African Americans are more likely to have the V > P discrepancy following right hemisphere damage than either Hispanics or

Native Americans. This is because Hispanics and Native Americans are more likely to have higher Performance than Verbal IQs before their injury or illness.

- *Gender.* The WAIS-R V > P discrepancy following right hemisphere lesions is more pronounced in males (13 points) than in females (7 points; A. Kaufman & Lichtenberger, 2002). This results partially from greater cerebral asymmetry in males. It is also possibly because of more verbally mediated strategies for Performance subtests by females, which serves to partially compensate for organically lowered performance abilities.
- *Recency of Lesion.* Acute (less than 12 months) unilateral right hemisphere lesions produce greater V > P effects than chronic lesions. This happens because, over time, patients are able to improve their performance abilities through both natural recovery of function and compensatory techniques. Even with chronic lesions, there is still an expected V > P discrepancy, but it is not as extreme as for acute lesions.
- *Type and Location of Lesion.* Especially right hemisphere strokes, but also tumors and, to a lesser extent, right temporal lobe epilepsy result in the expected V > P effect. Frontal lobe lesions have little effect on V – P differences, whereas posterior lesions do result in the expected V > P discrepancy.

Vocabulary

The Vocabulary subtest includes the following abilities or traits:

- Language development.*
- Word knowledge.*
- General verbal intelligence.
- Language usage and accumulated verbal learning ability.
- Rough measure of the subject's optimal intellectual efficiency.
- Educational background.
- Range of ideas, experiences, or interests that a subject has acquired.

The Vocabulary subtest is a test of accumulated verbal learning and represents an individual's ability to express a wide range of ideas with ease and flexibility. It may also involve the person's richness of ideas, long-term memory, concept formation, and language development. Vocabulary is noteworthy in that it is the most reliable Verbal subtest (WAIS-III test-retest reliability = .91; WISC-III test-retest reliability = .89) and, like Information, it is highly resistant to neurological deficit and psychological disturbance (Lezak, 1995; Reitan & Wolfson, 1993). Although the Vocabulary subtest holds up with age, it tends to fall off with those people for whom visual-spatial skills are far more important than verbal abilities. Vocabulary generally reflects the nature and level of sophistication of the person's schooling and cultural learning. Vocabulary is primarily dependent on the wealth of early educational environment, but it is susceptible to

* Abilities followed by an asterisk indicate specific abilities and traits strongly associated with the subtest under discussion.

improvement by later experience or schooling. It is the least variable of all the subtests, and subtest scores below the Vocabulary level sometimes imply a drop of efficiency in that function. Vocabulary is the best single indicator of general intelligence, with 69% of its variance accounted for by *g* on the WAIS-III and 80% of its variance accounted for by *g* on the WISC-III. Because of its high degree of stability, Vocabulary is often used as an indicator of a person's intellectual potential and to make an estimate of their premorbid level of functioning (see more precise methods in "Assessing Brain Damage" section).

The Vocabulary responses are similar to Comprehension and Similarities in that a qualitative analysis often provides useful information relating to the examinee's thought processes, background, life experiences, and response to frustration. It is often important to explore incorrect responses to determine whether they were guesses, clang associations (e.g., "ponder" meaning "to pound" or "assemble" meaning to "resemble"), concrete thinking, bizarre associations, or overinclusive reasoning. Even when a response is correct, a consideration of the style used to approach the word and specific content can be helpful.

High scores suggest high general intelligence and indicate that the examinee can adequately recall past ideas and form concepts relating to these ideas. Persons with high scores have a wide range of interests, a good fund of general information, and may have high needs for achievement. Clinical populations who score high on Vocabulary may use compulsive or intellectualizing defense mechanisms. Low scores suggest a limited educational background, low general intelligence, poor language development, lack of familiarity with English, and/or poor motivation.

Similarities

- Logical abstract reasoning.*
- Verbal concept formation or conceptual thinking.
- Distinguishing essential from nonessential details.
- Associative ability combined with language facility.

The Similarities subtest requires verbal concept formation and abstract reasoning ability. These functions mediate for the individual an awareness of the belonging-togetherness of objects and events of the day-to-day world. An essential aspect of adjusting to one's environment is the use of these abilities to clarify, reduce, and classify the style and manner to which a response is made. Inductive reasoning is required as the examinee must move from particular facts to a general rule or principle. Implicit in the test is the ability of individuals to use long-term memory and to apply elegant expressions in their responses. The more precise and abstract the expression, the higher the score, which indicates that verbal fluency is an important determinant. Correct responses to the last few items indicate a particularly high level of abstraction. Individuals with a good ability for insight and introspection tend to perform highly on this subtest; thus, it may be used as an indicator of favorable prognosis for psychotherapy. Scores decrease significantly in schizophrenics, rigid or inflexible thinkers, and patients with senile conditions. Examiners can, therefore, use this subtest to gain further information regarding the nature of an examinee's idiosyncratic or pathological form of concept formation.

High scorers show good verbal concept formation, which, if unusually high, may reflect intellectualizing tendencies. Low scorers show poor abstraction abilities, literalness, and inflexible thinking. The Similarities subtest in adult protocols is the most sensitive subtest to left hemisphere lesions, particularly lesions to the left temporal and/or left frontal regions (Dobbins & Russell, 1990).

Arithmetic

- Computational skill.*
- Auditory memory.
- Sequencing ability.
- Numerical reasoning and speed of numerical manipulation.
- Concentration and attention/low distractibility.
- Reality contact and mental alertness; that is, active relationship to the outside world.
- School learning (earlier items)/acquired knowledge.
- Logical reasoning, abstraction, and analysis of numerical problems (later items).

The Arithmetic subtest requires a focused concentration as well as basic mathematical skills and an ability to apply these skills. The skills required to complete this test are usually acquired by the time a person reaches junior high school; therefore, low scores are more likely to be the result of poor concentration. Arithmetic is likely to be more challenging and stressful than tests such as Information and Vocabulary, both because the task itself is more demanding and because the test is timed. Thus, persons who are susceptible to the disruptive effects of anxiety are likely to be adversely affected. However, examiners may want to establish whether the person simply lacked the necessary skills or had difficulty concentrating. This can be assessed by giving the person previously missed items a second time but allowing the use of paper and pencil without a time limit. Under these circumstances, persons with adequate mathematical knowledge who are distractible should be able to complete the items correctly.

Individuals from higher socioeconomic backgrounds, obedient teacher-oriented students, and persons with intellectualizing tendencies usually do well on this subtest. A helpful formula is that Information plus Arithmetic equals school achievement. Because numbers come from the outside environment and create rule and direction, some individuals react rebelliously. This is particularly true for antisocial personalities. Histrionic personalities, who do not readily accept outside direction and generally refuse to take responsibility for their behaviors, may likewise do poorly. This is not to suggest that lowered Arithmetic scores are diagnostic of these clinical groups, but rather, that this lowering may at times be consistent with the way these individuals interact with their environment.

High scorers show alertness, capacity for concentration, freedom from distractibility, and good short-term auditory memory and may use intellectualizing defenses. Low scorers show poor mathematical reasoning, lack of capacity to concentrate, distractibility, and poor auditory short-term memory. A poor educational background in which adequate mathematical skills have not been developed can also account for lowered performance.

Digit Span

- Immediate rote recall.*
- Reversibility; ability to shift thought patterns (from digits forward to digits backward).*
- Concentration and attention.
- Auditory sequencing.
- Rote learning.

Digit Span is considered to be a test of short-term memory and attention. The subject must recall and repeat auditory information in the proper sequence. Bannatyne (1974) has further described this as "auditory vocal sequencing memory." Correct responses require a two-step process. First, the information must be accurately received, which requires attention and encoding. Persons who are easily distractible have difficulty in this phase. Second, the examinee must accurately recall, sequence, and vocalize the information. Persons who can perhaps receive the information correctly may still have difficulty at this phase if they have short-term memory difficulties because they cannot hold the memory trace long enough. Sometimes, the previous digit is forgotten as they are attempting to vocalize a present one. Whereas Digits Forward is a simpler, more straightforward task requiring rote memory, Digits Backward is more complex. The examinee must usually hold the memory longer and also transform it before making a restatement. Thus, a good performance on Digits Backward is likely to reflect a person who is flexible, can concentrate, and is tolerant of stress. High Digits Backward scores may also involve the ability to form, maintain, and scan visual mental images formed from the auditory stimulus (Lezak, 1995; Wielkiewicz, 1990).

Passive, anxiety-free individuals seem to do best on this test. It requires an effortless and relatively unhampered contact with reality, which is characterized by open receptivity to incoming information. Performance is greatly hampered by increased anxiety or tension, and the Digit Span subtest is considered the most susceptible to the effects of anxiety. In addition to Digit Span, the other subtests that are sensitive to the effects of anxiety are Arithmetic, Digit Symbol-Coding, and Letter-Number Sequencing (WAIS-III). Collectively, these three subtests form the WAIS-III Working Memory Index and are (along with the Processing Speed subtests) sensitive tests to brain damage, mental retardation, and learning disabilities (Lezak, 1995; Psychological Corporation, 1997). Similarly, the Digit Span subtest (and Arithmetic) is included in the WISC-III Freedom from Distractibility Index, which is also sensitive to the effects of learning disabilities, ADHD, brain damage, and mental retardation (Bannatyne, 1974; A. Kaufman, 1994).

Persons who score high have good auditory short-term memory and excellent attention and may be relatively unaffected by stress and anxiety. However, just because a person has good short-term auditory memory for digits does not necessarily mean that his or her memory for more complicated information, such as music or verbally relevant information, is also good. These more complex features of memory may have to be assessed by other means. When Digits Backward is longer than Digits Forward, this rare event (3% to 10% of children's protocols; Wechsler, 1991; .9% of adult profiles, Psychological Corporation, 1997) suggests that the individual has excellent numerical abilities. Low scores on Digit Span indicate difficulty concentrating, which may be the result of anxiety or unusual thought processes. A large discrepancy (5 digits) in favor of Digits Forward versus Digits Backward can suggest the presence of an organic deficit, particularly if the overall backward Digit Span score is below scores for tests such as Information and Vocabulary. Whereas Digits Forward is fairly stable and resistant to deterioration, Digits Backward is a far more difficult task and is quite sensitive to deterioration (see subsection on estimating premorbid IQ in the "Assessing Brain Damage" section). Whereas Digits Forward is more likely to be lowered by left hemisphere lesions, lowered Digits Backward is more consistent with either diffuse or right frontal involvement (Lezak, 1995; Swierchinsky, 1978). Lowered performance for both Digit Span backward and Digit Symbol occur with the diffuse damage associated with exposure to solvents (Groth-Marnat, 1993; Morrow, Furman, Ryan, & Hodgson, 1988).

Information

- Range of general factual knowledge.*
- Old learning or schooling.
- Intellectual curiosity or urge to collect knowledge.
- Alertness to day-to-day world.
- Long-term memory.

The Information subtest samples the type of knowledge that average persons with average opportunities should be able to acquire. This knowledge is usually based on habitual, overlearned material, particularly in the case of older children and adults. Both Information and Vocabulary are highly resistant to neurological deficit and psychological disturbance (Lezak, 1995; Reitan & Wolfson, 1993) and are two of the most stable subtests. Because of this stability, Wechsler referred to them as "hold" tests as opposed to "no-hold" tests, which he theorized are more sensitive to deterioration and such situational variables as anxiety and fatigue (i.e., Arithmetic, Digit Symbol-Coding, Block Design). Furthermore, both these subtests are good measures of general intelligence and are highly correlated with educational level (A. Kaufman et al., 1988) and WAIS-III and WISC-III Full Scale IQs. Research has shown that the earlier WAIS-R Information and Vocabulary subtests have predicted college grade point average as accurately as well-established college aptitude tests (Feingold, 1983). It is for these reasons that Information (along with Vocabulary and Arithmetic) is included in Bannatyne's Acquired Knowledge category. It also loads most strongly (.84) on the Verbal Comprehension factor.

Although performance on the Information subtest involves remote memory and alertness to the environment, it is influenced only to a small extent by conscious effort and is believed to be only minimally affected by factors such as anxiety. To score well, the individual must have been exposed to a highly varied past environment, have an intact long-term memory, and possess a wide range of interests.

A high score on this subtest suggests that the examinee has good long-term memory, cultural interests, strong educational background, positive attitude toward school, good verbal ability, and possibly intellectualization as his or her most frequently used defense mechanism. Low scorers may show superficiality of interests, lack of intellectual curiosity, cultural deprivation, or lack of familiarity with Western (primarily American) culture (however, note the availability of numerous foreign country adaptations). Failing initial easy items combined with success on more difficult ones (high intrasubtest variability; see Level IV procedure) may suggest difficulties with retrieval, although research substantiating this hypothesis has been equivocal (E. Kaplan et al., 1991; Mittenberg et al., 1989; J. Ryan & Paul, 1999). High intrasubtest scatter may also suggest the possibility of malingering or poor motivation.

Comprehension

- Demonstration of practical knowledge.*
- Social maturity.*
- Knowledge of conventional standards of behavior.*
- Ability to evaluate past experience; that is, proper selection, organization, and emphasis of facts and relationships.*
- Abstract thinking and generalization (later items only).*
- Social judgment, common sense, or judgment in practical social situations.
- Grasp of social milieu; for example, information and knowledge of moral codes, social rules, and regulations.
- Reality awareness, understanding, and alertness to the day-to-day world.

Comprehension has often been considered to reflect the extent to which an examinee adheres to conventional standards, has benefited from past cultural opportunities, and has a well-developed conscience. However, formal studies have generally not supported a relationship between Comprehension and various measures of social intelligence (see Beebe, Pfiffner, & McBurnett, 2000). Comprehension is also, at least in part, a test of information, which is supported by its high correlation (low- to mid-70s, depending on age) with the Information and Vocabulary subtests. Comprehension involves an adaptive response by the individual to a situation that requires him or her to select the most efficient way of dealing with a specific problem. The examinee not only must possess relevant information but also must appropriately use this information for decision making. In this sense, the Comprehension subtest goes one step beyond the degree of complexity and synthesis required for the Information subtest. Like Vocabulary and Information, it measures general verbal ability—66% of its WAIS-III variance and 42% of its WISC-III variance are attributed to the Verbal Comprehension factor. The examinee must not only have the necessary information, but also apply it in a coherent, problem-oriented manner. Thus, a Comprehension score significantly below the Information score suggests that an examinee is not effectively using his or her knowledge.

In assessing an examinee's responses, it can be important to distinguish between actually dealing with the material to develop an original response and merely repeating overlearned concepts. For example, parroting answers to "forest," "parole system," or the proverbs does not indicate full comprehension and may simply be based on past experience rather than on accurate problem solving, good judgment, or abstract reasoning. Thus, basic rule-of-thumb answers can significantly increase the total number of correct responses. However, in the later items, a correct response requires higher-level problem solving, and these items, therefore, can still be a good measure of general intelligence instead of merely rote memorization.

Personality variables, especially those relating to judgment, are important areas to consider in this subtest. In particular, poor levels of adjustment can lower scores on Comprehension. Clinicians should note the pattern of responses, clichés, literalness, and any circumscribed responses. In contrast, good judgment involves the ability to engage in discriminative activity. Failure on the easy items indicates impaired judgment, even though later, more difficult items are passed. It is important to note emotional implications on this subtest because emotional responsiveness influences the way a person evaluates environmental events. For example, individuals who are highly analytical and use these analytical abilities to avoid emotions may have difficulty understanding the social components of situations as presented in Comprehension.

High scorers show reality awareness, capacity for social compliance, good judgment, and emotionally relevant use of information. Low scorers, especially if they have four or more subscale points below Vocabulary, might have poor judgment, impulsiveness, and hostility against their environment. Mentally disturbed persons often do poorly on Comprehension, which may be the result of disturbed perceptions, idiosyncratic thinking, impulsiveness, or antisocial tendencies.

Letter-Number Sequencing (WAIS-III only)

- Auditory short-term memory.
- Sequencing ability.
- Concentration and attention.

A good performance on Letter-Number Sequencing suggests that the person has good sequencing, attention, and concentration. It requires him or her to attend to a series of letters and numbers that have been read to him or her, hold them in memory, manipulate them into a new order, and repeat the new sequence. When combined with Arithmetic and Digit Span, it forms the Working Memory Index, but it is not used to calculate any of the IQs. Letter-Number Sequencing (along with Digit Span) is a subtest, which is also included on the Wechlser Memory Scale-III.

Psychometrically, Letter-Number Sequencing is good to adequate. Test-retest reliability has been found to range between .70 to .80, the SEM is 1.30, and it has a factor loading of .62 with the Working Memory Index.

Performance Scales

The Performance scales reflect:

- The individual's degree and quality of nonverbal contact with the environment.
- The ability to integrate perceptual stimuli with relevant motor responses.
- The capacity to work in concrete situations.
- The ability to work quickly.
- The ability to evaluate visuo-spatial information.

The Performance subtests are generally less affected by educational background than are the Verbal scales. If an individual does significantly (.05 level) better (9 points or more on the WAIS-III, 12 or more points on the WISC-III) on the Performance scales than on the Verbal subtests (P > V), this may indicate a number of interpretive possibilities, including superior perceptual organizational abilities, ability to work under time pressure, a tendency toward low academic achievement, possible acting out (juvenile delinquency), an individual who could be described as a doer rather than a thinker, a person from a relatively low socioeconomic background, presence of a language deficit, poorly developed auditory conceptual/processing skills, or that immediate problem solving is better developed than problem solving based on accumulated knowledge.

A number of studies, primarily with the WAIS/WAIS-R have found that a higher Performance than Verbal IQ (P > V) is consistent with unilateral left hemisphere lesions (A. Kaufman & Lichtenberger, 2002). There is, however, a complex relation between a number of relevant variables and P > V for unilateral left lesion patients. One issue is that the average Verbal IQ superiority of 4 points for unilateral left lesion patients across studies is not nearly as pronounced as the 9-point average for V > P with right hemisphere lesions.

Because the P > V effect is not as strong as the V > P discrepancy found with unilateral right hemisphere lesions, interpretations need to be quite tentative. In general, P > V discrepancies are most likely to occur for adult male patients with low educational attainment who have lesions in the posterior (versus frontal) regions (see A. Kaufman & Lichtenberger, 2002). These variables can be summarized as follows:

- *Age.* The P > V difference for left lesion adults is relatively small but has been found not to occur for children. Therefore, inferences regarding lateralization should be restricted to adults and adolescents.
- *Gender.* The laterality effect for P > V following unilateral left hemisphere lesions has been found to be greater for males (6 points) than for females (only 1 point; A. Kaufman & Lichtenberger, 2002; see also previous section on gender for V > P following right hemisphere lesions).
- *Education.* Individuals having less than a high school education generally score 2 to 3 points higher on their Performance IQ than Verbal IQ. Clinically, this means that persons with low educational attainment are more likely to have even greater P > V following unilateral left hemisphere lesions than persons with more education.
- *Type and Location of Lesion.* Posterior left lesions are likely to show the expected P > V difference. Frontal lesions, no matter what the cause, are not likely to demonstrate any V > P differences. Left hemisphere strokes tend to produce the clearest P > V effect and, to a lesser extent, left temporal lobe epilepsy. Left hemisphere tumors, as well as the relative recency of the lesion (acute versus chronic), have little effect on V > P discrepancies.

A further consideration related to P > V difference is that research with the WAIS-R/WISC-III indicates that certain population groups are likely to score higher on Performance subtests. In particular, children, adolescents, and adult Native Americans and Hispanics (especially if bilingual) have Performance scores that can be an average of nearly 15 points above their Verbal scores. As a result, Wechsler Intelligence scale interpretation, especially if related to Verbal or Full Scale IQs, should be made with extreme caution (if at all). Instead, the Verbal and Performance IQs should be considered separately. Additional correlates of P > V are autism, mental retardation, learning disabilities, illiteracy, delinquency, conduct disorder or psychopathy, bilingual populations, and individuals from occupations (especially blue-collar) emphasizing visual-spatial skills. Possible explanations for these differences include the challenges involved in learning two languages, the level to which the test instructions have been understood, attitudes and experience working within time constraints, degree of cerebral lateralization, and cultural or subcultural differences (i.e., extent that nonverbal communication is emphasized). Each of these correlates should be taken into account when making interpretations related to lateralization of brain lesions or any of the other possible interpretations consistent with P > V discrepancies.

Picture Completion

- Visual alertness.*
- Visual recognition and identification (long-term visual memory).*
- Awareness of environmental detail; reality contact.
- Perception of the whole in relation to its parts; visual conceptual ability.
- Ability to differentiate essential details from nonessential details.
- Visual concentration combined with an ability to visually organize material.

The Picture Completion subtest is a measure of visual concentration and is a nonverbal test of general information. It involves discovering consistency and inconsistency by paying close attention to the environment and accessing remote memory. It is dependent on, and also draws on, an individual's experience with his or her culture. Thus, a person who is unfamiliar with common features of American/Western society may make errors because of a lack of experience rather than a lack of intelligence. A person will also make errors if he or she is unable to detach himself or herself emotionally from the material, thereby making accurate discriminations difficult. For example, passive, dependent personalities might make errors because they notice the absence of people controlling the actions in the pictures. Typical responses might be that "there's nobody holding the pitcher," "there are no people rowing the boat," or "there's no flagpole." Sometimes negative, inflexible, oppositional individuals state that there is nothing missing in the pictures.

High scorers are able to recognize essential visual information, are alert, and demonstrate good visual acuity. Low scores indicate poor concentration and inadequate visual organization. Impulsiveness can often produce lowered performance because the examinee may make a quick response without carefully analyzing the whole picture.

Digit Symbol-Coding/Coding

- Psychomotor speed.*
- Ability to follow directions.*
- Clerical speed and accuracy.*
- Visual short-term memory.*
- Ability to follow directions.*

- Paper-pencil skills.*
- Ability to learn an unfamiliar task; capacity for learning and responding to new visual material.
- Some degree of flexibility; ability to shift mental set.
- Capacity for sustained effort, attention, concentration, and mental efficiency.
- Associative learning and ability to imitate newly learned visual material.
- Sequencing ability.

Visual-motor integration is implied by good performance on Digit Symbol-Coding. However, the most important functions necessary for a high score are psychomotor speed combined with good recall for the symbol-digit pairs. This test involves appropriately combining the newly learned memory of the digit with the symbol, as well as adequate spatial-motor orientation, followed by executing the half-habituated activity of drawing the symbol. The subtest also requires the ability to learn an unfamiliar task, accuracy of eye-hand coordination, attentional skills, short-term memory, and the ability to work under pressure. This is a delicate and complex interaction, which can be disturbed because of difficulties with any of the preceding skills. In contrast to Vocabulary, which is a highly stable subtest, Digit Symbol is extremely sensitive to the effects of either organic or functional impairment. In particular, depressed patients and patients with brain damage have a difficult time with this subtest. It is also the subtest that is most influenced by age. For example, a raw score required to achieve a subscale score of 10 for the 70- to 74-year-old group would obtain a subscale score of only 6 when compared with the 20- to 34-year-old reference group.

Digit Symbol-Coding pairs with Symbol Search to form the Processing Speed Index. Digit Symbol-Coding is a fair measure of *g* for the WAIS-III (35% of its variance) but only a poor measure of *g* for the WISC-III (20% of its variance). It has ample subtest specificity for both the WAIS-III and WISC-III.

Because visual-motor coordination (particularly visual acuity and motor activity) is implied, it is not surprising to find that those individuals with high reading and writing experience are among the high scorers. Functions that are implicit in the task are rapid visual, spatial, and motor coordination, as well as the executive action of drawing the symbol. Because this task requires sustained attention and quick decision making, anxious hesitancy, obsessiveness, deliberation, and perfectionism significantly lower scores. This difficulty might be somewhat counteracted by informing persons who appear perfectionistic and reflective that they need only make their responses legibly but not perfectly. Persons who are extremely competitive but also become highly anxious in competitive situations may also be adversely affected. Not only can Digit Symbol-Coding scores be lowered by anxiety, but also the psychomotor slowing found in depressive states or the confused orientation of schizophrenics likewise produces a decrease in performance. Thus, a rough index of the severity of a person's depression can be assessed by comparing the relative lowering of Digit Symbol-Coding with other more stable subtests. Of particular significance is that Digit Symbol-Coding is one of the most sensitive subtests to the effects of any type of organic impairment (Lezak, 1995; Psychological Corporation, 1997; Reitan & Wolfson, 1993), and it tends to be one of the lower scores found in learning-disabled individuals (Bannatyne, 1974; Groth-Marnat, 2002; A. Kaufman, 1994). Even with minimal brain damage, Digit Symbol-Coding is still likely to be the lowest subtest overall (Lezak, 1995; Reitan & Wolfson, 1993). In addition, patients with rapidly growing tumors are more likely to have lower scores than those with slow-growing tumors (Reitan & Wolfson, 1993).

Because Digit Symbol-Coding requires such a diverse range of abilities, high or low scores can potentially indicate a wide number of possibilities. This means that clinicians need to work particularly hard to extract the significance of scores by integrating scores with other relevant measures, behavioral observations, and medical/personal history. The WAIS-III has included two optional procedures to help parcel out whether an examinee's score was attributable primarily to visual memory, graphomotor speed, or a combination of both. The first procedure, Incidental Learning, assesses how intact visual memory is by first requesting patients to recall as many of the digit-symbol pairs as possible and, second, to simply recall as many symbols as possible (without the associated numbers). These two related tasks are untimed. In contrast, Digit Symbol-Copy assesses graphomotor speed by presenting the examinee with a series of symbols and then requests that he or she write down as many of the symbols as possible in boxes directly under the symbol. The examinee is given 90 seconds to write down as many of the symbols as possible. Various combinations of high and low scores can help to understand the underlying processes involved with Digit Symbol-Coding. For example, if a client did poorly on Digit Symbol-Coding and Incidental Learning was high (e.g., good visual memory) but Digit Symbol-Copy was low (e.g., slowed graphomotor speed), it suggests the reason for the poor performance was slow graphomotor speed.

High scorers potentially have excellent visual-motor ability, mental efficiency, capacity for rote learning of new material, and quick psychomotor reactions. Lower scorers may have reduced capacity for visual associative learning, impaired visual-motor functioning, and poor mental alertness.

Block Design

- Analysis of whole into component parts.*
- Spatial visualization.*
- Nonverbal concept formation.
- Visual-motor coordination and perceptual organization.
- Capacity for sustained effort; concentration.
- Visual-motor-spatial coordination; manipulative and perceptual speed.

The Block Design subtest involves nonverbal problem-solving skills because it emphasizes analyzing a problem into its component parts and then reintegrating these parts into a cohesive whole. The examinee must apply logic and reasoning in a manner that will solve spatial relationship problems. As a test of nonverbal concept formation, Block Design demands skills in perceptual organization, spatial visualization, and abstract conceptualization. The Block Design subtest is sturdy and reliable, correlating highly with general intelligence, and is not likely to be lowered except by the effects of depression or organic impairment. Also it has been found to relate to everyday measures of spatial abilities (Groth-Marnat & Teal, 2000). To perform well, examinees must be able to demonstrate a degree of abstraction that is free from literal concreteness. They must also make a distinction between part and whole by demonstrating both analytic and synthetic skills. This test involves an ability to shift the frame of reference while maintaining a high degree of flexibility. The examinee must also be able to inhibit his or her impulsive tendencies and to persist in a designated task.

An important feature of Block Design is that it enables an examiner to actually observe the examinee's response. Some subjects are easily discouraged and give up, while others insist on completing the task even if they have to work beyond the time limit. In approaching the task, one subject might impulsively place the blocks together in a nonrandom sequence, whereas another subject might demonstrate a meticulous sequential style, thereby revealing preferences for either a holistic simultaneous or a more sequential problem-solving style. Additional observations can reveal factors such as hand preference, motor coordination, speed of information processing, frustration tolerance, and ability to benefit from feedback. A highly reflective or compulsive style can lower scores because of the resulting extended time for completing the task. Placing blocks outside the 2×2 or 3×3 configuration is a further behavioral observation that reflects poor visuos-patial skills (J. H. Kramer, Kaplan, & Huckeba, 1999). Thus, potentially valuable information can be obtained by observing and recording differences in solving the Block Design tasks.

Block Design is also a nonverbal, relatively culture-free test of intelligence. It is reliable in that it correlates highly with general intelligence (approximately 52% of its variance may be attributed to *g*), but it has a relatively low correlation with education. Thus, the Block Design subtest is only minimally biased by an examinee's cultural or educational background. Block Design scores can, therefore, be an important tool in assessing the intellectual potential of persons from divergent cultural and intellectual backgrounds.

Block Design is an excellent indicator of right hemisphere brain damage and is especially sensitive to right parietal lesions (Lezak, 1995; Reitan & Wolfson, 1992, 1993). Right lesion patients tend to make errors because they might distort the designs, misperceive aspects of them, or become disoriented when attempting to complete them. In contrast, left lesion patients, particularly if the lesion is in the parietal lobe, are not nearly as likely to have a poor Block Design score. However, when they do, it is likely to be expressed in design simplification, confusion, and a concrete approach to reproducing the design (Lezak, 1995). Inattention (neglect) can be reflected by the examinee's failing to complete the right or left portion of the design. For example, only six or seven of the blocks might be used when attempting to complete a nine-block design (Lezak, 1995). Block Design is typically one of the lowest subtest in Alzheimer's patients. It is sensitive to the early phases of the disease and thus can be useful in differentiating between Alzheimer's and pseudodementing conditions such as depression (Fuld, 1984; La Rue & Jarvik, 1987).

High scorers show a good capacity for visual-spatial perception, visual-motor speed, a good ability to concentrate, and excellent nonverbal concept formation. Low scores suggest poor perceptual abilities, difficulties with visual integration, and problems in maintaining a sustained effort.

Matrix Reasoning (WAIS-III only)

- Visual-spatial reasoning.
- Abstract reasoning.
- Visual organization.
- Simultaneous processing of visual-spatial information.
- Analysis of wholes into component parts.

High scores on Matrix Reasoning suggest good visual information processing and nonverbal abstract reasoning skills. It is combined with Picture Completion and Block Design to form the Perceptual Organization Index. Matrix Reasoning is untimed and is, therefore, useful for persons from older age groups who might do poorly on some of the other timed tests. It also does not penalize those who have a reflective, cautious problem-solving style. Matrix Reasoning is relatively culture free and requires only a minimal amount of visual motor-coordination because the subject merely points to the correct response. Conceptually, Matrix Reasoning is similar to the Halstead Reitan Category Test and Raven's Progressive Matrices. However, future studies will need to determine the nature and degree of correspondence between these measures.

One of the rationales for Matrix Reasoning was to develop a visual-spatial subtest with good psychometric properties that could replace the psychometrically poor Object Assembly subtest. In many ways, this has been realized as Matrix Reasoning has been found to have test-retest stabilities ranging from .75 to .81, SEM of .97, a correlation with the Full Scale IQ of .75, and a factor loading of .61 on the Perceptual Organization Index. It is one of the best performance subtest measures of *g* (52% of its variance can be attributed to *g*). In contrast, Object Assembly has poorer psychometric properties with particular concerns related to its lack of stability (SEM = 1.66). As a result, Object Assembly is now an optional WAIS-III subtest.

High scores might indicate good nonverbal abstract reasoning abilities, a preference for simultaneous processing of information, and excellent visual information processing. Low scores might suggest low visual concept formation, poor or, at least, rigid visual reasoning, or poor concentration. Negativism might be indicated if the examinee seems unmotivated and replies with wording such as "none of them match."

Picture Arrangement

- Planning ability (comprehending and sizing up a total situation).*
- Anticipation of consequences.*
- Temporal sequencing and time concepts.*
- Accurately understanding nonverbal interpersonal situations.
- Ability to comprehend a total situation and evaluate its implications.
- Visual organization and perception of essential visual cues.
- Speed of associating and planning information.

The Picture Arrangement subtest is primarily a test of the ability to plan, interpret, and accurately anticipate social events in a given cultural context. Thus, an individual's cultural background can affect his or her performance on the test; normal subjects with poor or different cultural backgrounds often do poorly. This means that scores derived from such persons should be treated with caution. Wechsler (1958) stated that the test requires an examinee to use general intelligence in nonverbal social situations. In fact, each of the items requires a person to respond to some practical interpersonal interaction. Solving the correct sequence also requires at least some sense of humor. However, interpretive caution should be exercised because most research has not supported relationships between Picture Arrangement and measures of social intelligence (Beebe et al., 2000; Lipsitz et al., 1993). Both Picture Arrangement and Block Design are measures of nonverbal intelligence. However, Picture Arrangement is far more dependent on cultural variables than is Block Design. Picture Arrangement also requires the person to grasp or "size up" the complete situation before proceeding to a correct response. In contrast, persons can achieve good scores on Block Design by approaching the task in small segments and then contrasting their performance on each segment with the whole design.

Picture Arrangement is somewhat sensitive to the effects of brain damage, especially for those injuries that disrupt nonverbal social skills (Golden, 1979; Lezak, 1995). An unusually low Picture Arrangement score in a protocol in which there is little difference between Verbal and Performance IQs implies an organic impairment consistent with a static lesion to the right anterior temporal lobe (Reitan, 1974a; Reitan & Wolfson, 1993). More generalized right hemisphere lesions are likely to lower not only scores on Picture Arrangement, but also performance on Block Design and Object Assembly (Russell, 1979). There is also some evidence that patients with frontal lobe impairment do poorly on Picture Arrangement because of their tendency to respond impulsively and without considering the entire problem (Walsh, 1994).

Two approaches can be followed to obtain additional qualitative information from Picture Arrangement. The first is to observe and record how the person attempts to solve the problem. Does the client carefully consider the overall problem or rather impulsively begin altering the cards? Is the client easily discouraged or does he or she demonstrate a high degree of persistence? After the entire subtest has been completed, an examiner may also want to obtain a subject's description of the stories related to the pictures. This might be initiated by simply asking the examinee to "Tell me what is happening in the pictures" or "Make up a story about the cards." The following questions are especially important: Are the stories logical, fanciful, or bizarre? Are they original or rather stereotyped and conventional? Do examinees reveal any emotional attitudes relating either to themselves or to their interpersonal relationships? Were errors the result of incorrectly perceiving specific details or rather of neglect in even considering certain details? Did the examinee consider all the different relationships in the pictures or were important aspects omitted?

The previous information on Picture Arrangement applies primarily to the WAIS-III rather than the WISC-III because a substantial amount of extra credit for speed was given for the WISC-III revision of Picture Arrangement. It relates quite closely to the Processing Speed factor (along with Coding and Symbol Search; Hishinuma & Yamakawa, 1993; Wechsler, 1991). The practical implication is that WISC-III interpretation of Picture Arrangement scores should emphasize the speed component above or, at least in the context of, Picture Arrangement's other aspects (e.g., understanding nonverbal interpersonal situations).

Persons who score high on Picture Arrangement are usually sophisticated, have a high level of social intelligence, and demonstrate an ability to quickly anticipate the consequences of initial acts. Low scorers may have a paucity of ideas, difficulty planning ahead, slow processing of information, a poor sense of humor, difficulty in interpersonal relationships, and poor rapport.

Symbol Search

- Speed of visual search.*
- Speed of processing information.
- Planning.
- Encoding information in preparation for further processing.
- Visual-motor coordination.
- Learning ability.

Symbol Search was designed to be as pure a test as possible of information-processing speed. It pairs nicely with Digit Symbol-Coding because, conceptually, they assess similar areas, as is more formally indicated by relatively high correlations (WAIS-III, .65; WISC-III, .53) between the two subtests. Together, they form the Processing Speed factor. Symbol Search is psychometrically a relatively good subtest. Test-retest over a 2- to 12-week interval was .79 for the WAIS-III and .76 for the WISC-III. It correlates relatively highly with both Full Scale (WAIS-III, .66; WISC-III, .56) and Performance (WAIS-III, .69; WISC-III, .58) IQs.

High scores suggest that the individual can rapidly absorb information as well as integrate and respond to this information. In addition, it suggests good levels of visual-motor coordination, short-term visual memory, planning, general learning, and a high level of attention and concentration. Low scores suggest slow mental processes; visual-perceptual difficulties; possibly poor motivation and/or anxiety; difficulties with short-term visual memory; and a reflective, perfectionistic, or obsessive problem-solving style.

Object Assembly

- Ability to benefit from sensory-motor feedback.*
- Anticipation of relationships among parts.*
- Visual-motor organization.
- Simultaneous (holistic) processing.
- Synthesis; putting things together in a familiar configuration.
- Ability to differentiate familiar configurations.
- Manipulative and perceptual speed in perceiving the manner in which unknown objects relate to each other.

Object Assembly is a test of motor coordination and control, as are Digit Symbol-Coding and Block Design. It measures the ability to differentiate familiar configurations, and it also involves some anticipation and planning. However, scores are subject to a high degree of fluctuation, primarily because of the potential for accidentally fitting together parts. A related area that may create some confusion is that persons who are in the lower ranges of intelligence (60 to 75) sometimes do quite well, whereas persons with above-average IQs can do quite poorly. The preceding difficulties have resulted in only moderate test-retest reliabilities (WAIS-III, .76; WISC-III, .64 to .71). In addition, Object Assembly is only a moderate measure of general intelligence (WAIS-III, 38%, and WISC-III, 44% of its variance may be attributed to *g*) and is not highly correlated with Full Scale IQ scores (WAIS-III, .59; WISC-III, .58). Furthermore, its correlation with other subtests is generally low. This is why it became an optional subtest for the WAIS-III. Because it is psychometrically one of the poorest subtests, scores should be treated with caution. In addition, it generally lacks a sufficient amount of subtest specificity for adequate interpretation of the test's underlying abilities.

Despite these difficulties, an advantage of Object Assembly is that, as with Block Design and Picture Arrangement, an examiner can directly observe a person's problem-solving style and reactions to success or failure. The test presents an "open" situation, and those who can work freely in this context usually do well. However, those with rigid visual organizations stick with one clue without allowing themselves to change their frame of reference. This inflexibility is often seen with people who are obsessive-compulsive. On the other hand, a flexible visual organization permits a rapid integration of new clues and an adaptation of these clues toward completing the task. The same observations relevant for Block Design are appropriate for Object Assembly. These include persistence, concentration, hand preference, frustration tolerance, speed of processing information, reflectivity, impulsiveness, ability to benefit from feedback, and preference for a simultaneous versus a sequential problem-solving style. In particular, an overly cautious, reflective, and/or obsessive approach is likely to lower performances because of the loss of bonus points resulting from their slow completion of the task.

Persons scoring high on Object Assembly show good perceptual-motor coordination, have superior visual organization, and can maintain a flexible mental outlook. Low scorers show visual-motor disorganization, concreteness, and difficulties with visual concept formation. Like Block Design, Object Assembly is sensitive to right, especially right posterior, lesions (Lezak, 1995; Reitan & Wolfson, 1993). However, given the test's inadequate test specificity and low reliabilities, these interpretations should be somewhat more tentative than for other subtests.

Mazes (WISC-III only)

- Planning ability or foresight.
- Perceptual organization.
- Visual-motor coordination and speed.
- Nonverbal reasoning.

The Mazes subtest is an optional portion of the WISC-III and is not extensively used. Its correlation with the Full Scale IQ is unimpressive (.31), and it is also a poor measure of *g* (9% of its variance may be attributed to *g*). Despite these significant limitations, Mazes can at times provide an additional useful test, particularly with nonverbally oriented children or when a further assessment of planning, sequencing, and perceptual organization is required. Its main advantage is that it is a relatively pure measure of perceptual planning ability.

Individuals with high scores may have an efficient ability to plan ahead and maintain a flexible mental orientation, which further suggests an excellent ability to delay impulsive action (Ireland-Galman, Padilla, & Michael, 1980). Low scores reflect impulsivity and poor visual-motor coordination. Often, unusually low scores may suggest poor reality orientation or organic cerebral impairment, particularly to the frontal areas (Waugh & Bush, 1971).

ASSESSING BRAIN DAMAGE

General Principles

The WAIS-III and WISC-III measure many abilities that are likely to be lowered by brain damage. These include memory, learning, perceptual organization, problem solving, and abstract reasoning. As a result, the Wechsler intelligence scales are typically a core feature of any neuropsychological battery (Groth-Marnat, 2000b; Sullivan & Bowden, 1997). At one time, it was hoped that the Wechsler intelligence scales, along with other more specialized psychological tests, could be used in the actual diagnosis of brain damage. Despite some noteworthy success in this area, it is currently more typical for psychological tests to be used in the assessment of the effects a known lesion is likely to have on a person's cognitive and adaptive functioning. This further highlights the point that the Wechsler intelligence scales, along with other specific tests of neurocognitive ability, are not tests specifically sensitive to brain damage. Rather, they are tests that can reflect the effects of brain damage as well as a variety of other conditions.

During the earlier development of the WAIS and WISC, Wechsler (1958) hoped that brain damage could be discriminated based on relative lowerings in subtests that were most sensitive to neurological impairment. He referred to these brain-sensitive tests as *no-hold* tests (Digit Span, Digit Symbol, Similarities, Block Design) and contrasted them with *hold* tests, which were believed to be far more resistant to impairment (Information, Object Assembly, Picture Completion, Vocabulary). Although the distinction between hold and no-hold tests has some truth, the use of such a distinction in diagnosing brain damage has been found to result in too many misclassifications. Vogt and Heaton (1977) summarized the reasons for this lack of success by pointing out:

- There is no single pattern of brain damage, so it would be expected that highly variable test responses would occur.
- The hold/no-hold distinction does not account for other significant factors, such as the age when the brain damage occurred, environmental variables, education, location of the lesion, and whether the lesion is recent versus chronic.
- Many important abilities related to brain damage still are not measured by the Wechsler intelligence scales.

More recent work supports the theory that there is no specific brain damage profile (Aram & Ekelman, 1986; R. A. Bornstein, 1983; Groth-Marnat et al., 2000; Lezak, 1995; J. Todd, Coolidge, & Satz, 1977). Some persons with brain damage produce low

IQs, whereas for others, IQs are still high. Sometimes, there is a high level of subtest scatter, and, at other times, the scores on the subtests are quite even. Some persons with brain damage produce a high Verbal-Performance split and others do not. This is further complicated because a Verbal-Performance split is more likely to occur for males than for females (R. A. Bornstein & Matarazzo, 1982; A. Kaufman & Lichtenberger, 2002; Lezak, 1995) and for adults but not for children (A. Kaufman, 1994; A. Kaufman & Lichtenberger, 2002; Lezak, 1995). Brain damage may cause a general lowering on all or most subtests and, at other times, there may be a lowering of only specific subtests. The most general indicator for the detection of brain damage is whether a person's scores (either general or specific) are lower than expected given his or her socioeconomic status, age, education, occupation, and other relevant areas of his or her history.

One of the older conventional wisdoms about brain damage is that left hemisphere involvement is more likely to lower the Verbal Scales, whereas right hemisphere involvement results in relatively lower scores on the Performance Scales (see previous discussions under Verbal/Performance IQs, Verbal Scales, and Performance Scales). Reviews of this hypothesis have shown that sometimes this laterality effect has occurred and, at other times, it has not (Aram & Ekelman, 1986; R. A. Bornstein, 1983; A. Kaufman & Lichtenberger, 2002; Larrabee, 1986; Lezak, 1995). On average, right hemisphere lesions produce a V > P discrepancy of 9 points, whereas left hemisphere lesions produce a less marked P > V difference of 4 points (see review by A. Kaufman & Lichtenberger, 2002). Probably the safest approach is that a Verbal-Performance split is not diagnostic of either brain damage in general or, more specifically, damage to one or the other hemisphere. However, a Verbal-Performance split (especially if 15 points or greater) can at times be consistent with this hypothesis. This is especially true if the Verbal-Performance difference is 25 points or greater. More specifically, a lowered Verbal Scale (15 points or greater) suggests the possibility of language impairment. Noteworthy subtests within the Verbal Scales are Arithmetic, Digit Span, and Letter-Number Sequencing (WAIS-III) that, if lowered, suggest difficulties with attending and concentrating. A Performance Scale that is 15 or more points lower than the Verbal Scale suggests impaired perceptual organization abilities. Appropriate caution should be taken to avoid the risk of overinterpreting a person's results and to use further means of investigation, including knowledge of health status, medical history, and additional specialized psychological tests.

Another frequent belief is that brain damage is more likely to lower Performance than Verbal tests. Some good reasons can be given to suggest this may be true. The Performance subtests are timed and, because many persons with brain damage tire easily and have difficulties with concentration and attention, they would be expected to have a particularly difficult time with these tests. Support for this has been found because the Processing Speed Index (Digit Symbol-Coding and Symbol Search) has been lowered with several types of cognitive impairment (D. Fisher, Ledbetter, Cohen, Marmor, & Tulsky, 2000; K. Hawkins, 1998; Psychological Corporation, 1997). From a theoretical perspective, fluid intelligence is tied more to an intact brain structure and also is assessed more clearly by the ongoing problem-solving tasks presented in the Performance subtests. Thus, a destruction of brain tissue would be more likely to lower fluid intelligence, which would be reflected in lowered Performance subtest scores.

This hypothesis can be further assessed by calculating Horn's WAIS-III or WISC-III subtest groupings for fluid intelligence (see "WAIS-III/WISC-III Successive Level Interpretation Procedure" section, Level II, Step c). Although there is some basis for accepting the preceding assumptions, there are also many exceptions. Russell (1979) and Zilmer, Waechtler, Harris, Khan, and Fowler (1992) found that left hemisphere damage caused a lowering in both WAIS/WAIS-R Performance and Verbal subtests, whereas right hemisphere and diffuse damage resulted in the expected lowering, primarily in Performance subtests.

A. Kaufman and Lichtenberger (2002) suggest that an important reason for this relatively small V > P effect for unilateral left lesion patients is that different hemispheres do not so much process different types of information (verbal content versus visual-spatial content), but more that the left hemisphere processes information sequentially whereas the right hemisphere processes information simultaneously (see Springer & Deutsch, 1998). This is supported by the observation that adult left-lesion patients do worst on Arithmetic, Digit Span, and Digit Symbol-Coding, all of which require sequencing (and comprise the WAIS-III/WISC-III Working Memory/Freedom from Distractibility factor). The WAIS-R difference between unilateral left lesion patients' average subtest scores on Perceptual Organization (8.7) and Freedom from Distractibility (6.8) is nearly 2 subscale points. Thus, it might be more useful to assess the relative extent of lowering on unilateral left lesion patients' Freedom from Distractibility than to merely assess the extent of their P > V difference. Future research on the WAIS-III's Working Memory Index (Arithmetic, Digit Span, Letter-Number Sequencing) would also be likely to support these findings.

Many of the inferences related to brain damage depend on profile analysis. Useful material relevant to brain damage can be found in the discussion of Levels II through V under the "Interpretation Procedure" section in this chapter and in the relevant discussions for each subtest in the "Wechsler Subtests" section of this chapter. Much of this interpretation depends on hypothesis testing in which the practitioner integrates knowledge about the person, brain function, Wechsler subtests, and past clinical experience. Often, no clear, empirically based guidelines exist. Accuracy of any inferences are based partially on whether they make neuropsychological sense. However, one generally accepted principle is that intersubtest scatter is most likely to occur with focal lesions of recent origin (A. Kaufman & Lichtenberger, 2002). In contrast, general lowering of all abilities (low subtest scatter) is more likely with either chronic lesions or with diffuse degenerating diseases (e.g., exposure to neurotoxins; Groth-Marnat, 1993; L. Miller, 1993).

One useful strategy developed by Kaplan and her colleagues is to work toward parceling out the underlying processes responsible for scores on the Wechsler intelligence scales (Milberg et al., 1996). Alternative administration guidelines, error categories, useful tables, and interpretive procedures have been developed for both the WAIS-R (E. Kaplan et al., 1991; with plans for the WAIS-III) and WISC-III (E. Kaplan et al., 1999). For example, a clinician might be interested to know if a client's poor performance on Information or Vocabulary resulted from lack of knowledge or problems with retrieval. This might be determined by presenting him or her with multiple-choice formats that assist (recognition of correct answers) them with the retrieval process. If a client does significantly better on the multiple-choice format than the standard format, it suggests that the lowering was caused by retrieval difficulties. The new WAIS-III Digit Symbol-Coding optional procedures (Incidental Learning and Digit Symbol-Copy) were originally derived from Kaplan et al.'s (1991) WAIS-R as a Neuropsychological Instrument (WAIS-R NI) and, as discussed previously, can assist in determining if a poor performance resulted more from poor memory or graphomotor (psychomotor) slowing. Another strategy built in to the process approach is to carefully investigate various error categories (Groth-Marnat et al., 2000; E. Kaplan et al., 1991, 1999). For example, visual neglect might be indicated by not noticing details on the left (usually) side of pictures on Picture Completion or making errors on the left side of the designs for Block Design.

When the preceding strategies, principles, and cautions are taken into account, clinicians can generate and test useful hypotheses developed from different patterns of subtest scores. The following list summarizes some of the most frequently supported hypotheses about specific subtests or patterns of subtests:

- Digit Symbol-Coding is the most brain-sensitive Wechsler subtest and can be lowered by lesions in any location. A lowering implies difficulties with speed of information processing and/or learning, sequencing, rote learning, concentration (especially with lowerings in Digit Span and Arithmetic), visual-motor abilities, and speed of processing or learning (Lezak, 1995; Reitan & Wolfson, 1992). The WAIS-III combination of Digit Symbol-Coding and Symbol Search (Processing Speed Index) has been found to be the most frequently lowered group of subtests among a wide variety of brain-impaired populations (K. Hawkins, 1998; Psychological Corporation, 1997).
- Block Design is also brain sensitive, especially to either left or right parietal lesions (Golden, 1979; Lezak, 1995; McFie, 1960, 1969). A lowering implies visual-spatial problems (especially combined with a lowering in Object Assembly) and possible difficulty in constructing objects (constructional apraxia: note quality of drawings; J. H. Kramer et al., 1999; Zilmer, Bell, Fowler, Newman, & Stutts, 1991).
- Picture Arrangement lowering is consistent with right anterior temporal and possibly right frontal lesions (Reitan, 1974b; Reitan & Wolfson, 1993; Russell, 1979). In some cases, Picture Arrangement might also be lowered by left hemisphere lesions if there is a resulting impairment in following directions and/or conceptual skills.
- Both Digit Span and Arithmetic are frequently lowered in brain-damaged populations, particularly with left hemisphere lesions (A. Kaufman & Lichtenberger, 2002; Lezak, 1995; McFie, 1960, 1969). Lowering suggests poor concentration and attention and, if Digits Backward is significantly lower than Digits Forward (generally 5 or more digits), a significantly reduced level of mental flexibility and/or difficulty forming and maintaining a visual image of the digits. It may also suggest difficulties in a person's executive functions related to selecting a key stimulus, attending to it, and maintaining the information in short-term storage, while simultaneously performing other mental tasks (Wielkiewicz, 1990).
- Processing Speed (composed of Symbol Search and Digit Symbol-Coding) is the subtest that is most sensitive to the impact of most forms of cognitive impairment (D. Fisher et al., 2000; K. Hawkins, 1998; Psychological Corporation, 1997.)

- Vocabulary, Information, and Picture Completion have often been used as a rough estimate of a person's premorbid level of functioning because they are relatively unaffected by lesions. An important exception is that children who are brain damaged often score lowest on the Vocabulary subtest (Boll, 1974; Reitan, 1974b; Reitan & Wolfson, 1992). In addition, Information and Vocabulary are generally lowered (especially relative to Similarities) in patients with left temporal damage, suggesting difficulties with word comprehension, retrieval, and language expression (Dobbins & Russell, 1990). Another hold test, Picture Completion, while usually resistant to brain damage, might be lowered because of difficulties involving vision, especially visual agnosia (difficulty recognizing objects; E. Kaplan et al., 1991, 1999). Thus, always considering Vocabulary, Information, and Picture Completion as indicators of premorbid functioning can potentially result in incorrect inferences and should be interpreted in relation to what is known about brain-behavior relationships.
- The Similarities subtest, especially in relation to Information and Vocabulary, is most likely to be lowered with left frontal lesions and suggests difficulty with verbal reasoning and verbal concept formation (Dobbins & Russell, 1990).
- Qualitative responses, particularly related to error categories (even when the subtests are not lowered), can provide useful information related to brain damage. Some responses might suggest poor judgment and impulsivity, whereas others might indicate concrete thinking in which the person is bound by the stimulus value of the item (e.g., winter defined as "wet, cold" rather than the more abstract reference to a season; or the clang response that "ponder" means "to pound"). Other persons might report they once knew the answer but have forgotten, which can be assessed through WAIS-R NI/WISC-III PI multiple-choice options. Diffuse brain damage (but not focal) might also be consistent with a high degree of intratest scatter in which the client misses easy items but correctly answers later, more difficult ones (Mittenberg et al., 1989). This suggests retrieval failure and/or the random loss of previously stored information. This intrasubtest scatter is most likely to occur on Vocabulary, Comprehension, Information, Similarities, and Picture Completion.

Estimating Premorbid IQ

Neuropsychologists are frequently confronted with the need to estimate a client's premorbid level of functioning. In an ideal situation, previous IQ results derived before the injury could be obtained and compared with his or her current level of functioning. Even in this situation, clinicians should be aware that a decline in *overall* performance should not be inferred unless there is a significantly lower current IQ than had been obtained from a premorbid IQ assessment. A discrepancy of 12 or more WAIS-R Full Scale IQ points would result in an 80% accurate detection of adults (WAIS-III) who had actually suffered a cognitive decline (Graves, Carswell, & Snow, 1999). It should also be stressed that there still might be quite specific areas of decline that are not sensitive to the global measure of IQ scores.

In most cases, premorbid IQ results are not available; therefore, clinicians must rely on other strategies to infer premorbid ability. These strategies include historical achievement-based records, current measures of ability that are not sensitive to decline ("hold" measures), demographic-based regression equations, or a combination of these. Useful historical records might include grade point average, SAT scores, work achievement records, achievement tests, or peer ratings. The age of the person, as well as relevant aspects of the injury (i.e., size and location of the lesion, recency of injury), might also be important to consider.

A further strategy for estimating premorbid ability is to note performances on Wechsler subtests that are considered most resistant to neurological impairment (Information, Picture Completion, and especially Vocabulary). As discussed previously, these subtests have often been considered to reflect the person's past level of functioning and are, therefore, referred to as *hold* subtests. Administering an achievement test such as the Wide Range Achievement Test (WRAT-III) or Wechsler Individual Achievement Test (WIAT) might also accomplish a similar purpose. One difficulty is that for many clients, especially those who are well educated, this method is likely to overestimate premorbid IQ. In contrast, it would be likely to underestimate premorbid IQ for subgroups whose premorbid Performance Scales are typically greater than Verbal Scales (i.e., Native Americans, Hispanics, bilinguals, persons with low educational attainment, blue-collar workers).

A related technique is to consider the person's two or three highest subtests (regardless of whether the subtests are brain-sensitive or non-brain-sensitive) and then use these to estimate the person's premorbid level of functioning. Despite its occasional usefulness, this procedure is likely to result in a high number of misclassifications because it does not consider crucial factors such as the person's age, educational level, or location of the lesion (Matarazzo & Prifitera, 1989).

A variation of this hold procedure is to use a reading test such as the National Adult Reading Test (NART; H. Nelson & Williams, 1991) or Wechsler Test of Adult Reading (WTAR; Wechsler, 2001). The NART and WTAR were designed by selecting 50 irregularly spelled words (i.e., yacht, naive) that are unlikely to be pronounced correctly unless the client has previous knowledge of the words. This relatively pure recognition task places minimal demands on problem-solving abilities. A NART-estimated WAIS-R Full Scale IQ 20 points higher than a person's obtained IQ suggests intellectual decline (80% accuracy for those with actual decline; Graves et al., 1999). However, this assumes that the injury would not have affected the person's reading ability. The WTAR has the advantage that it has been co-normed with the WAIS-III and WMS-III. Despite their usefulness, the previous caveats related to demographics (ethnicity, education) would also be relevant for reading tests such as the NART/NART-R and WTAR.

Other efforts to determine premorbid IQ have used regression equations based on demographic variables (education, occupation, etc.). One of the most extensively researched is the Barona Index (Barona, Reynolds, & Chastain, 1984). To correctly classify (80% accuracy) clients with true cognitive decline, a discrepancy of 25 IQ points would be required (Graves et al., 1999). Unfortunately, this discrepancy is sufficiently large such that other more straightforward procedures (i.e., previous work performance, grade point average, medical records) would be likely to be more accurate. In addition, the index is likely to be inaccurate for persons with either extremely high (above 120) or extremely low (below 69) IQs (Barona et al., 1984; Graves et al., 1999; Veiel & Kooperman, 2001), and the formulas are likely to overestimate most premorbid IQ levels (Eppinger, Craig, Adams, & Parsons, 1987).

A final strategy is to combine various measures such as the NART and demographics or performance on specific subtests with demographics. Such procedures have generally resulted in slight incremental increases beyond the NART or demographics alone (Grave et al., 1999; Vanderploeg, Schinka, & Axelrod, 1996). Vanderploeg et al. found that the best predictor of WAIS-R Full Scale IQ could be made by calculating the following three regression equations and using the one that resulted in the highest IQ estimate (BEST 3 approach):

$$\text{WAIS-R FSIQ} = 3.55\ (\text{Information}) + 1.00\ (\text{SES}) + 58.70$$

$$\text{WAIS-R FSIQ} = 3.78\ (\text{Vocabulary}) + 0.70\ (\text{SES}) + 59.09$$

$$\text{WAIS-R FSIQ} = 2.94\ (\text{Picture Completion}) + 2.13\ (\text{SES}) + 1.62\ (\text{Age}) + 49.41$$

These calculations can be made by inserting the following variable codes:

Age: 16–17 years = 1; 18–19 = 2; 20–24 = 3; 25–34 = 4; 35–44 = 5; 45–54 = 6; 55–64 = 7; 65-69 = 8; 70–74 = 9

Education: 0–7 years = 1; 8 = 2; 9–11 = 3; 12 = 4; 13–15 = 5; 16+ = 6

Occupation: Unemployed = 1; farm laborers, farm foreman, and laborers (unskilled) = 2; operatives, service workers, farmers, and farm managers (semiskilled) = 3; craftsmen and foremen (skilled workers) = 4; managers, officials, proprietors, clerical, and sales workers = 5; professional and technical = 6

SES: Sum of education code and occupation code (If unemployed, SES = 2 × Education)

The correlation with the actual Full Scale IQ is .84, and the standard error of estimate was 9.10 using the equation with Information, 8.64 for Vocabulary, and 9.57 for Picture Completion. To infer overall cognitive decline, discrepancies of 18 points or more should be documented. This is clearly superior to the estimated 25-point discrepancy required for the Barona index. However, these formulas were calculated using the WAIS-R. Because WAIS-III Full Scale IQ scores are, on average, 3 points higher than scores derived from the WAIS-R (and, therefore, estimated by these equations), an additional 3 points (21 points in total) would be needed to infer cognitive decline if current IQ scores were obtained with the WAIS-III. In addition, the BEST 3 approach tends to slightly overpredict scores at the low IQ range but underpredict estimates in the high IQ range.

In contrast to adult BEST-3 premorbid estimates, research with children has found that an equation based on demographics alone is equally as effective in differentiating people with brain damage from non-brain-damaged persons as equations using a combination of demographics and WISC-III subtests (Vanderploeg, Schinka, Baum, Tremont, & Mittenberg, 1998). Thus, the following formula based on demographics alone is recommended:

$$\text{FSIQ} = 5.44\ (\text{Mean parental education}) + 2.80\ (\text{White/non-White}) - 9.01\ (\text{Black/non-Black}) + 81.68$$

This equation can be calculated by inserting the following variable codes:

Mean parental education: 0–8 years = 1; 9–11 = 2; 12 years (or GED) = 3; 13–15 years = 4; 16+ = 5

Ethnicity: Two coded variables: White/non-White (White = 1; non-White = 0) and Black/non-Black (Black = 1; non-Black = 0). Hispanics would be uniquely coded as 0 on both White/non-White and Black/non-Black (the regression equation should not be used for ethnic groups other than White, Black, or Hispanic).

However, when using a discrepancy cutoff of 13, only 64% of people with brain damage were correctly classified and 89% of normal controls were correctly classified (Vanderploeg et al., 1998).

As would be expected, estimating premorbid IQ has been a controversial procedure, particularly in a forensic context (see Veiel & Koopman, 2001). The following review points seem crucial. First, the previous equations should be used to supplement but not replace a careful evaluation of crucial information such as work history and medical records. In addition, formal cutoffs should be used. Rarely, for example, would an obtained IQ 5 to 10 points below the estimated "premorbid IQ" suggest actual cognitive decline in a person's *overall* ability. However, this still does not preclude the possible presence of quite specific deficits (i.e., facial recognition, short-term visual memory). The likelihood of errors increases when equations based on demographics or subtests are used with persons with IQs suspected of being extremely high or extremely low (below 80 or above 120).

Alzheimer's Disease

The initial symptoms of Alzheimer's disease are characterized by apathy, a decline in short-term memory, and difficulties with problem solving. Underlying these changes are reductions in cholinergic activity. Currently, neuropsychological assessment, particularly with the Wechsler intelligence scales, is one of a variety of diagnostic procedures to enhance diagnosis. Nonverbal abilities seem to be more sensitive to impairment than verbal abilities. Earlier research with the WAIS-R found that a full 52% of Alzheimer's disease patients had Verbal greater than Performance scores of 15 points or more (Fuld, 1984). Similarly, WAIS-III Verbal scores have been found to be 10 points higher than Performance subtests for a group of patients with "probable" Alzheimer's (Psychological Corporation, 1997). The lowest index scores were for Processing Speed (mean = 79.6) with some lowerings in Perceptual Organization (mean = 84.8) and Working Memory (mean = 87.2).

A specific WAIS-R Alzheimer's profile developed by Fuld (Fuld, 1983, 1984) found that Information and Vocabulary were relatively higher than Similarities and Block Design, and Digit Symbol and Block Design were lowest. This pattern makes conceptual sense in that Information and Vocabulary are relatively resistant to deterioration, reflect crystallized abilities, and are correspondingly the highest subtests in the profile. In contrast, Digit Symbol and Block Design are relatively sensitive to deterioration, reflect areas of fluid intelligence and, along with Object Assembly, are the lowest subtests in the profile. An extensive review of the Fuld profile using 18 studies concluded that sensitivity (proportion of true positives) to Alzheimer's disease was a very low 24.1% (Massman & Bigler, 1993). In contrast, the profile's specificity (proportion of true negatives) was 93.3%. This means that more accurate diagnoses are likely to be achieved through using the WAIS-III in combination with specific measures of memory (i.e., WMS-III) or specialized dementia batteries (i.e., CERAD battery). In addition, research on the Fuld or similar profiles needs to be conducted with the WAIS-III.

ASSESSING ADDITIONAL SPECIAL POPULATIONS

Learning Disabilities

Learning disabilities make up a complex, heterogeneous, loosely defined disorder with a wide variety of manifestations and many different theories regarding causation (A. Kaufman & Kaufman, 2002; Sattler, 2002; L. Siegel, 1999). A central component of all definitions is that learning disabilities involve difficulties in developing skills in reading (most commonly), writing, listening, speaking, reasoning, spelling, or math. This is sometimes summarized as poor information processing. Further essential features are these: Learning-disabled persons have adequate intelligence, show a significant discrepancy between achievement and intellectual ability, and have a disorder that is considered primarily intrinsic to the person, presumably because of central nervous system dysfunction. The underachievement cannot be primarily the result of an intellectual disability (mental retardation), brain damage, behavior problems, sensory handicaps, or environmental disadvantage.

The major purpose of learning disability assessment is to identify a client's strengths and weaknesses to be able to decide on an appropriate placement and to design an optimal program. Relevant areas to assess include developmental-cognitive processes, achievement, environmental demands, reactions of others to the client's difficulties, and the possible interaction of additional factors, such as fear of failure, overall level of interpersonal adjustment, and family history of similar difficulties. The Wechsler scales are typically considered essential as a means of identifying the client's overall level of functioning and specific cognitive strengths and weaknesses and to eliminate the possibility of intellectual disability (mental retardation). Other tests are usually required; for example, achievement tests, measures of adaptive behavior, visual-motor tests, assessments of auditory and visual processing, and measures of emotional and behavioral problems (see L. Siegel, 1999).

Considerable effort has been placed into searching for a specific Wechsler scale profile that is unique to learning-disabled populations (see Level IIIb in "Interpretation Procedure" section). There is some evidence for a WAIS-III ACID profile (Arithmetic, Coding/Digit Symbol, Information, and Digit Span) in that 24% of those diagnosed with learning disabilities had a partial (three out of the four subtests as the lowest scores) ACID profile and 6.5% had a full (all four of the subtests as the lowest) ACID profile (Psychological Corporation, 1997). This is higher than the standardization sample. The WAIS-III index scores of Working Memory and Processing Speed (compared to Perceptual Organization and Verbal Comprehension) were also found to be particularly low among a sample of adults diagnosed with reading disabilities (Psychological Corporation, 1997). This has led A. Kaufman and Lichtenberger (1999, 2002) to suggest the possible utility of combining the five subtests in these lowest indexes into a SCALD profile (Symbol Search, Digit Symbol-Coding, Arithmetic, Letter-Number Sequencing, Digit Span). The ACID profile has also received some support with the WISC-III in that most studies have found that approximately 20% of persons with learning disabilities had either a partial or full ACID profile (Mayes, Calhoun, & Crowell, 1998; A. Kaufman, 1994; A. Kaufman & Lichtenberger, 2002; Stanton & Reynolds, 1998).

A somewhat similar WISC-III profile substitutes the new Symbol Search subtest for Information, resulting in the SCAD (Symbol Search, Coding, Arithmetic, Digit Span) profile. These four subtests emphasize the functions of speed of information processing, visual short-term memory, and visual-motor coordination (Symbol Search and Coding), as well as number ability and sequencing (Arithmetic and Digit Span). These are specifically the types of functions that many learning-disabled individuals (as well as many other types of persons with brain dysfunctions) have difficulty with. Accordingly, children with learning disabilities and attention deficit disorder have been found to score particularly low on the SCAD profile (A. Kaufman, 1994; Mayes et al., 1998; Stanton & Reynolds, 1998). Similarly, children diagnosed with ADHD have performed relatively poorly on the WISC-III Freedom from Distractibility factor (Anastopoulos, Spisto, & Maher, 1994). This finding should be used with caution, however, because a relatively large proportion of children with ADHD still do not have this profile. In addition, S. Ward, Ward, Hatt, Young, and Mollner (1995) did not find support for the SCAD profile among learning-disabled children.

A further approach to understanding learning disabilities and related disorders is using Bannatyne's categories, which conceptualize learning-disabled performances as highest on subtests requiring spatial abilities (Object Assembly, Block Design, Picture Completion) in which little or no sequencing is required (Bannatyne, 1974). Conceptual skills are intermediate (Comprehension, Similarities, Vocabulary), and subtests requiring sequencing abilities (Digit Span, Digit Symbol-Coding, Picture Arrangement) are lowest. Thus, their spatial abilities are believed to be greater than their conceptual abilities, which, in turn, are greater than their sequential abilities. A fourth category, Acquired Knowledge (Information, Arithmetic, Vocabulary) is also sometimes used as a rough index of the extent to which the person has accumulated school-related facts and skills (see Level IIIc of "Interpretation Procedures" section). Even though these findings might suggest a greater degree of subtest scatter among learning-disabled persons, this has not been supported by research (Greenway & Milne, 1999).

Collectively, the preceding profiles suggest that many learning-disabled individuals perform best on tasks requiring holistic, right brain, simultaneous processing (Object Assembly, Picture Completion, Block Design) and worst on those requiring sequential processing (Digit Span, Digit Symbol/Coding, Picture Arrangement), which is expressed in difficulties with planning, reading, and numerical ability. Wielkiewicz (1990) has further suggested that these subtests indicate a poorly functioning executive ability in which the individual experiences difficulty attending to stimuli while simultaneously performing other mental tasks.

Reviews and cross-validation of Bannatyne's and ACID/SCAD profiles have produced inconsistent results (see Groth-Marnat, 2002). Only some groups of learning-disabled students in some studies showed the Bannatyne Spatial > Conceptual > Sequential pattern (Katz et al., 1993; A. Kaufman, 1994; A. Kaufman & Lichtenberger, 2002). This is not surprising given the many different modes of expression found under the umbrella term of "learning disabilities" (A. Kaufman & Kaufman, 2002). In addition, Bannatyne's pattern has not been found to be unique to learning disabilities, but frequently occurs in a diverse number of groups including juvenile delinquents and emotionally handicapped children (see Groth-Marnat, 2002). Although only minimal support exists for Bannatyne's categories as a diagnosis for learning disabilities, they are far from useless. The four categories (Spatial, Conceptual, Sequential, Acquired Knowledge) can be invaluable for interpreting relative strengths and weaknesses for learning-disabled persons as well as for other groups. While research has not been able to produce a unique "learning-disabled profile," the research invested in this effort has resulted in a useful means of analyzing Wechsler scale profiles.

Given the previous research, the following conclusions are warranted (adapted from Groth-Marnat, 2002):

- The Full Scale IQ can be most appropriately used in the assessment of persons with learning disabilities to estimate their overall potential and assist in excluding possible explanations for poor academic performance, other than learning disabilities (i.e., intellectual disabilities/mental retardation).
- There is moderate-to-equivocal evidence that some profiles (relatively low Processing Speed and Working Memory/Freedom from Distractibility, Spatial > Conceptual > Sequential, ACID, SCAD, SCALD) occur more frequently in learning-disabled populations compared to the general population.
- These profiles are not unique to learning disabilities but often occur in other groups as well (juvenile delinquents, ADHD, emotionally handicapped).
- If a person does have a "learning-disabled" Wechsler profile (ACID, etc.), it is *consistent with, although not necessarily diagnostic of,* learning disabilities.
- The majority of learning-disabled persons *do not* have Wechsler "learning disabled" profiles. Thus, the absence of one of the profiles *does not exclude* a diagnosis of learning disabilities.
- The various patterns of Wechsler subtests can, at times, be used to further understand individual cases of persons experiencing learning difficulties.

Mental Retardation (Intellectual Disability)

Mental retardation (intellectual disability) is a nonspecific, heterogeneous disorder that occurs during a person's early developmental stages (birth to 18 years; J. Jacobson & Mulick, 1996). It is defined in part as involving subaverage general intellectual performance, which in turn is defined as less than 2 standard deviations below average. Of equal importance are difficulties in adaptive behavior, and any assessment of intellectual disability must demonstrate both a low intelligence level (2 standard deviations below the mean) and evidence that the person cannot function independently or deal effectively with day-to-day life problems (American Psychiatric Association, 1994).

This must include at least two adaptive skill areas including communication, self-care, home living, social skills, community use, self-direction, health and safety, functional academics, leisure, and work (J. Jacobson & Mulick, 1996). Classification of mental retardation (intellectual disabilities) should identify the person's psychological and emotional strengths and weaknesses, overall physical health, and current environmental placement. The American Association of Mental Retardation (AAMR) guidelines (see J. Jacobson & Mulick, 1996) stress that this should lead to a profile that places less emphasis on describing the level of disability (mild, moderate, severe) and more on identifying the types and intensities of supports required by the person. These might be intermittent, limited, extensive, or pervasive. Thus, there has been a recent move away from describing the disability in favor of using information about the person to identify how the person's functioning could best be optimized by the best support available for the person. With appropriate supports, the person's functioning should be able to improve over time. In addition, assessment should take into consideration cultural and linguistic diversity, the context of the community environment, and balance out the individual's adaptive limitations with his or her adaptive skills and personal capabilities (see Schalock et al., 1994).

The AAMR guidelines emphasize the interaction of the person with the environment and, in particular, they encourage any assessment to focus on the level and intensity of required support with a philosophy of empowering the person. As such, there has been a relative deemphasis on the global IQ score, along with the elimination of person-oriented levels of disability. This does not mean that IQ scores are not important, but there is more of a focus on treatment and community-oriented descriptions. In somewhat of a contrast to this trend are the guidelines in the *DSM-IV* (1994), which continue to classify the degree of severity (and corresponding diagnostic code) based on the following IQ ranges: 50–55 to 70 (mild), 35–40 to 50–55 (moderate), 20–25 to 35–40 (severe), below 20–25 (profound) and severity unspecified (mental retardation is presumed to exist but intelligence is untestable by standard tests). The implications are that, for most contexts, clinicians should follow the AAMR guidelines because they are more useful, more clearly tied to recommendations, represent the most current thinking in the field, and are in accordance with national recommendations. However, there may be certain situations in some contexts where *DSM-IV* guidelines might be required.

Although mental retardation (intellectual disability) is a heterogeneous disorder, there is consensus that it consists of two general categories. Nonorganic (or familial) retardation is caused by low genetic inheritance, poor environment, and possibly some organic factors. Persons with familial retardation constitute the upper realms of intelligence (50 to 69) and adaptive functioning among persons with intellectual disabilities and can be educated. Organic retardation is frequently severe (IQ less than 50) and is more closely associated with neurological dysfunction and genetic impairment. Persons with this disorder typically require more supervision and care but are typically able to be taught to manage some routine day-to-day activities.

A typical assessment battery for the diagnosis and assessment of mental retardation (intellectual disability) includes the WISC-III or other individually administered intelligence tests (K-ABC, Stanford-Binet), an achievement test (Wide Range Achievement Test-III, Wechsler Individual Achievement Test, Kaufman Test of Educational Achievement), and measures of adaptive functioning (Adaptive Behavior Assessment System, AAMD Adaptive Behavior Scale, or Vineland Adaptive Behavior Scales).

Further information from interviews, behavioral observations, and medical records are also essential. An important purpose of a test such as the WISC-III is to establish the client's general intelligence level so that it can be placed into the context of other relevant information. The AAMR guidelines point out that, when determining the cutoff IQ for diagnosis, the range of error of the test should be taken into account. This means that the IQ cutoff criteria are somewhere between 70 and 75. The most difficult subtests for mentally retarded persons are Information, Arithmetic, and Vocabulary (primarily the Verbal Comprehension factor), while the easiest subtests are Picture Completion and Object Assembly (primarily the Perceptual Organization factor; Mueller, Dash, Matheson, & Short, 1984).

Gifted Children

Gifted children are frequently defined as having Verbal or Performance IQs of 130 or higher. Children who have a single outstanding ability such as art, music, or math are also frequently classified as gifted even though their IQs may not necessarily be above 130. A further caution is that the WISC-III places considerable emphasis on speeded performance. Thus, a person who was generally gifted, but did not express this giftedness in a rapid manner, may not do particularly well on the WISC-III. Although the WISC-III might be frequently used to identify giftedness, the Stanford-Binet may be somewhat more effective because it has a higher ceiling than the WISC-III. However, neither may be particularly good if a single outstanding ability is used to determine whether a particular child is gifted. Additional assessment strategies for children should include samples of their work, achievement tests, rating forms, or designation by a highly qualified person.

An essential goal of assessing for giftedness is to optimize (rather than "normalize") the child's abilities so that a greater likelihood exists that the child will eventually make a significant contribution to society. This implies that the assessment can recommend an appropriate educational placement and provide general guidelines for program planning. IQ, in itself, is in many ways a limited definition of giftedness. Many persons with extremely high IQs do not accomplish anything of significance. A high IQ (or outstanding talent in a specific area) is merely one of a variety of prerequisites. The interactions of internal motivation, discipline, and environmental opportunities, such as appropriate instruction, are of equal importance.

Caution should also be used with tests such as the WISC-III to assess gifted persons who demonstrate high creativity. Often, highly intelligent people are not particularly creative, which is supported by the low correlation between intelligence tests and creativity (Amabile, 1983). For abilities such as artistic or musical creativity, measures outside IQ testing may prove to be of greater importance. These might include a list of creative achievements, nomination by a qualified person, and specific tests of creativity.

Ethnic Minorities

Intelligence tests have frequently been criticized for being limited in assessing ethnic minorities. A detailed discussion of this issue is included in Chapter 2 (see "Test Bias and Use with Minority Groups" section). However, several additional guidelines should be noted. Often, it is essential to be familiar with the values and beliefs of the client's culture as well as relevant research. This is underscored by the observation that the degree of cultural difference between an interviewer and client has been found to be related to the amount of inaccurate perceptions (Malpass & Kravitz, 1969; P. Shapiro & Penrod, 1986). A clinician should determine the language most familiar to the client and establish the extent and manner in which any language difference might bias the test results. Of related and equal importance is the degree to which clients have assimilated into the dominant culture. Directions and pronunciation should be particularly clear. The examiner also needs to pay particular attention to the importance of rapport and motivation.

Probably the most important strategy is to maintain a flexible attitude, combined with the use of alternative assessment strategies. This strategy might include a variety of nonverbal techniques, such as the Universal Nonverbal Intelligence Test (Bracken & McCallum, 1998), Raven's Progressive Matrices Test, or emphasis on the Performance Scales of the WAIS-III/WISC-III. In addition, dynamic testing shows promise in assessing the extent to which a client can benefit from various ongoing learning opportunities (learning potential; Grigorenko & Sternberg, 1998). Material beyond merely tests should also have a greater significance (teacher reports, discussions with parents, history, behavioral observations).

SHORT FORMS

Dozens of short forms for the Wechsler intelligence scales have been developed to provide a more time-efficient means of estimating IQ. These short forms reduce administration time by either giving selected subtests or deleting specific items (early easy ones, odd or even items). Although time-efficient, these short forms tend to provide less information about a person's cognitive abilities, produce a wider band of error than a full administration, result in less clinical information, and are often of questionable accuracy when used for intelligence classifications (Silverstein, 1990). However, short forms can serve appropriately as screening devices, which are best used when the purpose of evaluation is other than for intellectual assessment. The results can be used either as a rough indicator of intelligence, or as a basis for determining whether a more complete cognitive assessment is necessary. None of the short forms should be confused with a full intellectual assessment or even with a valid indicator of IQ (J. Ryan & Ward, 1999). For this reason, it is important to clearly specify on the report that the indicated IQ is an estimate (indicate as *Est* next to the IQ score) and that a "brief WAIS-III/WISC-III" was given. If this is not specified, the IQ derived from the short form may be confused with a full administration and later decisions may be incorrectly based on the misleadingly described results.

The basic requirement for any short form is a minimum correlation of .90 with the full administration. Even at the .90 level, the band of error is considerably wider than for an IQ derived from a full administration. Calculations indicate that at a .90 correlation, two thirds of the IQs fall within 9 points of a person's actual IQ and a full one third are 10 or more points away from the actual IQ (L. Schwartz & Levitt, 1960). In addition to these psychometric considerations, short forms might be selected based on the type of clinical information needed, or special client characteristics (i.e., handicapped, non-English-speaking background).

Many clinicians calculate short form IQs by prorating the subtest scores—by calculating the mean subtest score for the subtests that were given. This mean can then be multiplied by the total number of Performance, Verbal, or Full Scale subtests to derive the equivalent of the Verbal, Performance, and/or Full Scale sum of scaled scores. Once this estimate of sum of scaled scores has been determined, relevant tables in the manual(s) can be consulted to determine the estimated IQs. Unfortunately, prorating may produce error by failing to consider the relative reliabilities of the different subtests that were used. To counter this, the WAIS-III manual allows examiners to prorate sums of scaled scores based on five Verbal and four Performance subtests (see Table A.10 in the *WAIS-III Scoring and Administration Manual*). In addition, Sattler (2001) has provided a formula (see Sattler, 2001, pp. 256–257, Exhibit 8–4) for obtaining deviation IQs from short forms. He has also provided tables for converting scores on various combinations of short forms into IQs (see pp. 828–835 for the WAIS-III and pp. 774–782 for the WISC-III).

Wechsler Abbreviated Measure of Intelligence (WASI)

The Psychological Corporation developed the Wechsler Abbreviated Scale of Intelligence (WASI; Psychological Corporation, 1997) as a means of providing clinicians and researchers with a short, reliable measure of intelligence linked to the WAIS-III (and WISC-III). The WASI includes four subtests (Vocabulary, Similarities, Block Design, and Matrix Reasoning), which have a similar format and similar content as the WAIS-III subtests with the same names. The selection of these subtests was based in part on high loadings on *g,* along with evidence suggesting bilateral hemispheric activation on most complex cognitive tasks (Springer & Deutsch, 1998). The WASI yields both Verbal and Performance IQs, as well as a Full Scale IQ. The WASI was nationally standardized using a population ranging between ages 6 and 89. Because the subtests were linked to the longer Wechsler intelligence scales, the WASI provides reliable estimates of full WAIS-III and WISC-III IQs. Administration time can be reduced even further by using a two-subtest form (Vocabulary and Matrix Reasoning), which takes approximately 15 minutes but yields only a Full Scale IQ estimate.

Best Two- and Three-Subtest Short Forms

One of the most frequently used two-subtest WAIS-III/WISC-III short forms uses Vocabulary and Block Design. Administration time is approximately 20 minutes and correlations with the full-administration Full Scale IQ are generally in the .90 range (Sattler, 2001). In two-thirds of the cases, IQs fall within 7 points of a person's actual IQ, and one-third of the scores have an error of eight points or greater. Conceptually, Vocabulary and Block Design are good tests to use because they are both good measures of *g,* are quite stable, and represent a sample subtest from both the Performance and Verbal scales. However, research with the WAIS-R suggests it might potentially underestimate the IQs of African Americans because these two subtests are typically their lowest scores (A. Kaufman et al., 1988). Furthermore, persons with high IQs are likely to have a greater margin of error when short forms are used to estimate their IQs because of the greater degree of subtest scatter among this subgroup (Matarazzo, Daniel, Prifitera, & Herman, 1988). If examiners wish to add a third subtest, the inclusion of Similarities, Information, Comprehension, Picture Arrangement, and Picture Completion have each been found to increase correlations into the low .90s (Sattler, 2001). An "amazingly short" form made up of the very short administration time subtests of Information and Picture Completion (WISC-III conversion to standard score = 2.9 (I + PC) + 42; A. Kaufman, Kaufman, Ramaswamy, & McLean, 1996) has been found to have correlations in the mid .80s to low .90s (Sattler, 2001).

Best Four-Subtest Short Forms

A possible four subtest combination includes Vocabulary, Arithmetic, Block Design, and Picture Arrangement. Correlations with the Full Scale IQ range from .93 to .95 for the WAIS-III and WISC-III (Sattler, 2001). Research with the WAIS-R indicated that these four subtests are usually excellent in detecting abnormal cognitive functioning (J. Ryan, Georgemiller, & McKinney, 1984). The inclusion of Arithmetic with Vocabulary and Block Design provides an assessment of auditory attention, along with an important indicator of how effectively the person functions in the real world. Picture Arrangement provides information on a person's knowledge of sequencing and his or her relative perceptiveness about common social situations. An important caution is that any short-form combination of Vocabulary, Block Design, Arithmetic, or Picture Arrangement is likely to overestimate the IQs of patients referred for neuropsychological evaluation (Roth, Hughes, Mankowski, & Crosson, 1984). Additional short forms using any four combinations of Vocabulary, Block Design, Arithmetic, Matrix Reasoning, Picture Arrangement, Information, Comprehension, Similarities, or Picture Completion are also likely to produce correlations with the Full Scale IQ in the low to mid-.90s (Sattler, 2001).

A. Kaufman et al. (1996) evaluated WISC-III four-subtest short forms based on clinical, practical, and psychometric considerations and recommended that the overall best tetrad was composed of Similarities-Arithmetic-Picture Completion-Block Design. Total administration time was approximately 27 minutes, scoring time was relatively brief, and it was found to be as psychometrically sound as other combinations. Conversion formulas to estimate Full Scale IQs use the sum of the four scaled scores (S + A + 1 + PC + 1 + BD; abbreviated as simply Xc) but vary according to the following age groups: ages 6, 8 to 14, 16, and total sample (1.6 + Xc + 1 + 36); age 7 (1.7 + Xc + 1 + 32); and age 15 (1.5Xc + 1 + 40).

Seven-Subtest Short Forms

One strategy is to delete the most time-consuming subtests and give as many of the shorter subtests as possible. J. Ryan and Ward (1999) developed a WAIS-III seven-subtest short form (Information, Digit Span, Arithmetic, Similarities, Picture Completion, Block Design, Digit Symbol-Coding), which takes 35 minutes to administer. A slight variation from this short form is to substitute Matrix Reasoning for Block Design. This has the advantage of providing a slightly more accurate estimate of Performance IQ

(Axelrod, Ryan, & Ward, 2001). The subtest scores can be prorated and the resulting scores can be used to develop estimates of Full Scale, Verbal, and Performance IQs. Alternatively, tables provided in Sattler (2001, p. 835) and J. Ryan (1999) can be used to develop estimated IQ scores. Performance and Full Scale IQ scores have been found to be nearly as reliable as for full-administration IQs with the average Full Scale standard error of measurement being 2.80 (and 2.72 for the version with Matrix Reasoning) versus 2.58 for the full WAIS-III Full Scale IQ (J. Ryan & Ward, 1999). Correlations between the J. Ryan and Ward (1999) seven-subtest short form and a full administration were .98 for the Full Scale IQ, .97 for the Verbal IQ, and .95 for the Performance IQ (.96 using Matrix Reasoning). Thus, the psychometric properties of the seven-subtest short form are excellent, and administration times are only marginally longer than for the Vocabulary-Arithmetic-Block Design-Picture Arrangement four-subtest short form.

The Satz-Mogel/Yudin Short Forms

An alternative to administering various combinations of subtests is to use every subtest but limit the number of items from each of the subtests. The most frequently used variation is the Satz and Mogel (1962) approach, which was originally developed for the WAIS but can also be used for the WAIS-III and WISC-III. The procedure is to administer every third item for Information and Vocabulary and multiply the scores by three to obtain the raw scores. Only odd items are administered for Similarities, Arithmetic, Comprehension, Block Design, Object Assembly, and Picture Completion, and each score is multiplied by two to obtain the respective scaled scores. Full administrations are given for Digit Span, Digit Symbol-Coding, Letter-Number Sequencing, Matrix Reasoning, and Symbol Search. The entire procedure takes approximately 40 minutes and the derived IQs have correlations similar to the best four-subtest variations. A distinct advantage over four-subtest variations is that the Satz-Mogel approach samples a wider range of areas. This is likely to increase the stability of scores over a wider variety of populations and allows clinicians to develop inferences over a larger number of behaviors. Research with the WAIS-III has indicated that IQs derived from the Satz-Mogel usually did not vary more than 6 points when compared with the full administration (J. Ryan, Lopez, & Werth, 1999). In addition, a full 86% of the clients had the same IQ classifications. A caution is that, although a score is provided for each subtest, it is inappropriate to attempt a profile analysis because the individual subtests are not sufficiently reliable (J. Ryan, Lopez, & Werth, 1999).

A WISC-R/WISC equivalent of the Satz-Mogel approach was developed by Yudin (1966) and has the same advantages and disadvantages and follows a nearly identical procedure. If adapted for the WISC-III, Digit Span, Mazes, and Symbol Search would not be administered because they are optional subtests; but Coding, like Digit Symbol-Coding on the WAIS-III, would be given in its entirety. However, if examiners did decide to use the Symbol Search subtest because of its good psychometric properties or clinical relevance, it would need to be given in its entirety.

Modified Format

A final approach is the elimination of early, easy items on each of the subtests. This is most appropriate for relatively bright subjects but should be used cautiously with persons of below-average intelligence. Cella (1984) has provided WAIS-R guidelines for the number of items to be omitted based on a subject's performance on the Information subtest. Research on the WAIS-III using this format has not yet been conducted. However, such a procedure with the WAIS-R has been found to have an almost exact correlation (.99) with a full administration and yet can reduce the total administration time by 25%. Despite this high correlation, some caution should be exercised toward Cella's Modified Format and the Satz-Mogel approaches. First, lowered internal consistency is likely to reduce subtest reliabilities sufficiently to render profile analysis questionable. Second, examinees are disadvantaged because they are not able to have as many previous subtest items to practice on (as items are skipped) before being administered more difficult items. The result may be that the norms for the full administration may not necessarily apply to the shortened versions.

RECOMMENDED READING

Gregory, R. J. (1999). *Foundations of intellectual assessment: The WAIS-III and other tests in clinical practice*. Boston: Allyn & Bacon.

Kaufman, A. S. (1994). *Intelligent testing with the WISC-III*. New York: Wiley.

Kaufman, A. S., & Lichtenberger, E. O. (1999). *Essentials of WAIS-III assessment*. New York: Wiley.

Kaufman, A. S., & Lichtenberger, E. O. (2000). *Essentials of WISC-III and WPPSI-R assessment*. New York: Wiley.

Kaufman, A. S., & Lichtenberger, E. O. (2002). *Assessing adolescent and adult intelligence* (2nd ed.). Boston: Allyn & Bacon.

Sattler, J. M. (2001). *Assessment of children: Cognitive applications* (4th ed.). San Diego, CA: Author.