Chapter 48

The Measurement of Giftedness

Linda Kreger Silverman

Abstract Instruments with the richest loadings on general intelligence (**g**) are the most useful for locating gifted children. Spearman's **g** represents **g**iftedness. Raven's Progressive Matrices, Stanford-Binet scales, and Wechsler scales are the most widely used IQ tests in selecting gifted children worldwide. All were founded on the conception of intelligence as abstract reasoning (**g**), but each may locate a different group of gifted children. If index scores vary significantly, Full Scale IQ scores should not be derived. The new General Ability Index (GAI) of the WISC-IV, based on only six subtests, is recommended by the NAGC Task Force for selection of students for gifted programs. The Verbal Comprehension Index and Perceptual Reasoning Index can also be used independently, as can the Verbal and Nonverbal IQ scores of the SB5. Ratio IQ methods and supplementary testing are recommended for locating exceptionally gifted children and for determining degree of acceleration needed.

Keywords IQ tests · General intelligence · Wechsler scales · Stanford-Binet scales · Raven's Progressive Matrices

Origins of the Measurement of Intelligence

The infinite variability of human beings is so remarkable that one would assume it had been one of the earliest topics of scientific interest. Yet, the exploration of individual differences began less than 140 years ago, in the latter part of the nineteenth century. The study of giftedness is imbedded historically and philosophically within the study of individual differences in intelligence. Sir Frances Galton (1869, 1883), "father" of the testing movement (Shouksmith, 1970), inaugurated this field of study. Considered a genius himself (Terman, 1917), Galton (1869) may have been among the first to use the term "gifted" to refer to individuals of higher intelligence. He was the first inquirer to furnish a comprehensive description of the traits of gifted children (Hildreth, 1966), as well as information about the origins and development of genius.

Until this time, it was commonly believed that everyone (except perhaps freaks of nature) had the same native endowment. Galton (1869) provided the first quantitative analysis of human intelligence in his book, *Heredity Genius.* He devised the use of percentiles for ranking individuals and demonstrated that there was enormous variability in the population. Employing the normal curve as a model, he showed that individuals at the extremes differ from each other to a greater extent than individuals near the average. In *Inquiries into Human Faculty and its Development*, Galton (1883) explored the feasibility of measuring mental capacities with discrimination tasks, such as discerning the sequence of a set of weights. This was followed by his creation of the first mental tests, which involved measures of sensory capacity. He set up the first mental testing center in 1884 at the International Health Exhibition in London and charged visitors a fee for measuring the acuity of their senses (DuBois, 1970). In this manner, he assessed nearly 10,000 individuals, ranging in age from 5 to 80 years. While Galton's

L.K. Silverman (✉)
Institute for Advanced Development, Denver, CO, USA
e-mail: gifted@gifteddevelopment.com

L.V. Shavinina (ed.), *International Handbook on Giftedness*,
DOI 10.1007/978-1-4020-6162-2_48,

experiments in testing failed, his influence was immense (Carroll, 1993) and a new field was born.

Galton's successors were Karl Pearson and James McKeen Cattell. Pearson refined Galton's concept of co-relation, creating the mathematical procedure for statistical correlation, essential in factor analysis and in determining test validity and reliability (DuBois, 1970). The Pearson product–moment formula for linear correlation is now in universal use, as well as his many other statistical procedures, including multiple correlation and the chi square test for goodness of fit. James McKeen Cattell, Galton's assistant in his Anthropometric Laboratory, was the first researcher in America to promote mental testing; he coined the term "mental test" (J. M. Cattell, 1890). Cattell set up the first testing laboratory in America. Building on Galton's methodology, he developed 50 tests of sensory capacity, discrimination and reaction time and attempted to use his measures to select superior individuals for responsible positions. As these simple measures failed to differentiate cognitive abilities, his endeavors, like his mentor's, did not prove fruitful (Cronbach, 1970). Galton was obsessed with the role of heredity in intelligence, but Cattell parted company with him on that issue; he maintained that environmental opportunity plays a vital role in the development of abilities (J. M. Cattell, 1915).

Spearman's Unsinkable g and Raven's Progressive Matrices

Theories of intelligence appeared at the turn of the century. Charles Spearman is often credited with having produced the first one in 1904. Like Galton, Spearman believed that every individual is endowed with a certain amount of mental energy and that it was possible to rank individuals according to intellectual power. He defined intelligence as the integrative capacity of the mind to understand one's own experience and extract relationships (Spearman, 1927). For Spearman, intelligence was not an elusive set of mysterious processes; it was a real entity. He set about the task of isolating intelligence from the contamination of other influences, such as learning, emotion, and temperament. He used the newly developed method of correlation in his attempt to ensnare this creature. Since measures of different abilities show high correlations, Spearman postulated that there must be a general function common to all abilities. He named his quarry the "general ability factor" or **g** factor, and the world heralded it as "the Spearman **g**."

Spearman believed that **g** manifested itself in varying degrees in a hierarchy of mental activities, with complicated mental activities containing the greatest amount. Tests of intelligence would have to contain large amounts of **g.** The two kinds of abilities Spearman (1923) most associated with **g** were the discovery of relationships between ideas ("eduction of relations") (p. 63) and the ability to see implications or new relations based on these relationships ("eduction of correlates") (p. 91). He considered tests of analytical reasoning ability ideal to capture the pure element of intelligence.

Raven's Progressive Matrices

Inspired by his work, two of Spearman's fellow Englishmen produced such a test in 1938. John C. Raven, a psychologist, and Lionel S. Penrose, a geneticist, invented the *Progressive Matrices Test* as an assessment of pure **g**. Little is written about J. C. Raven. In a letter to David Watt, Raven's son remarked regarding the *Progressive Matrices*, "The items have all the hallmarks of my father's personality. One sees the same attention to detail, the same concern with design, the same concern with aesthetics, and the same progression in thinking that one sees in his gardening" (Watt, 1998, p. 144).

Spearman considered the *Progressive Matrices* to be the best measure of **g**. "In keeping with Spearman's theoretical analysis of *g*, this test requires chiefly the eduction of relations among abstract items" (Anastasi, 1988, p. 302). R. L. Thorndike (1986) also saw Raven's *Progressive Matrices* as exemplifying a method of measuring **g** as purely as possible, uncontaminated by any other influences. And this instrument has withstood the test of time. In his detailed analysis of factor analytic studies of intelligence, John Carroll (1993) wrote "Our evidence suggests that the Progressive Matrix test is a good measure of g. . ." (p. 696). It also appears to measure cognition of figural relations, spatial ability, accuracy of discrimination, reasoning by analogy, logical relations, and inference. The test is constructed of figural analogies of

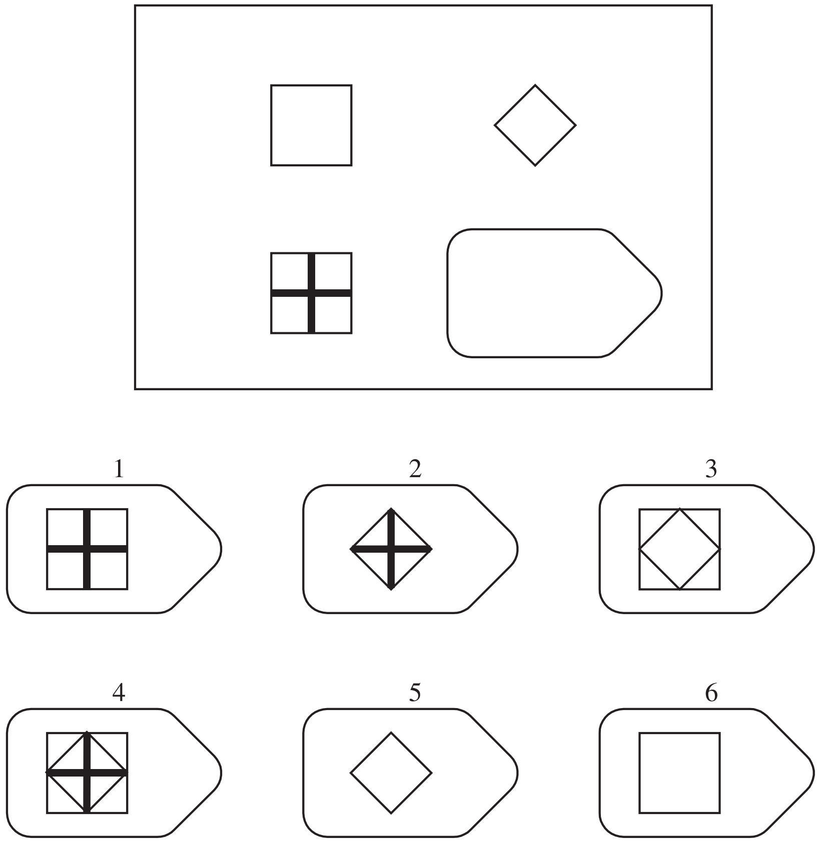

Fig. 48.1 Simulated item similar to those in the *Raven's Progressive Matrices – Standard Progressive Matrices* (Copyright 1998 by NCS Pearson, Inc. Reproduced with permission. All rights reserved)

varying degrees of complexity. Each matrix has a missing element in the lower right corner. The examinee must select the part that will complete the matrix from among six or eight alternative segments or symbols (Fig. 48.1).

The three forms of the *Progressive Matrices* have differing levels of difficulty: the 36-item *Coloured Progressive Matrices* (J. C. Raven, 1965) for children from 5 to 11 years; the 60-item *Standard Progressive Matrices* (J. C. Raven, 1958) for ages 6–17; and a 48-item *Advanced Progressive Matrices* (J. Raven, J. C. Raven, & Court, 1998) for older adolescents, adults, and the gifted. The *Standard Progressive Matrices* was standardized in Great Britain with 3,250 children aged 6 through 16 and is considered representative of all seven regions (J. Raven, 1981). Norms for American children were established in 1986 (J. Raven & Summers, 1986). Data have also been collected in other countries. Correlations tend to be higher with nonverbal than verbal tests (Anastasi, 1988). It is possible to solve the items through visual apprehension or analytically. The test can be self-administered by older students and a concrete version with moveable pieces on form boards is available for very young children and the disabled (J.C. Raven, 1965). A tactile form of the *Progressive Matrices* has shown promise for blind children between the ages of 9 and 15 (Rich & Anderson, 1965). The instructions can be pantomimed for the deaf and it has been adapted for orthopedically handicapped children, as it can be responded to with only a head nod (Anastasi, 1988). Therefore, the test is excellent for locating gifted children with severe disabilities, with the exception of those with visual processing disorders or dyslexia.

Today, *Raven's Progressive Matrices* is one of the most popular tools worldwide for assessing gifted children of diverse cultural backgrounds as it (1) is a nonverbal test, (2) is culturally reduced, (3) is simple to administer, (4) can be administered by teachers individually or in groups, (5) is inexpensive, (6) is untimed, and (7) is quick to administer (between 15 and 45 minutes). Its main drawback is its low ceiling. While item difficulty is excellent on the *Advanced Progressive Matrices*, scores range only as high as the 99th percentile (135 IQ) on all forms, so it cannot be used for discriminating at the highly gifted levels and above. The "intellectually superior" range is set at the 95th percentile. It has found renewed popularity through the international efforts of J. C. Raven's son, J. Raven.

With remarkable foresight, J. C. Raven recommended that his *Progressive Matrices* be given in conjunction with a vocabulary test, such as the *Mill Hill Vocabulary Scale*, (J. Raven, 1983a) and recent standardizations have been conducted with both scales (J. Raven, 1981). In his Introduction to *The Coloured Progressive Matrices,* J. C. Raven wrote "*By itself,* it is not a test of 'general intelligence', and it is always a mistake to describe it as such. For this purpose it should be used in conjunction with a vocabulary test" (p. 3). The combination of the *Progressive Matrices* and a vocabulary test has been used successfully to select culturally diverse gifted children in the United States and abroad. Vocabulary is a robust measure of ability and vocabulary subtests have high **g** loadings on most intelligence tests (Anastasi, 1988; Flanagan & Kaufman, 2004; Kaufman, 1994).

Nonverbal or spatial abilities are becoming increasingly recognized as an important part of intelligence. International research has shown that these abilities have improved over the last few generations (Flynn, 1984, 1999). "The Flynn effect was originally derived from research with tests that are good measures of general intellectual ability (Spearman's *g* factor), such as *Raven's Progressive Matrices*" (Wasserman, 2007a, p. 2). (More information about the Flynn effect can be found toward the end of the chapter.)

Multifactor Theories

As important as **g** was, it could not, however, account for all mental abilities. It was necessary to add another construct. Spearman (1927) solved the problem by devising a two-factor theory, which states that every intellectual activity consists of a general (**g**) factor, which is shared with all other intellectual activities, and a specific (**s**) factor, which is unique. The (**s**) factor represents that portion of each ability not correlated with other variables.

In his pursuit of **g**, Spearman laid the groundwork for the factor analytic method, which has lived a long and healthy life in the field of psychological testing. It provided the basis for the work of Thurstone (1938), Burt (1949), Vernon (1951), Guilford (1956), R. B. Cattell (1963), Horn (1988), and a host of other theorists and researchers (see Carroll, 1993). Sir Cyril Burt (1949) and Philip E. Vernon (1951) extended Spearman's two-factor model into more complex hierarchical models of intelligence. They accepted the basic principle of an overarching general factor, but focused their attention on categorizing the specific factors into groups. These models were the forerunners of the modern theories of intelligence that undergird current intelligence tests.

For the first few decades of the 20th century, Spearman's concept of a general ability factor dominated psychometric theory and practice (R. L. Thorndike 1986); then a backlash against **g** ensued in the 1930s. Some of Spearman's successors used his factor analytic techniques to demonstrate that intelligence could not be regarded as unitary. They espoused a multidimensional perspective of the nature of intelligence. The most vocal opponent of **g** was Louis Thurstone (1938), who constructed a theory of primary mental abilities with nine parallel factors. He was followed by a series of researchers and theorists who sought to extinguish **g**. The most ambitious undertaking was by Guilford (1956), who proposed a model of 120 intelligences; it mushroomed to 180 before he died (Guilford, 1988). In a similar vein, not employing factor analysis, Howard Gardner (1983) postulated seven intelligences, which continue to expand.

But **g** is a tenacious animal; it refuses to become extinct. For decades, American psychologists have vigorously sought to replace the concept with specific factors, but no matter how hard they try, they have been unsuccessful in eradicating **g**. The larger the number of items assessed, the more likely the items are to correlate with each other, giving evidence of an underlying construct that supports all mental abilities (R. L. Thorndike, 1986). Thorndike reported that among the 60 tests in Louis and Thelma Thurstone's battery of primary mental abilities, 97% of the correlations were positive—further demonstration of **g**. "Still, the general ability factor, **g**, refuses to die. Like a phoenix, it keeps rising from its ashes and will no doubt continue to be an enduring part of our psychometric theory and psychometric pratice" (R. L. Thorndike, 1986, p. 6). Carroll's (1993) popular three-stratum theory was an expansion of Spearman's, among others. He concluded, "There is abundant evidence for a factor of general intelligence, G. . ." (p. 624). The unsinkable **g,** along with myriad specific factors, continues to be the foundation of psychometric assessment.

The unitary model of intelligence was preserved in the work of Binet and his successors. Spearman discovered through factor analytic techniques that independent test items, which might be somewhat "untrustworthy" by themselves, yielded highly valid results when they were pooled together (DuBois, 1970). The concept was empirically discovered by Binet as well—without the aid of factor analysis—and it became the basis of the first intelligence scale. It remains central to test construction.

The Binet–Simon Scales

While Spearman was developing his theory in England, Alfred Binet was experimenting with various mental tests in France, and the results of their labors appeared within a year of each other. Spearman supplied a unitary model of intelligence, and Binet produced the first workable method by which it could be assessed. The *Binet–Simon* scale, released in 1905, was a major advancement in the assessment of intelligence because it was the first time items of a scale had been combined to yield a composite measure of a complex function (DuBois, 1970). Spearman's **g** was now quantifiable. In 1908, Binet and Simon presented a major revision of their scale, utilizing a new invention: mental age. They defined mental age as the age at which a given number of test items were passed by an average child. These instruments were the first age-based scales. Binet continued to improve the scale, changing the arrangement of some of the items, and provided one further revision in 1911, shortly before his untimely death at the age of 54.

Mental age served as an excellent basis for scoring, but there was no way to distinguish between normal and abnormal development on the basis of mental age alone. German psychologist, William Stern, found the solution. Stern (1910/1911) recommended that the mental age be divided by the chronological age in order to obtain a "mental quotient," which would be more constant than mental age alone. Lewis Terman (1916b) incorporated this concept into his adaptation of the Binet scale, renaming it the "intelligence quotient," and the IQ became a permanent part of testing vocabulary.

Although his ideas varied from time to time, essentially Binet (1909) viewed intelligence as judgment or reasoning ability, a perspective which was to be carried forward by Terman. Binet's name has become linked with the identification of children with limited mental abilities, but his work was particularly important for detecting giftedness in children. Like Galton, Binet had an intense interest in prodigies. He studied children with extraordinary talent in calculation, mathematics, chess, writing, and other areas (Hildreth, 1966).

Galton, Spearman, Binet, and Terman all shared the belief that there was a single unifying factor of intelligence, but for Binet, that factor could not be isolated from the rest of experience. He viewed intelligence as a rich, complex, multi-faceted gestalt—a myriad of dynamically interrelated abilities. Emotion and personality also played critical roles in his conception of intellectual ability. He believed that intelligence was highly influenced by the environment and that it could be *improved* through appropriate instruction. He viewed intelligence as a continuously evolving process, *not* as a static amount of raw material that stayed the same throughout life (Seagoe, 1975). Yet, intelligence testing is viewed today as a method of rigidly determining the limits of one's abilities—quite different from Binet's intent. In adopting the numerical criterion that his test provided, we lost sight of Binet's essential philosophy, which was his greater legacy.

The Stanford-Binet Scales

Binet gave Lewis Terman his blessing to create an American version of his scale, selling him the rights "for a token of one dollar" (Wasserman, 2003, p. 425). Alfred Binet was Terman's role model. He dedicated his book, *The Measurement of Intelligence* (1916a), to Binet. In 1910, Terman began the process of revising the *Binet–Simon* scale. Although there were other revisions and translations (e.g., Goddard, 1910), after six painstaking years, Terman was able to build a better mousetrap. "The revision introduced so many changes and additions as to represent virtually a new test. Over one-third of the items were new, and a number of old items were revised, reallocated to different age levels, or discarded" (Anastasi, 1988, p. 239). The scale was standardized on 1,000 children and 400 adults. Seagoe (1975) claimed that it was not mere modesty that prevented Terman from naming his revision the "Terman-Binet," but rather a profound appreciation for all of the students who had assisted him in gathering the data and a desire to share the credit.

The first edition received a more enthusiastic welcome than Terman dared to dream. He had not expected interest in intelligence testing to catch on so quickly. The *Stanford-Binet Intelligence Scale* (Terman, 1916b) became the major mental test of the next two decades. Within 15 years, it was translated into 20 languages, and between 15 and 20 million tests were given annually. The success of the test appears to have been due to Terman's skills in selecting and placing items that reflected the appropriate order of development of the abilities tested and to the internal consistency and cohesiveness of the test as a whole (Seagoe, 1975).

The 1937 revision (Terman & Merrill, 1937) was greatly expanded and consisted of two equivalent forms to assess reliability: Form L (for Lewis Terman) and Form M (for Maud Merrill). At that time, two alternate forms were necessary, as no other well-constructed IQ test was available (Anastasi, 1988). It was standardized on over 3,000 children, at least 100 at each age level, and efforts were made to improve the geographical and socioeconomic stratification of the sample. "The 1937 edition of the *Stanford-Binet* had remarkable breadth, developmentally appropriate procedures, and a highly varied administration pace so that examinees performed many different kinds of activities. It may have constituted an early high point for intelligence testing" (Wasserman, 2003, p. 425).

In 1960, 4 years after Terman's death, the *Stanford-Binet (Form L-M)* appeared, merging the best test items of the two 1937 forms based on the performance of approximately 4,500 individuals. The scale was not restandardized, but the new samples served to determine changes in item difficulty (Anastasi, 1988). The ratio method of computing IQ (dividing the mental age by the chronological age and multiplying by 100) was replaced by the inclusion of a series of norm tables provided by Samuel R. Pinneau. Three new levels were added to the test—Superior Adult I, II, and III—and the age for IQ calculation was extended to 18. New research had shown that mental age continues to increase beyond the age of 16, the point at which Terman had originally thought intelligence reached its peak.

The Stanford-Binet Intelligence Scale (Form L-M)

In 1972, a unique approach to standardization was used by R. L. Thorndike. He selected 2,100 participants from the 20,000 in the standardization sample of his group IQ test, the *Cognitive Abilities Test* (R. L. Thorndike & Hagen, 1971). His goal was to represent all ability levels. While no attempt was made to stratify the sample by socioeconomic status, geographic region, and ethnic group, "the sample ultimately was more inclusive and diverse than that used with any previous *Stanford-Binet* edition" (Wasserman, 2003, p. 425). African American, Mexican American, and Puerto Rican children were included (Anastasi, 1988). Scoring was based on deviation IQs, and the maximum possible score was reduced to 163 (or lower for some age levels). Only a couple minor changes were made to the content.

Although the Stanford-Binet received a good share of criticism over the years, it provided the most stable means of predicting school success and assessing giftedness. It dominated intelligence testing in the United States until the 1960s (Lubin, Wallis, & Paine, 1971). In an article describing the endurance of the Stanford-Binet, R. L. Thorndike (1975) wrote

> (It) has been for most of the past 60 years the workhorse of psychometric appraisal of cognitive development, the standard against which other tests of cognitive abilities have been evaluated, and more recently a prime target for the social critics of ability testing. (pp. 3–4)

The Stanford-Binet scale maintained international popularity, as it had been translated into many languages and adapted for use in England and Australia. While the United States continued to develop new IQ tests and new editions of tests, clinicians in other countries were not quick to retool each time a new IQ test hit the market. First, there was the question of cost. Psychologists in private practice had limited funds to spend on new instruments. Second, there was the lack of adaptations of these new instruments to the monetary system and culture of their countries. Third, each new assessment tool required a substantial learning curve before the examiner felt comfortable both administering the test and interpreting the results. Fourth, the Stanford-Binet had a wide age range, making it useful for the majority of individuals tested, whereas the purchase and mastery of three different Wechsler scales would have been required to replace it.

Researchers and psychologists who specialize in the exceptionally gifted still use the *Stanford-Binet (Form L-M)* as a **supplementary** test for many reasons: (1) it is the only IQ test that can discriminate children in the exceptionally and profoundly gifted range of intelligence; (2) the ratio-based scoring used to derive formula IQ scores beyond the norms in the manual allows a greater spread of scores; (3) its high ceiling makes it ideal for out-of-level testing (Stanley, 1990); (4) it is an excellent measure of abstract reasoning; (5) the complex verbal absurdities, proverbs and analogies are exciting for gifted children; (6) it tests the limits of children's knowledge, as they receive credit for all correct answers beyond the discontinue criterion; (7) it is not necessary to be advanced in every area; it is possible to proceed to the highest levels on the basis of strengths in verbal, spatial, or mathematical reasoning; (8) it has a strong research base, high predictive validity, and a long history of successful use with gifted populations; (9) as the test is virtually untimed, it provides a fairer assessment of children with visual-motor or processing speed issues; (10) its adaptive-testing format: examinees are given a range of tasks best suited to their ability level, which leads to better rapport (R. L. Thorndike et al., 1986); (11) examinees, particularly preschoolers, find the test more engaging because the items are short and more varied (Canter, 1990; Vernon, 1987); (12) the tests can be administered out of order in accordance with the engagement of the child; (13) the wide range of items (Bayley, 1970); (14) the rapid movement from one type of item to another prevents a gifted child from succeeding on an item due to practice effects; (15) the extensive age range (R. L. Thorndike et al., 1986); (16) its emphasis on abstract verbal abilities enhances identification of African American children (Kearney & LeBlanc, 1993); and (17) it is an excellent clinical tool in the hands of a skillful examiner as it respects clinical judgment.

The Stanford-Binet Intelligence Scale, Fourth Edition (SB-IV)

In the 1986 revision, the Stanford-Binet scale took on an entirely new personality. Point scales were substituted for the developmental age scale format that characterized all of its predecessors. Instead of the global score yielded previously, four broad areas of cognitive ability were assessed: Verbal Reasoning, Abstract/Visual Reasoning, Quantitative Reasoning, and Short-Term Memory. The overall IQ score was renamed the Standard Age Score. The Vocabulary test, which had informally been used as a routing test in prior editions, became codified as the method for determining the starting point for each test. Wasserman (2003) indicates that the fourth edition was innovative in many ways: the first to use item response theory (IRT) (Rasch, 1960) to determine starting and stopping points; the first to use "differential item functioning to minimize item bias" (p. 425); and "the first contemporary test to operationalize Horn and Cattell's (1966) fluid and crystallized model of intelligence" (p. 426). It offered flexibility in selecting which subtests to administer. Much easier to administer and score than earlier versions, the scale employed an easel-based format. Composite scores correlated highly with Wechsler scales. The standardization sample consisted of over 5,000 individuals, from age 2 through 23, tested in nearly every state. The sample was stratified for geographic region, community size, ethnic group, gender, and socioeconomic status (R. L. Thorndike et al., 1986).

Despite all of these improvements, the *Stanford-Binet Intelligence Scale, Fourth Edition* (SB-IV) was plagued with difficulties and not warmly received. The *Technical Manual* and test norms were delayed. The authors could not agree about the factor-scoring procedures, so several different methods were generated (Wasserman, 2003). The test had a higher floor and lower ceiling, making it less usable at both extremes (Wilson, 1992). Whereas Terman had designed his original scale with a very high ceiling in the hope of locating future geniuses (Terman, 1925), it was clear that the gifted were not a top priority in designing this revision. Originally, the scale was to end at 148—the 3rd standard deviation above the norm. Upon hearing that the new test had a ceiling of 148, the author made personal calls to all three test constructors and each one explained that there were not enough highly gifted children in the normative sample to warrant going beyond 148. The reason for this was given by Elizabeth Hagen:

> In constructing a cognitive abilities test you are always faced with constraints. You have to produce an instrument that will adequately appraise the full range of individual differences in a chronological age group from the very slowest level of development to the most rapid. At the

> same time, you have to produce an instrument that can be administered fairly easily and within a reasonable amount of time. The compromise is to produce an instrument that is most effective in the range of 4 s.d.'s; therefore you can't use tasks that are successfully completed by 99.99 percent of an age group or that are failed by 99.99 percent of an age group. In the construction of the Binet [Revision IV], I was working with some nonverbal items that could only be solved by children who were in classes for the gifted. You can't put items like that in an intelligence test because they aren't functional for a wide enough group. (Hagen, interviewed in Silverman, 1986, p. 171)

Higher scores could only be generated by means of extrapolation. Robert Thorndike agreed to extrapolate to the fourth standard deviation—164 (R. L. Thorndike, personal communication, October 1985), but it was a destination in theory only, as it was virtually unattainable.

Two samples of children who had scored in the gifted range on other instruments were reported in the *Technical Manual*; their composite scores were, respectively, 121.8 and 116.3. The first group had attained a mean score on Form L-M of 135.3 and the second group had a mean WISC-R Full Scale IQ score of 117.7. Strangely, the gifted or very superior range was still set at 132 (R. L. Thorndike et al., 1986). Psychologists in the United States who were using the WISC-R for placement decisions for gifted programs saw no advantage of switching to the SB-IV. Those who were using Form L-M to locate extremely advanced children could not find them on the SB-IV.

Perhaps the most vocal critic of the SB-IV was Philip Vernon (1987), a contemporary of Lewis Terman, and a leading figure in psychological assessment. In "The Demise of the Stanford-Binet Scale," Vernon wrote, "It is misleading to call the battery a Stanford-Binet scale. Indeed it marks the end of Terman's 70-year era" (p. 257). He protested that "The L-M provides a richer sampling of verbal abilities. In the [Binet IV] there are no rhymes, analogies, sentence completions, ideational fluency, interpreting proverbs, comprehension of paragraphs, or verbal reasoning problems" (p. 255).

The Stanford-Binet Intelligence Scale, Fifth Edition (SB5)

The newest revision of the Stanford-Binet appeared in February 2003, and, once again, a whole new animal was created. The SB5 measures five factors: Fluid Reasoning, Knowledge, Quantitative Reasoning, Visual-Spatial Processing, and Working Memory. Each factor has a verbal and nonverbal scale. It generates three IQ scores: Verbal, Nonverbal, and Full Scale. It is not necessary to calculate a Full Scale IQ; it is permissible to use either the Verbal IQ or the Nonverbal IQ independently. The SB5 assesses the widest age range of all the Stanford-Binet scales, from 2 to 85. To accommodate this range, there are six levels of the test. The battery employs Item Response Theory as pioneered by Rasch (1960) and incorporates items from all previous editions (Roid, 2003a). Each item is arranged by level of difficulty. Like the Wechsler scales, the mean IQ is 100, the standard deviation is 15 (on all prior editions, it was 16), and the scaled scores on the subtests range from 1 to 19.

Rather than completing items that are all alike, as in the Wechsler scales, the examinee encounters a variety of different testlets measuring each of the five factors. The scale was normed on 4,800 individuals ranging in age from 2 through 99. An additional 3,000 individuals participated in a variety of special studies (Carson & Roid, 2004). Scores range from 40 to 160 in the standard range. Extended IQ scores range as low as 10 and as high as 225 (Roid, 2003b).

The flexibility of being able to administer only the Nonverbal scale makes the SB5 useful in measuring the abilities of non-English speaking children. Its emphasis on mathematical and visual-spatial reasoning and its liberal time limits make it appealing for locating visual-spatial learners (Silverman, 2002). As 20% of the test measures mathematical reasoning, while only 10% measures abstract verbal reasoning (Wasserman, 2006), the battery is an enormous departure from earlier versions, which were criticized for emphasizing verbal abilities. Children who like mathematics and science tend to score fairly well on this instrument, but the scale appears to identify a different group as gifted.

The driving force behind this instrument is Cattell–Horn–Carroll (CHC) theory. The field has increasingly put pressure on test constructors to have a theoretical basis for the scales of intelligence they create. The latest, greatest model was devised during the development of the third edition of the *Woodcock-Johnson Cognitive Battery* (WJ-III) and the fifth edition of the *Stanford-Binet* (SB5). It was based on the monumental work of John B. Carroll (1993), who created what McGrew (2007) has called a periodic table of psychological elements. After meticulously analyzing over 460 archival data sets, Carroll (1993) confirmed that the hierarchical theories of fluid and crystallized abilities espoused by R. B. Cattell (1971) and Horn (1988) were empirically sound. He expanded upon Cattell and Horn's theories by creating a three-tiered model, with **g** at the apex, followed by eight broad ability factors and numerous second-order factors. The two models were complementary (with the main exception that John Horn refused to accept the existence of **g**). Test constructors have integrated the work of Cattell, Horn, and Carroll as the basis for several IQ tests (e.g., WJ-III, SB5, *Differential Abilities Scales* [DAS and DAS2], *Kaufman Assessment Battery for Children* [KABC-2]). A series of meetings with Richard Woodcock, Kevin McGrew, John Horn, and John Carroll resulted in the Cattell–Horn–Carroll (CHC) theory (McGrew, 1997), which has become the most popular theory today in test construction.

Historically, attempts to break intelligence down into narrow, discrete abilities date back to Binet and Simon (1905), who concluded that individual "faculties" could not be purely and efficiently measured. Independent investigations in the United States (Sharp, 1899) also challenged support for the differentiation of mental faculties, and Binet decided to use complex tasks that might be influenced by several mental faculties at once. Even as Thurstone was introducing his nine distinct primary abilities, E. L. Thorndike (1941) was urging caution:

> Mental abilities are not an orderly retinue of a few easily defined and unitary faculties or powers, somewhat like the chemical elements, for each of which a mental meter or test can be found by sufficient labor and ingenuity. A mental ability is a probability that certain situations will evoke certain responses, that certain tasks can be achieved, that certain mental products can be produced by the possessor of the ability. It is defined by the situations, responses, products, and tasks, not by some inner essence. (p. 504)

Lewis Terman (1921) defined intelligence simply as abstract reasoning ability: "An individual is intelligent in proportion as he is able to carry on abstract thinking" (p. 128). Observing that those with strong abstract reasoning (**g**) were more successful in school and work, and those with weak abstract reasoning were less successful, he created a scale that measured as much abstract reasoning as practicable. The original scale had excellent predictive validity (R. L. Thorndike, 1975). According to a survey by Snyderman and Rothman (1988), "abstract thinking or reasoning" ranked first place as a component of intelligence by 661 "scholars with any expertise in the area of intelligence and intelligence testing" (p. 250); there was **99.3%** agreement as to its importance. The authors comment, "In many ways, Terman's 1921 definition of intelligence as abstract thinking remains at the heart of current thought about intelligence" (p. 57).

These two diverse viewpoints create a quandary in the assessment of gifted children. Is giftedness essentially the amount of general intelligence (**g**) an individual possesses or is it an aggregate of specific abilities? Perhaps it is both, depending on one's perspective. Gifted children have high abstract reasoning (Silverman, 1993a). Traditionally, the intelligence tests used to select gifted children have been designed as good measures of abstract reasoning. IQ tests based on multifactored models define different children as gifted and those with high abstract reasoning ability may be missed. If we are comfortable that the children defined as gifted throughout the last century are the ones we want to continue to be able to identify, then tests based on **g** are the most useful. The main value of CHC theory for assessing the gifted may be its demonstration of an overarching **g** factor.

The test constructors of the SB5 created a careful balance of items in accordance with CHC theory, but is the battery effective in assessing gifted children? The answer depends on many variables: if the program is seeking children with strong visual-spatial and mathematical abilities rather than children with strong verbal reasoning abilities; if the cut-off score for admission is set at 120; and if Rasch scoring methods are used. When the Full Scale IQ is derived from the norm tables, scores are considerably lower for gifted children than on other instruments or on prior versions of the Stanford-Binet. Using the traditional cut-off score of 130, traditionally used for admission to many gifted programs in the United States, approximately one third of the students who qualified on other instruments would not qualify on the SB5 (Lovecky, Kearney, Falk, & Gilman, 2005). The mean IQ score of 202 children in the gifted validation sample of the SB5 was 124 (Roid, 2003a). The mean IQ score of a group of 25 profoundly gifted children, whose IQ scores ranged from 170 to 235+ on the SBL-M, was 130 on the SB5 (K. Kearney, personal communication, August 1, 2003). The highest score recorded in the standardization of the SB5 was 148 (Roid, 2003a).

To remedy this situation, it is recommended that the threshold for admission to gifted programs be lowered to 120 (Ruf, 2003). In addition, researchers have been experimenting with several other methods of scoring. Carroll (1993) found that tasks involving induction, reasoning, visualization, and language comprehension were most likely to be correlated with **g**, and that measures of speed of information processing and capacity of working memory were also likely to be correlated, "though at a low level" (p. 624). Therefore, it seemed reasonable to remove Working Memory from the Gifted Composite score; two alternate methods have been provided (Roid, 2003b; Roid & Carson, 2003). There is also a Nonverbal Gifted Composite Score, in which Working Memory is eliminated. Another method involves generating Rasch-Ratio scores, which are closer to the scores of the SBL-M, albeit somewhat higher in some cases (Carson & Roid, 2004; Gilman & Kearney, 2004; Lovecky et al., 2005). Some of the scoring modifications for the gifted originated with examiners of gifted children, Kathi Kearney, Deborah Ruf, and Deirdre Lovecky, who were involved in validation studies of the SB5 (Lovecky et al., 2005). Scoring modifications for the gifted can be found in several documents (see Carson & Roid, 2004; Roid & Barram, 2004; Roid & Carson, 2003; Ruf, 2003).

A table of Extended IQ scores is available in the Interpretive Manual for those who score beyond 150 on the Full Scale IQ score (Roid, 2003b). Based on Rasch scoring, the examinee is credited with all raw score points beyond the requirement to obtain the ceiling score of 19. To date, few have qualified for the Extended IQ score, but when Rasch-Ratio scores are derived for students who score 120 and above on the SB5, it holds great promise for identifying highly, exceptionally and profoundly gifted students.

At the time of this writing, apparently it is permissible to use the SBL-M as a supplemental test, as long as examiners acknowledge that the scores are on a different metric and, therefore, not comparable to deviation IQs (Carson & Roid, 2004). The publisher prefers that the SBL-M be co-administered with the SB5 so that three types of composite scores might be contrasted: standard scores, Rasch-Ratio scores, and SBL-M scores. As these scores are often radically discrepant, it may be difficult for professionals to determine which of these scores to use in selecting students for services. One option is to select students for gifted programs who attain a score of 120 or above on the SB5 and to qualify students for services for the highly and exceptionally gifted or determine the degree of acceleration needed on the basis of Rasch-Ratio scores. Then only one instrument is needed, with different scoring methods for different purposes.

The Wechsler Intelligence Scales

Just as the success and longevity of the Stanford-Binet scales are attributable to Lewis Terman, the phenomenal success of the Wechsler tests is due to the sagacity of their creator, David Wechsler. A gifted clinician and test constructor (Wasserman, 2003), Wechsler was skillful in determining which tasks would be most effective in probing the depths of human intelligence. For the last half century, the name Wechsler has become synonymous with intelligence testing. The Wechsler scales won over the American market in the 1960s (Lubin et al., 1971), and today, translated into 25 languages, they are the most widely used individual IQ test for identifying gifted children internationally.

Wechsler's introduction to testing was at Columbia University, where he was instructed by James McKeen Cattell and Edward L. Thorndike, among others. His passion was kindled during World War I. In his tribute to Wechsler, Matarazzo (1981) described the critical events during the war that shaped the man's future. While awaiting induction to the army, Wechsler worked with E. G. Boring at an army camp in Long Island, administering and scoring thousands of Army Alpha intelligence tests, a group-administered offshoot of the Stanford-Binet. He was later assigned to a post in Ft. Logan, Texas, where his major responsibility was individually assessing recruits who could not pass the written Army Alpha and Army Beta tests. The US government was concerned that too many immigrants and men of different ethnic backgrounds failed to qualify for the armed forces. Wechsler noted that despite the very low scores these men had received on the written tests, most of them functioned adequately in their civilian lives. He used the Stanford-Binet and nonverbal tests to gain a better indication of their abilities and found that the nonverbal tests reflected their adaptive capabilities much more accurately than the verbal tests.

Later, he sought clinical training with many notables, such as Henri Pieron and Louis Lapicque in Paris, and Anna Freud in Vienna.

Wechsler's conceptions became crystallized through his associations with both Spearman and Pearson when he was assigned to the University of London for 3 months as an army student. He had profound respect for Spearman's contributions. Later in life, when he saw that Spearman's unitary conception of intelligence had been largely abandoned by psychologists, he wrote, "Spearman's demonstration of the existence of at least one pervasive factor in all performances requiring intellectual ability remains one of the great discoveries of psychology" (Wechsler, 1958, p. 9). His exposure to the innovative correlational methods of Pearson provided the background he was to need for his own test construction (Matarazzo, 1981). These experiences gave Wechsler valuable insights into the nature and assessment of intelligence.

As much as he appreciated Spearman, Wechsler found Spearman's concept of general intelligence inadequate to encompass the non-intellective factors of intelligence, such as emotion and motivation. Wechsler had a grander vision: he was determined to formulate a more comprehensive picture of intelligence. In addition to the ability to reason, deal with symbols, and conceptualize abstractly, Wechsler believed that intelligence also involved individuals' abilities to deal effectively with their environment, which was related to motivation and emotional commitment. For Wechsler, like Binet, intelligence was inseparable from personality factors. He carefully differentiated between testing and assessment; he felt that testing could never completely capture the full range of a person's intellectual capacity because it overlooked the personality factors, whereas clinical assessment took these factors into account.

Wechsler's definition of intelligence serves as the foundation of his scales:

> *Intelligence is the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment*. It is global because it characterizes the individual's behavior as a whole; it is an aggregate because it is composed of elements or abilities which, though not entirely independent, are qualitatively differentiable. (Wechsler, 1939, p. 3)

The definition seems to bypass the issue of whether intelligence is a single factor or a group of factors; it appears to be both. He called it an aggregate of qualitatively different elements functioning as an integrated whole. He went on to say that all that can be asked of an intelligence scale is that it measure sufficient portions of intelligence to offer a "fairly reliable index of the individual's global capacity" (Wechsler, 1958, p. 15). Here, Wechsler reveals his affinity with **g**, in the tradition of Spearman, Binet, and Terman.

In 1939, while chief psychologist at Bellevue Psychiatric Hospital, Wechsler created the *Wechsler-Bellevue Intelligence Scale*. The scale consisted of 11 subtests that were modified from currently available tests. It covered a broader spectrum of capabilities than had been tapped by the Stanford-Binet. Wechsler felt that the Binet scale was overburdened with verbal items, so he added a "Performance" dimension, which included a number of nonverbal items. His interest in measuring intelligence in adults required a new approach, since the content of tests for children were of little interest to adults. Further, developmental increments useful for assessing children would not apply to adults. Mental age was replaced by a comparison of each person's mental abilities with the norm for his or her age, thereby initiating the deviation IQ into intelligence testing. He also worked with a clinical population and sought to develop an instrument that could be used to assist in the diagnosis of mental disorders. His clinical interpretations of test scores to suggest different types of mental illness have not withstood the test of time.

Ten years later, in 1949, a downward extension of the original Wechsler-Bellevue (Anastasi, 1988) resulted in the *Wechsler Intelligence Scale for Children* (WISC). In 1955, the Wechsler-Bellevue was revised and renamed the *Wechsler Adult Intelligence Scale* (WAIS). Wechsler published the *Wechsler Preschool and Primary Scale of Intelligence* (WPPSI) in 1967. In 1974, he introduced a revision of the children's scale, the *Wechsler Intelligence Scale for Children-Revised* (WISC-R). The revision of the adult scale, the *Wechsler Adult Intelligence Scale-Revised* (WAIS-R) was published in 1981, the year of Wechsler's death.

Each of Wechsler's tests was extremely popular. All followed a similar format, giving near-equal emphasis to verbal and nonverbal components. Each test yielded three IQ scores: Verbal, Performance, and Full Scale. Unlike the Binet scales, the WISC could be broken down into subtests to obtain a profile of specific abilities. The Verbal subtests included Information, Comprehension, Arithmetic, Digit Span, Similarities and

Vocabulary. The Performance subtests included Picture Arrangement, Picture Completion, Block Design, Digit Symbol (Coding), and Object Assembly. Some of these items were borrowed from the Stanford-Binet scales. These subtests formed clusters or factors that could be analyzed to diagnose relative strengths and weaknesses (e.g., Bannatyne, 1974).

In addition, the Wechsler scales were easier to administer and score than the Stanford-Binet. The learning curve was dramatically shortened. Whereas it took months of practice to become fluid at administering the Stanford-Binet, the WISC could be learned very quickly. Nearly all clinical judgment was removed. The Wechsler scales exacted uniformity in administration and scoring to assure that the results of all tests, regardless of examiner, were comparable. The directions and scoring criteria were so clear and detailed that there was little margin of error. Ironically, Wechsler never gave his own tests the same way twice! He felt that the standardization was "imposed on him by the test publisher" (Wasserman, 2003, p. 430).

Another major difference between the WISC-R and the Stanford-Binet was the standardization sample. Rigorous standardization has always been a hallmark of the Wechsler scales (Wasserman, 2003). The 1937 Stanford-Binet had been standardized entirely on a white middle-class population and the 1960 version had not been restandardized. The WISC-R was standardized on a group of 2,200 children, including African Americans, Puerto Ricans, Mexican Americans, American Indians, and Asians, in approximately their proportions in the American population. This made the test more applicable to diverse cultures. The standardizations of the WISC-III and WISC-IV were even more scrupulous.

Clinicians appreciated the measures of nonverbal reasoning and became enamored with the Verbal and Performance discrepancies, a value not shared by their originator. To Wechsler, Verbal and Performance did not represent two dimensions of intelligence—they were just different ways of measuring **g**. "The subtests are different measures of intelligence, not measures of different kinds of intelligence, and the dichotomy into Verbal and Performance areas is only one of several ways in which the tests could be grouped" (Wechsler, 1958, p. 64).

In addition, Wechsler placed little value on the multifactor interpretation of his instruments. He was certain that general intelligence (**g**) accounted for much more of the variance than verbal, spatial, memory, speed, and other factors.

> Wechsler also rejected the separation of abilities because he saw intelligence as resulting from the collective integration and connectivity of separate neural functions. He believed that intelligence would never be localized in the brain and observed, "While intellectual abilities can be shown to contain several independent factors, intelligence cannot be so broken up"... (Wasserman, 2003, pp. 428–429)
>
> The Wechsler scales are the most widely used intelligence tests for identification of intellectually gifted ... its principal value is still based upon its measurement of the general factor **g**. (p. 430)

Unfortunately, after his death, the publisher veered from Wechsler's basic philosophy, and the Wechsler scales moved inexorably toward multifactor models (Wasserman, 2003). In the latest version of the WISC, even the familiar Verbal and Performance sections have been dispensed with in favor of four factors, but the degree to which they measure general intelligence (**g**) varies. And an IQ test or a portion of such a test is only useful for the assessment of giftedness to the extent that it measures **g**.

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

It was necessary to describe in some detail the last three versions of the Stanford-Binet scales, because they are all based on different theoretical models, measure dissimilar components, and are all in current use in various parts of the world. While earlier editions of the Wechsler scales for children are still in use throughout the globe, differences between the WISC, the WISC-R, and the WISC-III are relatively minor. Structurally, they are very much alike; however, the WISC-IV has been substantially altered. The WISC-III is still in use in many countries where it has been translated. Therefore, the next two sections will focus on these two editions.

The third edition of the *Wechsler Intelligence Scale for Children* (WISC-III) was released in 1991. It consists of six Verbal subtests and seven Performance subtests. In school settings, only the 10 required subtests are usually administered, but in clinical settings, all 13 of the subtests have been found valuable. The WISC-III introduced the concept of indices, paving the road for their acceptance in lieu of Verbal and Performance scores in the WISC-IV. The Verbal Comprehension Index consists of Information, Similarities, Vocabulary, and Comprehension. The Perceptual Organization Index is composed of Picture Completion, Picture Arrangement, Block Design, and Object Assembly. The Freedom from Distractibility Index includes Arithmetic and Digit Span, an optional test. The Processing Speed Index is made up of Coding and Symbol Search, an optional test. One optional subtest, Mazes, does not appear in any of the indices, foretelling of its demise in the WISC-IV. The latter two indices can only be calculated when two of the optional subtests are given. The Full Scale IQ score—a combination of the Verbal and Performance IQ—is weighted toward the higher of the two.

The WISC-III was designed for children between the ages of 6 and 16 years. Children as young as 3 take the *Wechsler Preschool and Primary Scale of Intelligence* (WPPSI) and youth 16 or older take the adult version, the WAIS. Even though, technically, 16-year-olds can be tested on the WISC-III, their scores are likely to be severely depressed due to the fact that they are hitting the ceiling of the test and because of the excessive bonus points for speed at the higher age levels (Kaufman, 1993). Like most other instruments, the mean is 100, and the standard deviation is 15. Scores range from 40 to 160, although it is quite unusual to obtain an IQ score above 150 on this scale. The Verbal subtests are administered orally and are untimed, with the exception of part of the Arithmetic subtest. The Performance subtests are heavily timed—a major limitation of this edition. The examiner alternates between administering a Performance subtest and a Verbal subtest.

Clinicians usually administer Digit Span, as it is an important diagnostic tool for assessing working memory. Symbol Search, introduced in the WISC-III, is an assessment of visual discrimination, tracking and processing speed, with minimal reliance on motor skills. All the child needs to do is make a slash mark. It is diagnostically useful for separating mental processing speed from visual-motor processing speed; it is often associated with reading ability. Both Digit Span and Symbol Search became required subtests on the WISC-IV. Mazes is the weakest test psychometrically, but it yields useful information about visual-motor planning and impulsivity. It is also an enjoyable way to end the test with gifted children, particularly those with high visual-spatial abilities.

The scaled scores range from 1 to 19 on each subtest, with a mean of 10, and a standard deviation of 3 points. Scores from 8 to 11 are in the average range; 12 and 13 are high average; 14 and 15 superior; 16, 17 and 18, gifted; and 19, highly gifted. Scores of 6 and 7 are low average; 4 and 5 borderline; and 1, 2, and 3 disabling. The following chart compares IQ scores, scaled scores, percentiles and ranges (Table 48.1).

The average range is quite broad, as low as the 9th percentile. This makes it difficult to identify children who are twice exceptional (both gifted and learning disabled). To be highly gifted, one must attain scores at the ceiling of the test, in the top one percentile, corresponding to the designation of disabled, which is the bottom one percentile. While there are exceptionally gifted children (160 -174 IQ) and profoundly gifted children (175+), there are no scores on the WISC-III to reflect these higher ranges.

Gifted children attain high scores on the subtests most richly loaded on general intelligence (**g**). The chart on p. 960 places the subtests of the WISC-III in order of **g**-loadings (Kaufman, 1994). The first five subtests listed are the best measures of giftedness (Table 48.2).

The **g-**loadings reveal why the Verbal Comprehension Index score, composed of three of the "good" measures (Vocabulary, Information, and Similarities),

Table 48.1 Relation of IQ scores and scaled scores to percentile ranks and ranges (Wechsler, 1974, p. 25, adapted)

IQ score	Scaled score	Percentile rank	Range
145	19	99.9	Highly gifted (ceiling)
140	18	99.6	Gifted
135	17	99	Gifted
130	16	98	Gifted
125	15	95	Superior
120	14	91	Superior
115	13	84	High average
110	12	75	High average
105	11	63	Average
100	10	50	Average
95	9	37	Average
90	8	25	Average
85	7	16	Low average
80	6	9	Low average
75	5	5	Borderline
70	4	2	Borderline
65	3	1	Disabling
60	2	0.4	Disabling
55	1	0.1	Disabling

Table 48.2 g-loadings on WISC-III (Kaufman, 1994, p. 43)

Good measures of g	
Vocabulary	(0.80)
Information	(0.78)
Similarities	(0.76)
Arithmetic	(0.76)
Block Design	(0.71)
Fair measures of g	
Comprehension	(0.68)
Object Assembly	(0.61)
Picture Completion	(0.60)
Symbol Search	(0.56)
Picture Arrangement	(0.53)
Poor measures of g	
Digit Span	(0.47)
Coding	(0.41)
Mazes	(0.30)

and one that is close, Comprehension, would be the best predictor of success in a gifted program. The combination of high scores in Arithmetic and Block Design usually indicates mathematical talent (Silverman, 2001). High Block Design and Object Assembly scores, combined with low Arithmetic and Digit Span, often signifies a visual-spatial learning style with auditory-sequential weaknesses (like the brilliant physicist who cannot calculate) (Silverman, 2002).

The greatest weakness of the WISC-III is its increased emphasis on bonus points for speed (Kaufman, 1992, 1993). Alan Kaufman, who assisted Wechsler in the renorming of the WISC-R, criticized both the WPPSI-R and the WISC-III for this reason.

> The biggest negatives for gifted assessment are the new emphasis on problem-solving speed on the WPPSI-R; the substantially increased stress on performance time in the WISC-III compared to the WISC-R; and the low stability coefficients for a majority of WPPSI-R and WISC-III subtests. (Kaufman, 1992, p. 158)

> The impact of problem-solving speed on a person's WISC-III IQ is substantial. . . . Children with reflective cognitive styles will be penalized on highly speeded items, as will children with coordination difficulties. Gifted children may score well below the cutoff needed to qualify for an enrichment program when the WISC-III is administered if they tend to be reflective or have even a mild coordination problem. Similarly, learning-disabled children may fail to score in the average range, even if they have normal intelligence, because of the speed factor. (Kaufman, 1993, p. 350)

The problems created by bonus points for speed increase with the age of the child. If an 8-year-old gets every single item correct, but fails to earn any bonus points for speed, it is only possible to obtain a score of 14 on Picture Arrangement (PA), 14 on Object Assembly (OA), and 16 on Block Design (BD). At the age of 12, the *highest possible scores without bonus points* are 8 on PA, 9 on OA, and 9 on BD—all below the 50th percentile. At the age of 16, the highest possible scores are 6 on PA, 7 on OA, and 7 on BD (Kaufman, 1993).

A study conducted at the Gifted Development Center compared IQ scores of 20 children on the WISC-III and the *Stanford-Binet Intelligence Scale* (*Form L-M*). The children were in the highest IQ ranges, obtaining scores from 151 to 191 on the SBL-M, with a mean of 173. On the WISC-III, they scored between 116 and 150, with the following means: Verbal IQ 141, Performance IQ 120, and Full Scale IQ 134. Only three of the children scored in the highly gifted range on the WISC-III, attaining Full Scale IQ scores of 146, 148, and 150—lower than the lowest score on the SBL-M. Discrepancies between the two IQ tests ranged from 14 to 60 points (mean difference 37 IQ points) (Silverman, 1995). The mean Verbal IQ was well within the gifted range. The excessively timed mean Performance IQ was 21 points lower and depressed the mean Full Scale IQ.

When using the WISC-III for the identification of English speaking gifted children, it is recommended that the Verbal IQ or the Verbal Comprehension Index be used as the criterion, rather than the Full Scale IQ. If the child's Performance IQ is higher than the Verbal IQ, which often occurs with students for whom English is not their primary language, the Performance IQ should be used for identification, as the Verbal IQ is strongly culturally loaded.

Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV)

The fourth edition of the Wechsler scales was released in August 2003. It is quite a departure from the WISC-III and all prior editions. The traditional Verbal and Performance scores have vanished. Four indices now take center stage: Verbal Comprehension, Perceptual Reasoning, Working Memory, and Processing Speed. The Verbal Comprehension Index and the Perceptual Reasoning Index can be thought of as rough equivalents to Verbal and Performance IQ scores (Wechsler, 2003). The Perceptual Reasoning Index is superior to the Performance IQ as a measure of abstract visual reasoning and, therefore, likely to be a better predictor of success in a gifted program—particularly for children from diverse cultures, bilingual children, twice exceptional children, and visual-spatial learners. Verbal Comprehension and Perceptual Reasoning are emerging as the most important factors to weigh in the selection to gifted programs.

The greatest changes have occurred in the nonverbal section of the test. The former Perceptual Organization Index, composed of Picture Completion, Picture Arrangement, Block Design, and Object Assembly, has been completely overhauled. Of the four subtests, only Block Design remains as a core subtest. Picture Completion is now an optional test, not administered very often, due to time constraints. Picture Arrangement and Object Assembly, along with Mazes—good measures of visual processing deficits—have all disappeared. They have been replaced by new subtests: Picture Concepts, a visual version of the Similarities subtest, and Matrix Reasoning, which measures spatial perception. The triad of Block Design, Matrix Reasoning, and Picture Concepts enables the WISC-IV to provide a better assessment of visual-spatial abilities than the WISC-III, particularly since the WISC-IV has generous time limits. More weight should be given to performance on Block Design and Matrix Reasoning, as they have much higher g-loadings than Picture Concepts. Visual-spatial learners may respond to Picture Concepts in terms of visual associations (e.g., "I saw the plant on the table" or "they're yellow"), rather than with the expected categorical reasoning.

The WISC-IV offers greater flexibility than the WISC-III, as it allows the substitution of two optional tests in calculating the Full Scale IQ. There are 10 core subtests and five optional tests. Most examiners administer 10 subtests, since it takes too long to give the entire battery. As 30% of the test measures verbal abilities, the scale remains a strong test of abstract verbal reasoning. It is now a stronger measure of nonverbal reasoning, as well. There are higher ceilings on some subtest items. And it is an excellent diagnostic tool for assessing strengths and weaknesses in twice exceptional children. Its major flaw is the increased weight of processing speed and short-term memory in the IQ score, accounting for 40% of the Full Scale IQ; this is double their weight on the WISC-III. However, there is an easy solution to the problem: the *General Ability Index* (GAI).

The GAI is an alternative to the Full Scale IQ for children whose Processing Speed and Working Memory indices are considerably lower than their Verbal Comprehension and Perceptual Reasoning indices. Composed of just the Verbal Comprehension Index and the Perceptual Reasoning Index, the Processing Speed and Working Memory indices are not used in calculating the score. It is becoming standard practice to derive the General Ability Index, especially when there are large disparities among the index scores (Flanagan & Kaufman, 2004; Weiss, Saklofske, Prifitera, & Holdnack, 2006). Dawn Flanagan and Alan Kaufman (2004), in *Essentials of WISC-IV Assessment*, emphasize that the Full Scale IQ (FSIQ) should *not* be reported if the variance from the highest to lowest composite score is 23 points (1.5 s.d.) or greater. Differences of this magnitude render the Full Scale IQ uninterpretable, as it is not a unitary construct.

The gifted group in the WISC-IV *Technical Manual* showed a 14-point discrepancy. Their mean Full Scale IQ score was 123.5 (Wechsler, 2003). Working Memory and Processing Speed were substantially lower than Verbal Comprehension and Perceptual Reasoning. Gifted Development Center's (GDC) research with 103 children (Falk, Silverman, & Moran, 2004) yielded much larger disparities (Table 48.3).

Of the four indices, the Verbal Comprehension Index is clearly the best indicator of giftedness in both groups and the Perceptual Reasoning Index is the second best indicator for American children. Working Memory was elevated in the GDC sample by the substitution of Arithmetic for Letter–Number Sequencing. The mean Full Scale IQ scores of both gifted samples was definitely depressed below the gifted range,

Table 48.3 Performance of two gifted samples on WISC-IV indices compared to the control group (WISC-IV Technical Manual, p. 77)

	GDC	Gifted Norm Group	Control Group
Verbal Comprehension Index	131.7	124.7	106.6
Perceptual Reasoning Index	126.4	120.4	105.6
Working Memory Index	117.7	112.5	103.0
Processing Speed Index	104.3	110.6	102.8
Full Scale	**127.2**	**123.5**	**106.7**

even though the mean Verbal Comprehension Index in GDC's sample was high enough to qualify these students for gifted services.

There is little variation in the indices of the control group: less than four IQ points between highest and lowest subtest scores. By comparison, the mean discrepancy between highest and lowest subtest scores in GDC's gifted sample was **27.4** points. Nearly 60% of the sample had disparities between the Verbal Comprehension Index and the Processing Speed Index of 23 points and some ranged as high as 69 points—over four standard deviations. Only four children scored above 130 on Processing Speed (Rimm, Gilman, & Silverman, 2008). It is also revealing that while the gifted group demonstrated a 25-point advantage over the control group in verbal abstract reasoning, differences in Processing Speed were negligible: less than 2 points.

Other samples of gifted children reveal similar patterns. Sylvia Rimm found that while the mean Verbal Comprehension Index of her clients was in the gifted range, the mean Processing Speed Index was in the high average range (Rimm, Gilman, & Silverman, 2008). Wasserman (2006) reports "our data suggests that over 70% of the students applying for gifted placement have Processing Speed Index scores in the average range or below" (p. 2). This confirms findings of earlier studies of Wechsler scales (Reams, Chamrad, & N. Robinson, 1990). Gifted students do not perform faster on processing speed tasks than average students and scores on Processing Speed depress their Full Scale IQ scores. *Therefore, on the WISC-IV, it is **not** appropriate to require a Full Scale IQ score in the gifted range for selection to gifted programs.*

When a Full Scale IQ score is derived, Arithmetic could be considered as a substitute for Letter–Number Sequencing or Digit Span in measuring Working Memory for students seeking placement in gifted programs (Rimm, Gilman, & Silverman, 2008; Silverman, Gilman, & Falk, 2004). Children in the gifted norm sample did better on Arithmetic than on any other subtest except Vocabulary (Wechsler, 2003). Arithmetic also ranked fourth as a good measure of general intelligence (**g**) (Table 48.4):

Table 48.4 g-loadings on the WISC-IV (Flanagan & Kaufman, 2004, p. 309)

Good measures of g	
Vocabulary	(0.82)
(Information)	(0.79)
Similarities	(0.79)
(Arithmetic)	(0.74)
(Word reasoning)	(0.70)
Comprehension	(0.70)
Fair measures of g	
Matrix Reasoning	(0.68)
Block Design	(0.67)
(Picture Completion)	(0.63)
Letter–Number Sequencing	(0.60)
Symbol Search	(0.58)
Picture Concepts	(0.57)
Digit Span	(0.51)
Poor measures of g	
Coding	(0.48)
(Cancellation)	(0.25)

Letter–Number Sequencing has a higher **g** loading than Digit Span, but Digit Span has a venerable history in the measurement of auditory short-term memory for nonmeaningful material, and is easier for the clinician to interpret. Therefore, at the Gifted Development Center, Arithmetic has been substituted for Letter–Number Sequencing in most assessments. However, there are cases when it would be best not to substitute Arithmetic: (1) if the child is mathephobic (Gilman & Falk, 2005); (2) if the examiner is attempting to diagnose weaknesses in working memory to document the need for accommodations; (3) if the district or agency requires a 23-point discrepancy among indices before a General Ability Index (GAI) score can be generated; or (4) if the child has AD/HD, Nonverbal Learning Disorder, or memory problems.

Unfortunately, the Coding subtest, which has never been a good predictor of giftedness in previous versions of the WISC (Kaufman, 1992), continues to be a required subtest on the WISC-IV. As can be seen from the factor loadings, Coding is a poor measure of **g** and serves to diminish scores of gifted students, whose speed of performance on clerical paper and pencil tasks is rarely as well developed as their conceptual abilities. This asynchrony or dyssynchrony in development has been found in various studies of the gifted internationally (Silverman, 1993b; Terrassier, 1985) and is another reason why processing speed should not play a role in the assessment of giftedness.

The National Association for Gifted Children (NAGC) in the United States established a Task Force in 2006 to create policies regarding the assessment of gifted students. The goals of the Task Force are to provide guidelines for the interpretation of the major individual IQ tests, as well as guidelines for differentiating among exceptionally gifted students.

As a basis for these policies, the Task Force has been conducting studies of gifted children's performance on different instruments and with various scoring methods (see Rimm, Gilman, & Silverman, 2008).

The NAGC Task Force concurs that the General Ability Index should be accepted as a basis for placement in gifted programs. The GAI norm tables are available on Harcourt Assessment's website: [http://harcourtassessment.com/hai/Images/pdf/wisciv/WISCIVTechReport4.pdf]. In addition, there are children for whom the Verbal Comprehension Index **or** Perceptual Reasoning Index provides the best estimate of ability; these indices are independently appropriate for selection to programs. Children with deficits in Working Memory or Processing Speed can succeed in programs for advanced students, provided that instructional modifications are made to fit their needs (National Association for Gifted Children, 2008).

Funds for assessing gifted students are usually limited. For this reason, many districts employ short forms of individual intelligence scales, such as the *Wechsler Abbreviated Scale of Intelligence* (WASI) (Wechsler, 1999). A reasonable alternative would be to administer only the six core subtests from which the General Ability Index can be derived: Vocabulary, Similarities, Comprehension, Block Design, Matrix Reasoning and Picture Concepts (Silverman et al., 2004). As most of these subtests are richly loaded in general intelligence (**g**), they are likely to locate the students who would be most successful in a gifted program. Another recommendation is to lower the cut-off score for gifted programs to 120, as the mean score for the gifted group in the Technical Manual was 123.5 (Wechsler, 2003).

Assessing Higher Ranges of Giftedness

Within the gifted population, there is a vast range of ability, corresponding to some degree to the range at the opposite end of the intellectual spectrum. The fields of psychology and education recognize that there are mildly, moderately, severely, and profoundly delayed children. They have not been as quick to recognize that the variability in the gifted range is as great or greater. Hal Robinson (1981) wrote:

> Within the top 1% of the IQ distribution, then, there is at least as much spread of talent as there is in the entire range from 1st to 99th percentile. Moreover, those we might call the "supergifted," (those with IQs 4 or more standard deviations above the mean) tend to be as unlike the "garden-variety gifted" (with IQs 2 or 3 standard deviations above the mean) as the "garden-variety gifted" are unlike children with scores clustered within 1 standard deviation of the mean of the population. (p. 71)

Camilla Benbow tested a boy who scored 199 on the SBL-M at the age of 7 years and 203 on a second administration. Julian Stanley (1990) reported that as a 14-year-old eleventh grader, the same young man earned perfect scores on the Verbal and Mathematical portions of the *Preliminary Scholastic Aptitude Test* (PSAT) and was 320 points above the 99th percentile of college-bound seniors on his National Merit Scholarship type index. "Truly, an IQ of 200 can be far more powerful than any of 150!" (p. 167). Hal Robinson (1981) observed that children in the very highest ranges of intelligence "may not fare as well in many respects as those with more moderate gifts" (p. 75). Bobbie Gilman (2008) concurs, "As a group, our most gifted students are not our highest achievers" (p. 5). "Exceptionally gifted children are our highest risk gifted population—not our most successful" (Rimm, Gilman, & Silverman, 2008, p. 184).

While there is a limit to how delayed a person can be, the degree of giftedness is limited only by our measurement tools. Julian Stanley (1990) pointed out that we have been content to measure six-foot individuals with five-foot rulers. The highest IQ found to date in the GDC population is 262+ on the SBL-M. Different methods need to be employed to locate children who are this advanced. And common nomenclature would be helpful.

The manuals for assessment instruments do not use the word "gifted." Since the earliest days of testing, IQ tests have yielded scores in the "superior" and "very superior" ranges. The term, "superior," is even more problematic than "gifted." Superior usually refers to examinees who score anywhere between the 91st and the 97th percentile, or in the 120–129 range on tests with a 15-point standard deviation. Very Superior is equivalent to gifted and applies to the 98th percentile and above. In January 1999, John Wasserman, recognizing that there was inconsistent terminology to describe the various levels of giftedness, requested that an international group of experts on children in the highest IQ ranges determine new nomenclature. The following designations were agreed upon (Table 48.5):

Table 48.5 Levels of giftedness (Wasserman, 2003, p. 435)

Level	IQ range	Standard deviations
Profoundly gifted	above 175	+5 SD
Exceptionally gifted	160–174	+4 SD
Highly gifted	145–159	+3 SD
Gifted	130–144	+2 SD

These levels were based on a standard deviation of 15 points. Use of the term "exceptionally gifted" for children in the 160–174 range came from work of Miraca Gross (e.g., Gross, 2004). There was no agreement on the nomenclature for gifted children below 130 or significantly above 175. Alternatives offered for the Superior range included "advanced" or "mildly gifted." Many programs use an IQ score of 120 as a cut-off for admission, which is appropriate, given the scores of gifted children on the new IQ tests. Also, creatively gifted, twice exceptional children, bilingual children, and gifted children from diverse cultural heritages are more likely to score on American scales in the 120 range than the 130 range. No name was designated for children in the 190 and above range as they are nearly impossible to locate on modern tests. However, the experimental scoring methods being developed may result in a need to re-examine this issue.

Robinson (1981) considered it entirely possible "that there are many more truly exceptional young children in the population than would be predicted on the basis of the normal curve alone" (p. 73). Terman observed this phenomenon with his original test as early as 1925, finding many more children above 170 IQ in his longitudinal study than predicted. In large-scale studies in several countries, J. C. Raven (1959) and J. Raven (1983b) found distributions with the *Progressive Matrices Intelligence Test* which sharply depart from the Gaussian curve. Even Wechsler (1939) noted "the wide-spread but mistaken belief that mental measures distribute themselves according to the normal curve" (p. 128). If the scores at the high end of the curve do not follow the normal curve, then deviation IQ scores become questionable and ratio IQs may be a reasonable alternative.

One approach employed by examiners of exceptionally gifted children since 1989 is similar to the method used in the Talent Searches. Julian Stanley had the ingenious idea of offering the *Scholastic Aptitude Test* (SAT) as an above-level test to middle-school students in order to differentiate among the most advanced students (Benbow & Lubinski, 2006). Now, each year, more than 300,000 gifted youth throughout the world participate in Talent Searches (Webb, Lubinski, & Benbow, 2007). In these programs, seventh and eighth grade students who achieve at the 95th (or 97th) percentile in grade-level reading or mathematics achievement tests are allowed to take US college board examinations (e.g., SAT-1 or ACT). College board exams were designed to differentiate among capable high school seniors for college placement purposes. When such a difficult examination is given to 12 and 13-year-olds, those who appear on achievement tests to be similar in abilities are discovered to have vastly different levels of ability. For example, two students with 97th percentile scores in math achievement may obtain scores on the SAT-Math anywhere between 200 (lowest score possible) and 800 (highest score possible). Talent Searches enable highly gifted youth the opportunity to demonstrate their full capabilities, perhaps for the first time, and it becomes apparent that they are ready for considerably advanced work.

Just as the Talent Searches employ a two-step process to differentiate the most able students, younger children who are exceptionally or profoundly gifted can be found by using a combination of two different measures: one comparing them with others their own age and one with a higher ceiling, like the SAT, that compares their abilities to those of older children. Because it is organized by age levels, with increasing levels of difficulty, all the way to Superior Adult III, the "Binet-type age scale might be considered the original examination suitable for extensive out-of-level testing" (Stanley, 1990, p. 167).

Talent Searches provide out-of-level testing to children at the 95th or 97th percentile in reading and mathematics; examiners who assess the exceptionally gifted offer out-of-level testing to children who achieve on a standard IQ test (e.g., the WISC-IV, WPPSI-III, SB5, DAS-2, KABC-2, etc.) at or above the 99th percentile on at least two subtests (particularly those subtests that are good measures of **g**). In the case of the *Raven's Progressive Matrices,* a score at the 97th percentile would probably warrant an out-of-level supplementary test. A policy put forth in 1989 and adopted by examiners of the exceptionally gifted, the **SBL-M is given purely as a supplemental measure** to test the limits of children's abilities when they achieve ceiling-level scores on tests with lower ceilings (Silverman & Kearney, 1989; see also, Silverman & Kearney, 1992; Wasserman, 2007b). When the child exceeds the scores in the norm table of the SBL-M, a formula IQ is derived according to the instructions on page 339 in the manual (Terman & Merrill, 1973). The formula IQ is a ratio metric.

To document that the child's abilities exceed the measuring tool, examiners such as Betty Meckstroth track the number of raw score points earned beyond the minimum score required to attain a 19 (the highest possible subtest score on a Wechsler scale). Children have been found who scored 13 or more raw score points beyond the ceiling on Vocabulary (19 + 13) and 8 extra points on Similarities (19 + 8) (Rimm, Gilman, & Silverman, 2008). It is obvious that the abilities of such children are not fully tapped on the WISC-IV. However, Pearson Assessment has created extended norms for the WISC-IV with maximum IQ and index scores in the 180–210 range for children under 10, depending on age and composite score. The extended norms were posted February 7, 2008, on the Pearson website in *WISC-IV Technical Report #7:* http://harcourtassessment.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=015-8979-044&Mode=resource.

Riverside Publishing provides guidelines for the current use of the SBL-M as a supplemental test (Carson & Roid, 2004). Wasserman (2007b) considers this practice warranted:

> The SBL-M remains unmatched in its breadth of procedures and is probably truer to the changing nature of cognitive-intellectual abilities over development than any test subsequently published. Its unique age-scale format and liberal discontinue rules enable testing to continue far beyond one's chronological age, thereby providing examinees with an opportunity to demonstrate considerably advanced competencies.
>
> I consider this test appropriate only after an examinee has approached the ceiling of a more recently normed test (such as the WISC-IV or SB5. . .), as a method of resolving just how far above the ceiling the examinee's true abilities may lie. When reported in an appropriately conservative manner (because of its limitations), the ratio IQ approach provides the only available means of estimating intelligence in exceptionally and profoundly gifted ranges that has any prior foundations in research (e.g., the work of Terman and Hollingworth). (Wasserman, 2007b, p. 51)

Whenever the continued use of the SBL-M is discussed, the inevitable question arises, "But what about the Flynn effect?" James Flynn (1984) discovered that IQ scores for the last three generations have increased 3 IQ points per year across the globe. This information made a giant wave in the testing industry and has led to grave concerns about using old norms that might inflate scores. Flynn (2006) recommends that IQ scores be corrected for the Flynn effect, which has been done by some psychologists, such as Brian Start in Australia and Sylvia Rimm in the United States (Rimm, Gilman, & Silverman, 2008). However, there is counter evidence to Flynn's findings. In his recent update on the Flynn effect, John Wasserman (2007a) states:

> [Flynn] advocates that the Flynn effect, which was derived from large group studies, be used to generate corrected scores for individual test findings, in spite of the likelihood that such corrections are likely to contain much more error than accurate prediction.
>
> My January, 2007 examination of psychological research databases suggests that the Flynn effect has not yet been adequately demonstrated for all levels of ability . . . there is no substantive evidence for its validity with high ability individuals (particularly those who are intellectually gifted) . . .
>
> I have yet to see any sound empirical studies of the Flynn effect in gifted samples. (p. 1)

Newer studies suggest that the Flynn effect may have tapered off at the beginning of the 1990s (Teasdale & Owen, 2005).

A concern sometimes expressed is that scores beyond the norm table will give parents excessive expectations of the child's abilities, and the parents may make unreasonable demands of the child's school. Educators may be uncomfortable with extraordinarily high scores if they do not see evidence of these abilities in the classroom. Classrooms have ceiling effects. Exceptionally gifted children often know more than the teacher is teaching or the classroom tests are testing, but they have no chance to display their advanced knowledge. The antidote to ceiling effects in assessment and in the classroom is to provide these children with opportunities to demonstrate advanced problem-solving abilities. At the very least, it is essential to be aware that these children do exist.

Another means of differentiating children in the highest IQ ranges is by using ratio IQ scores, such as the Rasch-Ratio scores on the SB5. These are already in use in deriving the Extended IQ score. Sylvia Rimm (2006) has been experimenting with a ratio method of scoring the WISC-IV, utilizing test-age equivalents for subtests provided in the manual (p. 253), converted into months, to determine a child's mental age. Age-equivalent scores reflect all correct items for each subtest. Similar to the original ratio IQ scores derived by Terman (1916b), the child's mental age is divided by the chronological age and multiplied by 100 to create a Rimm Ratio. Rimm Ratios are helpful for identifying young children in the highly, exceptionally, and profoundly gifted ranges; they appear to approximate SBL-M scores (Rimm, 2006). One drawback of Rimm Ratios is the necessity of hand scoring the test-age equivalents of each subtest and then averaging them to obtain a mental age. Also, since the ceiling of the Test-Age Equivalents is low, the older the child, the harder it is to obtain an accurate Rimm Ratio without extrapolating beyond the ceiling of the chart. After discussions with Sylvia Rimm, Pearson Assessment developed an extended norm table for the WISC-IV. The table does not generate scores as high as Rimm Ratios, but it is the first time that Wechsler scales have exceeded 160.

While there is controversy about the use of ratio IQ scores, interest in alternative scores like the ratio IQ and other indices is growing in popularity, even in nonverbal tests (Wasserman, 2005). It is gratifying that test constructors are now aware of the existence of exceptionally and profoundly gifted children and that the major test publishers are attempting to create new scoring methods in order to locate our brightest children.

Assessing Gifted Children with Learning Disabilities

Twice exceptional children are more complicated to assess than gifted children without learning disabilities. Their strengths mask their weaknesses and their weaknesses pull down their IQ scores so that they often do not qualify for gifted programs. The interaction of giftedness with various learning disabilities is covered elsewhere (Silverman, 2003), but it is necessary to mention a series of guidelines to consider when assessing the growing number of twice exceptional children worldwide.

Misdiagnosis of gifted children with learning disabilities frequently occurs because:

- their scores are averaged, masking both their strengths and their weaknesses;
- they are compared to the norms for average children instead of to their own strengths;
- their lower scores may not be significantly below the norm;
- their ability to compensate often inflates their lower scores; and
- the magnitude of the disparities between their strengths and their weaknesses is not fully taken into account. (Silverman, 2000, p. 158)

Diagnosticians are trained to look at test scores from a normative perspective. A child's scores are compared to the norms in the manual for children that age. The diagnostic question usually raised is "How does this child's performance compare to the norm?" If the child scores within the average range, no disability is detected. To understand gifted children with learning disabilities, it is necessary to ask an entirely different question: "*To what extent does the discrepancy between this child's strengths and weaknesses cause frustration and interfere with the full development of the child's abilities*?" This is an intrapersonal rather than normative view of test interpretation (Silverman, 1998), and it is critical in working effectively with this population.

A subtest scatter of 9 points or greater (3 s.d.) usually signals a potential disability. However, gifted children with even more dramatic discrepancies may be overlooked if their lowest scores fall within the broad range of "average." For example, if an average child had a high subtest of 13 and a low subtest of 4 on a WISC (3 s.d.), the lower score would be at the 2nd %, in the borderline range, which would gain attention. But a highly gifted child with 19s (99.9th %) on Similarities, Vocabulary, or Block Design, who attains a 6 on Coding (> 4 s.d. disparity), would not be identified as learning disabled, because 6 (9th %) is at the low end of the average range (Silverman, 2001). This normative interpretation of test data prevents twice exceptional children from gaining services for either exceptionality.

Two key diagnostic indicators of giftedness combined with learning disorders are significant (2 s.d.) disparities between strengths and weaknesses, with some subtest scores in at least the superior range, and succeeding on the harder items, while missing the easier items. When both are present, it is likely that the child is twice exceptional.

Gifted children with learning disabilities have much more erratic scores over time than other children. Some twice exceptional children achieve higher scores as they get older, some have lower scores, and some have scores that vary dramatically in unpredictable directions on different tests. Many factors affect their performance. They tend to do poorer on timed tests and on tasks that require handwriting. And their ability to demonstrate the true level of their strengths will be considerably different on days when their compensation mechanisms work well and on days when they do not (Silverman, 2000). Therefore, the highest score attained at any age on any test may well be the best estimate of a twice exceptional child's abilities (Rimm, Gilman, & Silverman, 2008; Silverman, 2003).

Conclusion

To select gifted children for special services or programs, it is essential to use the right tools. Those instruments or portions of instruments that have the richest loadings on general intelligence (**g**) are the most useful for locating gifted children. *Raven's Progressive Matrices,* the Stanford-Binet scales, and the Wechsler scales were all founded on the conception of intelligence as abstract reasoning (**g**). Abstract reasoning and general intelligence (**g**) are synonymous. Giftedness is high abstract reasoning (Silverman, 1993a). Therefore, **g** could as easily stand for *giftedness* as for *general intelligence*.

The best measure of **g**, according to Spearman, is the *Raven's Progressive Matrices.* The *Progressive Matrices* are also an excellent way to find gifted visual-spatial learners. They should be administered in conjunction with a vocabulary test. Individual IQ tests are more valid and reliable than group tests. The newest editions of the major individual tests are the *Wechsler Intelligence Scale for Children, Fourth Edition* (WISC-IV), and the *Stanford-Binet Intelligence Scale, Fifth Edition* (SB5). Older versions of these scales still have utility. The new scales offer flexibility in administration and scoring. It is not necessary to use Full Scale IQ scores for the purpose of program selection, and, when factor scores differ widely, it is inappropriate to derive Full Scale scores. The General Ability Index (GAI) of the WISC-IV is suitable for identifying gifted children. It requires the administration of only six subtests.

The WISC-IV is a better instrument for locating students with high abstract verbal reasoning, and it is less timed than the WISC-III. The SB5 is a better instrument for locating students with mathematical and scientific talent and visual-spatial abilities, although gifted visual-spatial learners will often have high scores on the Perceptual Reasoning Index of the WISC-IV, as well. It is permissible to use portions of either instrument to locate gifted children with different strengths or from diverse cultures. The Verbal Comprehension Index and the Perceptual Reasoning Index can be used independently. The same is true for the Verbal and Nonverbal IQ scores on the SB5.

Given the fact that the two major IQ tests yielded mean Full Scale IQ scores of 123.5 and 124 on their gifted samples, IQ cut-off scores for gifted programs should probably be set at 120 (Falk, Silverman, & Moran, 2004). The highest index or factor score is usually the best predictor of success in the gifted program, if the program is responsive to the learning strengths of the students. When a child tops out on a test with a low ceiling, achieving two or more subtests at the 99th percentile, or a Raven's score at the 97th percentile, it is recommended that a test with a higher ceiling, such as the SBL-M, be administered as a supplemental test, or a ratio-based scoring method be used, to determine the full strength of the child's abilities.

Ratio IQ scores may vary considerably from deviation IQ scores, as they are based on a different metric. The higher the child's IQ, the greater the discrepancy is likely to be. For placement in gifted programs, GAI scores on the WISC-IV or a Gifted Composite score on the SB5 may be sufficient. However, when making decisions about the degree of acceleration needed or to qualify a child for programs for the exceptionally gifted, supplemental testing or ratio-based scoring is recommended. Given the low ceilings on most tests, the highest score a gifted child achieves on any measure is probably the best estimate of his or her abilities.

In modern times, we have gained in efficiency and ease of test administration and scoring, in representation of diverse cultures, and in the assessment of multiple factors of intelligence. We have improved our measurement of visual-spatial abilities. However, in the bargain, we seem to have lost the capacity to assess the full range of gifted children's abilities. Ratio-based scoring methods hold promise in locating our most advanced children.

Acknowledgments The author would like to thank Kathi Kearney and John Wasserman for their assistance in the preparation of this manuscript.

References

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.

Bannatyne, A. (1974). Diagnosis: A note on recategorization of the WISC scaled scores. *Journal of Learning Disabilities, 7*, 272–273.

Bayley, N. (1970). Development of mental abilities. In P. H. Mussen (Ed.), *Carmichael's manual of child psychology* (3rd ed., Vol. 1). New York: John Wiley.

Benbow, C. P., & Lubinski, D. (2006). Julian C. Stanley, Jr. (1918–2005). *American Psychologist, 61*, 251–252.

Binet, A. (1909). *Les idees modernes sur les enfants*. Paris: Flammarion.

Binet, A. (1911). Nouvelles recherches sur la mesure du niveau intellectual chez les enfants d'ecole. *L'Annee Psychologique, 17*, 145–201.

Binet, A., & Simon, Th. (1905). Application des methods nouvelle au diagnostic du niveau intellectual chez des enfants normaux et anormaux d'hospice et d'ecole primaire. *L'Annee Psychologique, 11*, 191–244.

Binet, A., & Simon, Th. (1908). Le developpement de l'intelligence chez les enfants. *L'Annee Psychologique, 14*, 1–94.

Burt, C. (1949). The structure of the mind: A review of the results of factor analysis. *British Journal of Educational Psychology, 19*, 100–111; 176–199.

Canter, A. (1990). A new Binet, an old premise: A mismatch between technology and evolving practice. *Journal of Psychoeducational Assessment, 8*, 443–450.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge: Cambridge University Press.

Carson, D. & Roid, G. (2004). *Acceptable use of the Stanford-Binet Form L-M: Guidelines for the professional use of the Stanford-Binet Intelligence Scale, Third Edition (Form L-M).* Itasca, IL: Riverside Publishing.

Cattell, J. M. (1890). Mental tests and measurements. *Mind, 15*, 373–381.

Cattell, J. M. (1915). Families of American men of science. *Popular Science Monthly, 86*, 504–515.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*, 1–22.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Boston: Houghton Mifflin.

Cronbach, L. J. (1970). *Essentials of psychological testing* (3rd ed.). New York: Harper & Row.

DuBois, P. H. (1970). *A history of psychological testing*. Boston: Allyn & Bacon.

Falk, R. F., Silverman, L K., & Moran, D. (2004, November). *Using two WISC-IV indices to identify the gifted*. Paper presented at the 51st Annual Convention of the National Association for Gifted Children, Salt Lake City, UT.

Flanagan, D. P., & Kaufman, A. S. (2004). *Essentials of WISC-IV assessment*. Hoboken, NJ: John Wiley & Sons.

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin, 95*, 29–51.

Flynn, J. R. (1999). Searching for justice: The discovery of IQ gains over time. *American Psychologist, 54*, 5–20.

Flynn, J. R. (2006). Tethering the elephant: Capital cases, IQ, and the Flynn effect. *Psychology, Public Policy, and Law, 12*, 170–189.

Galton, F. (1869). *Hereditary genius: An inquiry into its causes and consequences.* London: Macmillan.

Galton, F. (1883). *Inquiries into human faculty and its development*. London: J. M. Dent & Sons.

Gardner, H. G. (1983). *Frames of mind: A theory of multiple intelligences*. New York: Basic Books.

Gilman, B. (2008). *Challenging highly gifted learners*. Waco, TX: Prufrock Press.

Gilman, B. J., & Falk, R. F. (2005, August). *Research-based guidelines for use of the WISC-IV in gifted assessment*. Paper presented at the 16th Biennial Conference of the World Council for Gifted and Talented Children, New Orlean, LA.

Gilman, B. J., & Kearney, K. (2004, November). *From conceptual to practical: Making gifted testing relevant.* Paper presented at the 51st annual convention of the National Association for Gifted Children, Salt Lake City, UT.

Goddard, H. H. (1910). A measuring scale of intelligence. *Training School, 6*, 146–154.

Gross, U. M. (2004). *Exceptionally gifted children*. (2nd ed.). London: Routledge Falmer.

Guilford, J. P. (1956). The structure of intellect. *Psychological Bulletin, 53*, 267–293.

Guilford, J. P. (1988). Some changes in the structure-of intellect model. *Educational and Psychological Measurement, 48*, 1–4.

Hildreth, G. H. (1966). *Introduction to the gifted.* New York: McGraw-Hill.

Horn, J. L. (1988). Thinking about human abilities. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., pp. 645–685). New York: Plenum.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology, 57*, 253–276.

Kaufman, A. S. (1992). Evaluation of the WISC-III and WPPSI-R for gifted children. *Roeper Review, 14*, 154–158.

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

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

Kearney, K., & LeBlanc, J. (1993). Forgotten pioneers in the study of gifted African-Americans. *Roeper Review, 15*, 192–199.

Lovecky, D. V., Kearney, K., Falk, R. F., & Gilman, B. J. (2005, August). *A comparison of the Stanford-Binet 5 and the Stanford-Binet Form L-M in the assessment of gifted children.* Paper presented at the 16th Biennial Conference of the World Council for Gifted and Talented Children, New Orleans, LA.

Lubin, B., Wallis, R. R., & Paine, C. (1971). Patterns of psychological test usage in the United States: 1935 – 1969. *Professional Psychology, 2*, 70–74.

Matarazzo, J. D. (1981). David Wechsler (1896–1981). *American Psychologist, 36*, 1542–1543.

McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf-Gc framework. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 151–179). New York: Guilford.

McGrew, K. S. (2007). *Cattell-Horn-Carroll CHC (Gf-Gc) theory: Past, present & future. 6.1 A. Conclusions and caveats*. Institute for Applied Psychometrics. (Retrieved from http://www.iapsych.com/CHCPP/A.ConclusionsandCaveats.html on January 25, 2007)

National Association for Gifted Children (Jan. 2008). *Use of the WISC-IV for gifted identification*. Retrieved January 23, 2008, from http://www.nagc.org/index.aspx?id=375

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.

Raven, J. C. (1938). *Progressive Matrices*. London: H. K. Lewis.

Raven, J. C. (1958). *Standard Progressive Matrices (Revised): Sets A, B, C, D, & E*. Oxford: Oxford Psychologists Press.

Raven, J. C. (1959). A note on Burt's "The distribution of intelligence." *British Journal of Psychology, 47*, 95–100.

Raven, J. C. (1965). *Guide to using The Coloured Progressive Matrices: Sets A, Ab, B*. London: H. K. Lewis.

Raven, J. (1981). *The 1979 British standardization of the Standard Progressive Matrices and Mill Hill Vocabulary Scale. Research Supplement No. 1 to the manual for Raven's Progressive Matrices and vocabulary scales*. London: H. K. Lewis.

Raven, J. (1983a). The Progressive Matrices and Mill Hill Vocabulary Scale in Western societies. In S. H. Irvine & J. W. Berry (Eds.), *Human assessment and cultural factors* (pp. 107–114). New York: Plenum.

Raven, J. (1983b). Raven Coloured Progressive Matrices Intelligence Test in Thailand and in Denmark: A response. *School Psychology International, 4*, 173–176.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Advanced Progressive Matrices* (1998 edition). Oxford, England: Oxford Psychologists Press.

Raven, J., & Summers, B. (1986). *Manual for Raven's Progressive Matrices and Vocabulary Scales*—Research supplement no. 3. London: H. K. Lewis.

Reams, R., Chamrad, D., & Robinson, N. M. (1990). The race is not necessarily to the swift: Validity of the WISC-R bonus points for speed. *Gifted Child Quarterly, 34*, 108–110.

Rich, C. C., & Anderson, R. P. (1965). A tactual form of the Progressive Matrices for use with blind children. *Personnel & Guidance Journal, 43*, 912–929.

Rimm, S. (2006, November). *Breaking the ceiling scores for profoundly gifted children using the WISC-IV.* Paper presented at the 53rd annual convention of the National Association of Gifted Children, Charlotte, NC.

Rimm, S., Gilman, B. J., & Silverman, L. K. (2008). Non-traditional applications of traditional testing. In J. VanTassel-Baska (Ed.), *Alternative assessments with gifted and talented students* (pp. 175–202). Waco, TX: Prufrock Press.

Robinson, H. B. (1981). The uncommonly bright child. In M. Lewis & L. A. Rosenblum (Eds.), *The uncommon child* (pp. 57–81). New York: Plenum Press.

Roid, G. H. (2003a). *Stanford-Binet Intelligence Scales, Fifth Edition, Technical Manual*. Itasca, IL: Riverside.

Roid, G. H. (2003b). *Stanford-Binet Intelligence Scales interpretive manual: Expanded guide to the interpretation of SB5 test results*. Itasca, IL: Riverside.

Roid, G. H., & Barram, R. A. (2004). *Essentials of Stanford-Binet (SB5) assessment*. New York: Wiley.

Roid, G. H., & Carson, A. (2003). *Special composite scores for the SB5*. Assessment Service Bulletin No. 4. Itasca, IL: Riverside.

Ruf, D. L. (2003). *Use of the SB5 in the assessment of high abilities*. Assessment Service Bulletin No. 3. Itasca, IL: Riverside.

Seagoe, M. V. (1975). *Terman and the gifted*. Los Altos, CA: William Kaufman.

Sharp, S. E. (1899). Individual psychology: A study in psychological method, *American Journal of Psychology, 10*, 329–391.

Shouksmith, G. (1970). *Intelligence, creativity and cognitive style*. New York: Wiley.

Silverman, L. K. (1986). An interview with Elizabeth Hagen: Giftedness, intelligence and the new Stanford-Binet. *Roeper Review, 8*, 168–171.

Silverman, L. K. (1993a). A developmental model for counseling the gifted. In L. K. Silverman (Ed.), *Counseling the gifted & talented* (pp. 51–78). Denver: Love.

Silverman, L. K. (1993b). The gifted individual. In L. K. Silverman (Ed.), *Counseling the gifted & talented* (pp. 3–28). Denver: Love.

Silverman, L. K. (1995). Highly gifted children. In J. Genshaft, M. Bireley, & C. L. Hollinger (Eds.), *Serving gifted and talented students: A resource for school personnel* (pp. 217–240). Austin, TX: Pro-Ed.

Silverman, L. K. (1998). Through the lens of giftedness. *Roeper Review, 20*, 204–210.

Silverman, L. K. (2000). The two-edged sword of compensation: How the gifted cope with learning disabilities. In K. Kay (Ed.). *Uniquely gifted: Identifying and meeting the needs of twice exceptional learners* (pp. 153–159). Gilsum, NH: Avocus.

Silverman, L. K. (2001). Diagnosing and treating visual perceptual issues in gifted children. *Journal of Optometric Vision Development, 32*(3), 153–176.

Silverman, L. K. (2002). *Upside-down brilliance: The visual-spatial learner.* Denver: DeLeon.

Silverman, L. K. (2003). Gifted children with learning disabilities. In N. Colangelo & G. Davis (Eds.), *Handbook of gifted education* (3rd ed., pp. 553 – 543). Boston: Allyn & Bacon.

Silverman, L. K., Gilman, B. J., & Falk, R. F. (2004, November). *Who are the gifted using the new WISC-IV?* Paper presented at the 51st annual convention of the National Association for Gifted Children, Salt Lake City, UT.

Silverman, L. K. & Kearney, K. (1992). Don't throw away the old Binet, *Understanding Our Gifted, 4*(4), 1, 8–10.

Silverman, L. K., & Kearney, K. (1989). Parents of the extraordinarily gifted. *Advanced Development, 1*, 41–56.

Snyderman, M., & Rothman, S. (1988). *The IQ controversy, the media and public policy*. New Brunswick, NJ: Transaction Books.

Spearman, C. E. (1904). "General intelligence": Objectively determined and measured. *American Journal of Psychology, 15*, 201–292.

Spearman, C. E. (1923). *The nature of intelligence and the principles of cognition.* London: Macmillan.

Spearman, C. E. (1927). *The abilities of man*. New York: Macmillan.

Stanley, J. C. (1990). Leta Hollingworth's contributions to above-level testing of the gifted. *Roeper Review, 12*(3), 166–171.

Stern, W. (1910). Das ubernormal kind. *Der Saemann*, 67–72. [Translated and reprinted in 1911 as The supernormal child. *Journal of Educational Psychology, 2*, 143–148, 181–190.]
Teasdale, T. W., & Owen. D. R. (2005). A long-term rise and recent decline in intelligence test performance: The Flynn Effect in reverse. *Personality and Individual Differences, 39*, 837 – 843.
Terman, L. M. (1916a). *The measurement of intelligence*. Boston: Houghton Mifflin.
Terman, L. M., (1916b). *The Stanford revision of the Binet-Simon tests*. Boston: Houghton Mifflin.
Terman, L. M. (1917). The intelligence quotient of Francis Galton in childhood. *American Journal of Psychology, 28*, 209–215.
Terman, L. M. (1921). In E. L. Thorndike, et al. Intelligence and its measurement: A symposium. *Journal of Educational Psychology, 12*, 127–133.
Terman, L. M. (1925). *Genetic studies of genius, Vol. 1: Mental and physical traits of a thousand gifted children*. Stanford, CA: Stanford University Press.
Terman, L. M., & Merrill, M. A. (1937). *Measuring intelligence: A guide to the administration of the new revised Stanford-Binet tests of intelligence*. Boston: Houghton Mifflin.
Terman, L. M., & Merrill, M. A. (1960). *Stanford-Binet Intelligence Scale: Manual for the third revision. Form L-M*. Boston: Houghton Mifflin.
Terman, L. M., & Merrill, M. A. (1973). *The Stanford-Binet Intelligence Scale: 1973 norms edition*. Boston: Houghton Mifflin.
Terrassier, J-C. (1985). Dyssynchrony-uneven development. In J. Freeman (Ed.). *The psychology of gifted children* (pp. 265–274). New York: John Wiley.
Thorndike, E. L. (1941). Mental abilities. *Proceedings of the American Philosophical Society Symposium on Recent Advances in Psychology, 84*, 503–513.
Thorndike, R. L. (1975). Mr. Binet's test 70 years later. *Educational Researcher, 4*, 3–4.
Thorndike, R. L. (1986). Historical and theoretical perspectives. In R. L. Thorndike, E. P. Hagen, & J. M. Sattler (1986). *The Stanford-Binet Intelligence Scale: Fourth edition. Technical manual* (pp. 1–7). Itasca, IL: Riverside.
Thorndike, R. L., & Hagen, E. P. (1971). *Cognitive Abilities Test*. Boston: Houghton Mifflin.
Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *The Stanford-Binet Intelligence Scale: Fourth edition. Technical manual*. Itasca, IL: Riverside.
Thurstone, L. L. (1938). Primary mental abilities. *Psychometric Monographs, No. 1*.
Vernon, P. E. (1951). *The structure of human abilities*. New York: John Wiley.
Vernon, P. E. (1987). The demise of the Stanford-Binet scale. *Canadian Psychology/Psychologie Canadienne, 28*(3), 251–258.
Wasserman, J. D. (2003). Assessment of intellectual functioning. In J. R. Graham & J. A. Naglieri (Eds.), *Handbook of psychology, Volume 10: Assessment psychology* (pp. 417–442). Hoboken, NJ: Wiley.
Wasserman, J. D. (2005, November). *Utility of nonverbal measures with the gifted*. Paper presented at the 16th Biennial Conference of the World Council for Gifted and Talented Children, New Orleans, LA.
Wasserman, J. D. (2006). *Tips for parents: Intellectual assessment of exceptionally and profoundly gifted children*. Unpublished manuscript available online at http://www.gt-cybersource.org/Record.aspx?NavID=2_0&rid=14107
Wasserman, J. D. (2007a). The Flynn effect in gifted samples: Status as of 2007. Unpublished manuscript available online at http://www.gifteddevelopment.com/Whats_New/flynn.htm
Wasserman, J. D. (2007b). Intellectual assessment of exceptionally and profoundly gifted children. In K. Kay, D. Robson, & J. F. Brenneman (Eds.), *High IQ kids: Collected insights, information and personal stories from the experts* (pp. 48–65). Minneapolis, MN: Free Spirit Press.
Watt, D. C. (1998, January). Lionel Penrose, F. R. S. (1898–1972) and eugenics: Part One. *Notes and Records of the Royal Society of London, 52*(1), 137–151.
Webb, R. M., Lubinski, D., & Benbow, C. P. (2007). Spatial ability: A neglected dimension in talent searches for intellectually precocious youth. *Journal of Educational Psychology, 99*, 397–420.
Wechsler, D. (1939). *The measurement of adult intelligence*. Baltimore: Williams & Wilkins.
Wechsler, D. (1949). *Wechsler Intelligence Scale for Children Manual*. New York: The Psychological Corporation.
Wechsler, D. (1955). *Wechsler Adult Intelligence Scale Manual*. New York: The Psychological Corporation.
Wechsler, D. (1958). *The measurement and appraisal of adult intelligence* (4th ed.). Baltimore: Williams & Wilkins.
Wechsler, D. (1967). *Wechsler Preschool and Primary Scale of Intelligence*. New York: The Psychological Corporation.
Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children—revised*. New York: The Psychological Corporation.
Wechsler, D. (1981). *WAIS-R manual: Wechsler Adult Intelligence Scale—Revised*. San Antonio, TX: The Psychological Corporation.
Wechsler, D. (1989). *Manual for the Wechsler Preschool and Primary Scale of Intelligence–Revised*. San Antonio, TX: The Psychological Corporation.
Wechsler, D. (1991). *Manual for the Wechsler Intelligence Scale for Children, Third Edition*. San Antonio, TX: The Psychological Corporation.
Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: The Psychological Corporation.
Wechsler, D. (2003). *The WISC-IV technical and interpretive manual*. San Antonio, TX: Psychological Corporation.
Weiss, L. G., Saklofske, D. H., Prifitera, A., & Holdnack, J. A. (2006). *WISC-IV Advanced clinical interpretation*. Burlington, MA: Academic Press.
Wilson, W. M. (1992). The Stanford-Binet Fourth Edition and Form L-M in assessment of young children with mental retardation. *Mental Retardation, 30*(2), 81–84.