

Assessment of the Gifted: Making Sense of the New Tests

Linda Kreger Silverman, Ph.D.
Gifted Development Center

There has always been more perplexity in the assessment of gifted children than in the assessment of any other population, due to surprising discrepancies in the IQ scores they attain on various tests. Average children and developmentally delayed children usually obtain rather consistent IQ scores on different instruments. However, a profoundly gifted child can score 120 on one IQ test and 220 on another—a discrepancy of *100 points*! These discrepancies are attributable to varying ceilings of tests.

In recent years, the perplexity has been compounded by the fact that the major individual IQ tests have been completely reformulated. Scores generated by the latest revisions are not comparable to scores derived from previous versions of the same instrument. It is not only more difficult for gifted students to qualify for special programming, it is also more difficult for learning disabled children to qualify for services (not to mention twice-exceptional children). In attempts to address these issues, the companies producing the IQ tests have created several ways of scoring the same test. Needless to say, this has caused major confusion in the field.

Which test is best for selecting gifted students? Which version of the test? Which method of scoring should be used? Unfortunately, there are no easy answers. The purpose of this symposium is to share the findings of those who have compared the performance of gifted children on different instruments, as well as to share alternative methods of identifying gifted children. The answer to what is the best method ultimately will need to be decided by the reader.

The New IQ Tests

The *Stanford-Binet Intelligence Scales, Fifth Revision (SB5)*, was released in February, 2003. Unlike the *Wechsler Intelligence Scale for Children, Third Edition (WISC-IV)*, which is designed for ages 6 through 16, the SB5 was designed for individuals between 2 and 85 years of age. It generates three IQ scores: Verbal, Nonverbal and Full Scale. In addition, there are five factor scores: Fluid Reasoning, Knowledge, Quantitative Reasoning, Visual-Spatial Reasoning and Working Memory. Each factor has a verbal and nonverbal scale. To accommodate the wide age range, there are six levels of the test. For the gifted, all of the scales except Working Memory seem to be useful in program selection. The emphasis on visual-spatial reasoning and the liberal time limits make it appealing for locating visual-spatial learners.

This fifth edition is an enormous departure from the earlier versions of the test: the *Stanford-Binet Intelligence Scale, Fourth Edition (SB-IV)*, released in 1986, and the *Stanford-Binet Intelligence Scale (Form L-M) (SBL-M)*, released in 1960 and renamed in 1972. All three of the Binet tests are based on different conceptions of intelligence. Lewis Terman, who developed the SBL-M with his colleagues, defined intelligence as abstract reasoning. The power of the original test came from its attempt to measure this unitary concept of intelligence. The SB-IV was based on

the conception of intelligence as fluid and crystallized abilities (Thorndike, Hagen & Sattler, 1986). It generated four area scores: Verbal Reasoning, Abstract/Visual Reasoning, Quantitative Reasoning and Short-Term Memory. These collectively composed a total score, called a Composite IQ score.

The Wechsler Intelligence Scale for Children, Fourth Edition

The most popular individual IQ test used in American schools is the *Wechsler Intelligence Scale for Children (WISC)*. The fourth edition was released in August, 2003. It is quite a departure from its predecessor, the WISC-III. The traditional Verbal and Performance scores have vanished. Four indices now take center stage: Verbal Comprehension, Perceptual Reasoning, Working Memory and Processing Speed. For the gifted, Verbal Comprehension and Perceptual Reasoning are emerging as the most important factors to take into consideration in the selection to gifted programs.

The greatest changes have occurred in the nonverbal section of the test. The former Perceptual Organization Index was composed of Picture Completion, Picture Arrangement, Block Design and Object Assembly. Of the four subtests, only Block Design remains as a core subtest. Picture Completion is now an optional test, and it is not administered very often, due to time constraints. Picture Arrangement and Object Assembly have been eliminated. In their place are two new subtests, Picture Concepts, which is a visual version of the Similarities subtest, and Matrix Reasoning, which measures spatial perception. The triad of Block Design, Picture Concepts and Matrix Reasoning enable the WISC-IV to provide a better assessment of visual-spatial abilities than the WISC-III. However, what has been lost in the bargain is the power to diagnose visual processing deficits. It is also important to note that the WISC-IV is far less timed than the WISC-III.

Dawn Flanagan and Alan Kaufman (2004), in *Essentials of WISC-IV Assessment*, argue that the Full Scale IQ (FSIQ) should *not* be reported if the variance from the highest to lowest composite score is 23 points or greater. The gifted group in the WISC-IV normative sample showed a 13-point discrepancy, suggesting the likelihood of many gifted children whose FSIQs should not be used.

The Gifted Development Center’s (GDC) research with 103 children (Falk, Silverman & Moran, 2004) yielded even larger discrepancies. Their significance is evident when the results are compared with those of a control group in the normative sample:

| | GDC | Control Group in Norm Sample |
|----------------------------|------------|-------------------------------------|
| Verbal Comprehension Index | 131.7 | 106.6 |
| Perceptual Reasoning Index | 126.4 | 105.6 |
| Working Memory Index | 117.7 | 103.0 |
| Processing Speed Index | 104.3 | 102.8 |

Full Scale

127.2

106.7

(WISC-IV Technical Manual, p. 77)

There is little variation in the indices of the control group: less than 4 IQ points between highest and lowest subtest scores. Note that the Working Memory and Processing Speed Indices do not deflate the Full Scale IQ scores. On the contrary, the Full Scale mean is actually higher than the highest index. By comparison, the mean discrepancy between highest and lowest subtest scores in the gifted sample was **27.4** points. *Nearly 60%* of the sample had discrepancies between the Verbal Comprehension Index and the Processing Speed Index of 23 points. Discrepancies ranged as high as 69 points—over 4 standard deviations. It is also revealing that while the gifted group demonstrated a 25-point advantage over the control group in verbal abstract reasoning, their differences in Processing Speed were negligible: less than 2 points. It is obvious that gifted students do not perform faster on these processing speed tasks than average students.

Of the four indices, the Verbal Comprehension Index is clearly the best indicator of giftedness and the Perceptual Reasoning Index is the second best indicator. The mean Full Scale IQ score of the gifted sample was definitely depressed below the gifted range, even though the mean Verbal Comprehension Index was high enough to qualify these students for gifted services. By these scores, a general rule might be to eliminate consideration of the Full Scale IQ for gifted identification.

Flanagan and Kaufman (2004) suggest using the General Ability Index (GAI) instead, which, like the DWI-1 of Dumont and Willis, utilizes only the Verbal Comprehension and Perceptual Reasoning scores. This is now being supported by trainers for Harcourt Assessments (PsychCorp). If the GAI table from Flanagan and Kaufman's book were used to combine the mean Verbal Comprehension and Perceptual Reasoning Indices from the Gifted Development Center study (131.7 + 126.4), the resulting mean GAI of the gifted group would be 132, which qualifies for gifted services.

Subtests Most Appropriate for Gifted Assessment

The following chart indicates the strongest subtests for the gifted population in two different studies. (Subtests in parentheses are optional.)

WISC-IV Subtest Means of 63 Gifted Child in the Norm Sample compared with 103 Gifted Children from GDC

| | Gifted Norm Group | GDC |
|-------------------------|--------------------------|-------------|
| Similarities | 14.1 | 15.8 |
| Vocabulary | 14.6 | 15.4 |
| Comprehension | 14.1 | 14.8 |
| Matrix Reasoning | 13.4 | 14.7 |
| Picture Concepts | 12.7 | 14.6 |

| | | |
|--------------------------------|-------------|-------------|
| (Arithmetic) | 14.2 | 14.1 |
| (Information) | 13.9 | 14.1 |
| Block Design | 13.8 | 13.2 |
| (Word Reasoning) | 13.2 | 12.9 |
| Letter-Numb. Sequencing | 12.6 | 12.9 |
| (Picture Completion) | 13.0 | 12.5 |
| Symbol Search | 12.1 | 11.5 |
| Digit Span | 12.0 | 12.3 |
| (Cancellation) | 11.0 | 10.3 |
| Coding | 11.5 | 9.9 |

(WISC-IV Technical Manual, p. 77)

Similarities, Vocabulary and Comprehension make up the three required subtests to derive the Verbal Comprehension Index. Please note that, in both studies, these three subtests emerged among the highest scores for the gifted groups. Matrix Reasoning, Picture Concepts and Block Design, which make up the three required subtests to derive the Perceptual Reasoning Index, appear to be among the next strongest set of required (not optional) subtests. Picture Concepts fared considerably better with the Gifted Development Center sample than with the gifted group in the norm sample.

In the norm sample, Arithmetic surpassed all but Vocabulary, and in the Gifted Development Center sample, it ranked in sixth place. Additional information about Arithmetic can be found in the factor loadings on general intelligence. Note that it holds the highest rank as a measure of general intelligence.

Good Measures of g

| | |
|--------------|------|
| Arithmetic | .768 |
| Vocabulary | .751 |
| Information | .748 |
| Similarities | .733 |

Fair Measures of g

| | |
|--------------------|------|
| Matrix Reasoning | .687 |
| Block Design | .672 |
| Word Reasoning | .648 |
| Comprehension | .646 |
| Letter-Number Seq. | .621 |
| Picture Completion | .616 |
| Picture Concepts | .582 |
| Symbol Search | .568 |
| Digit Span | .525 |

Poor Measure of g

Coding .454

Poorest Measure of g

Cancellation .209

(Keith, Fine, Taub, Reynolds, & Kranzler, 2004)

Combining information from the performance of two sets of gifted students with the factor loadings on general intelligence, it becomes clear that Arithmetic is a much stronger measure of giftedness than Letter-Number Sequencing or Digit Span, the two required subtests for deriving the Working Memory Index. While Letter-Number Sequencing has a higher rank than Digit Span in loading on general intelligence, Digit Span produces more predictable and interpretable responses from students. Letter-Number Sequencing involves listening to a random list of letters and numbers, separating them and manipulating them in a prescribed way. An occasional response to the task is, “You want me to do *what?*” One boy took over 3 minutes each for the last few items and stated that he felt nauseated afterward.

Therefore, at the Gifted Development Center, we substitute Arithmetic for Letter-Number Sequencing in most assessments. If a child appears to be mathephobic, we do not do the substitution. Two substitutions are allowable to derive a Full Scale IQ score, as long as they reflect an *a priori* judgment before the test is administered (or unless a subtest becomes spoiled in administration).

The new Cancellation item is not particularly useful in assessing giftedness, as can be seen from both the performance of the two gifted groups and its extremely low standing as a measure of general intelligence. It is even less correlated with general intelligence than the Mazes subtest, which was removed from the WISC-IV. The Gifted Development Center uses this optional subtest only on rare occasions for diagnostic purposes. 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 both studies and the factor loadings, Coding is a poor measure of general intelligence 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 in development is typical of the gifted population (Silverman, 1993), and another reason why processing speed should not play a role in the assessment of giftedness.

Other Measures

A new version of the *Wechsler Preschool and Primary Scale of Intelligence*, the WPPSI-III, was released in 2002. While this test also varies considerably from its predecessor, the WPPSI-R, and appears to be effective with the gifted, we will not be discussing the WPPSI-III during this symposium. Selection for gifted programs is more often based on the WISC-IV and SB5 than on the WPPSI-III, due to the fact that the majority of gifted programs are designated for older children.

The *Differential Ability Scales (DAS)* has also been used for selecting students for gifted programs, and we have found it particularly useful with visual-spatial children. A new version of the DAS is expected in the near future. The *Woodcock-Johnson III Tests of Cognitive Abilities (WJ-III)*, has a higher ceiling, beyond 200 IQ in some age ranges.

An international staple in the selection of students for gifted programs has been Raven's *Progressive Matrices* (also the *Coloured Progressive Matrices*). A number of nonverbal tests have been generated in recent years, including the *Naglieri Nonverbal Abilities Test (NNAT)* and the *Universal Nonverbal Intelligence Test (UNIT)*. These tools have been particularly valuable for assessing children from culturally diverse backgrounds.

The New Norms

Norms are periodically updated to reflect the increase in intelligence in the general population. This increase is called the "Flynn Effect." Flynn (1999) suggests that the entire population of the planet is increasing in intelligence at the rate of .3 IQ points per year. For this reason, newer, more stringent test norms are usually considered more accurate than old norms. However, the Flynn Effect appears to apply to the mid-range of the IQ spectrum, with "only very minimal changes at the extremes of ability," (J. D. Wasserman, personal communication, December 23, 1997). Given this information, one would expect that scores for the gifted would vary only slightly from one set of norms to another. Yet, closer scrutiny has revealed enormous differences in the gifted range—more than three times the Flynn Effect (Silverman, 1989). How could this happen?

The normal curve appears to fit the distribution of scores within 3 standard deviations of the mean, but does not apply at the extremes (Jensen, 1980). Unusually high frequencies of highly, exceptionally and profoundly gifted children have been observed since the research of Lewis Terman (1925). However, as modern IQ tests are constructed with the assumption of a normal distribution of scores, the larger portion of gifted students actually found in the population still has to force fit within the tail of the normal curve in order to derive standard scores. This compresses the scores in the gifted range. The more extraordinary the child's ability, the greater the impact of this compression. Each new set of norms exacerbates the problem.

While psychologists are trained to believe that newer norms are more accurate for everyone, these norms do not seem to be appropriate for the gifted population. Here are some examples of how new norms differentially affect scores in the average and gifted ranges. In 1960, even though the SBL-M used ratio-based scoring, when a child of 8 years 0 months achieved a mental age of 8.0 on the SBL-M, the resulting IQ score was 100. In 1972, when the test used deviation scoring, the same raw score yielded an IQ score of around 96 (it varied by age). This is a difference of approximately 4 points or .36 per year—closely approximating the Flynn Effect. By 1986, the same raw score would have resulted in an IQ score of 92, a total discrepancy of 8 points.

By way of contrast, in 1960, a five year old achieving a mental age of 8.0 would have had a ratio-based IQ score of 165. In 1972, using deviation IQs, the same raw score yielded an IQ score

of 153, a difference of 12 points. Differences between the SB-IV, published in 1986, and the 1972 norms appear to be at least 13.5 points in the moderately gifted range (Thorndike, Hagen & Sattler, 1986), which, hypothetically, would bring the child's score down below 140 (rendering the child ineligible for a program for the highly gifted). Discrepancies of the same magnitude were found in a study of gifted children using the WISC-R, released in 1974, and the WISC-III, released in 1991. "The mean difference between WISC-R and WISC-III Full Scale IQ was 13.5 points" (Bryant, 1992, p. 13). This is an apparent *loss of one IQ point per year* for children in the gifted range. In the 26-year period from 1960 to 1986, average students needed to obtain only 8 more points to make up for the average gains in intelligence of the general population, whereas gifted children needed to obtain **over 25** more points to match previous scores (Silverman, 1989).

Discrepancies greater than would be predicted by the Flynn Effect continue in current testing (Falk, Moran & Silverman, 2003). According to the WISC-IV *Technical Manual*, the mean Full Scale IQ score of 63 children who scored in the gifted range on another IQ test was **123.5** (Wechsler, 2003). The mean IQ score of 202 in the gifted validation sample of the SB5 was **124** (Roid, 2003). 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, 2003).

Given the fact that the validation studies with gifted children on the two major IQ tests yielded mean Full Scale IQ scores of 123.5 and 124, it seems apparent that the cut-off scores for gifted programs need to be adjusted. When the WISC-IV or SB5 are used for program selection, it would be wise to set the entrance criterion at **120** IQ, rather than the traditional 130 cut-off score (Falk, Moran & Silverman, 2003). Setting the cut-off score at 120 IQ takes into account the error of measurement. This practice would locate many children who would have scored 130 and above on earlier versions of the tests. In addition, the highest score attained at any age on any test may well be the best estimate of a gifted child's abilities. I am aware that this flies in the face of conventional wisdom in the field, which dictates that the latest IQ score should be used for eligibility purposes. However, as this section reveals, for the gifted, the newest norms do not necessarily generate the best estimates of ability.

The New Scoring Methods

To complicate matters further, we now have a plethora of methods for deriving IQ scores. Traditionally, gifted programs relied upon Full Scale IQ scores or composite scores for eligibility. The reliance on Full Scale scores will need to be revisited, given the research presented at this symposium. Until 2003, the Wechsler scales generated Verbal and Performance scores, in addition to the Full Scale IQ scores. Some states in the U.S. mandated that the Verbal *or* Performance IQ score on the Wechsler scales be used in the eligibility for gifted programs, rather than the Full Scale IQ score. Now that these composite scores are no longer being generated, the Verbal Comprehension Index (VCI) and Perceptual Reasoning Index (PRI) represent reasonable alternatives (Wechsler, 2003).

As the WISC-IV has included measures of Working Memory and Processing Speed in the calculation of Full Scale IQ scores, and these are not often strong suits for the gifted, Full Scale IQ scores are not as strong indicators of potential for success in gifted programs as the Verbal Comprehension and Perceptual Reasoning Indices. There is a new method of deriving a child's IQ score from these two indices alone. The resulting IQ score is called the General Ability Index (GAI). Authorities offer different criteria on when it is appropriate to derive the General Ability Index and which table to use. This will be discussed in more detail in Bobbie Gilman and Frank Falk's presentation. In any event, the VCI, PRI or GAI are all better predictors of success in gifted programs than the traditional Full Scale IQ score.

The SB5 also has several scoring methods. The table scores generally run lower than the WISC-IV scores. Rasch-ratio scores tend to run higher than SBL-M scores. Alternatively, it has been suggested that eliminating Working Memory provides a better estimate of the abilities of gifted children.

The new conceptions of intelligence are multi-faceted. There are many ways to be gifted. The WISC-IV, SB5, DAS and the nonverbal scales all recognize the importance of visual-spatial abilities. In the past, it was easier to find children with very advanced verbal abilities, and gifted program models were developed to serve these children. Now that we can identify children with advanced visual-spatial abilities, it will be necessary to adapt gifted programs to serve other types of giftedness.

Ceiling Effects

The perennial problem encountered in assessing the gifted is **ceiling effects**. Most people are unaware of the extent to which low test ceilings can depress IQ scores in the gifted range. Ceiling effects occur when the child's knowledge goes beyond the limits of the test. In order to assess the full strength of a gifted child's abilities, test items must be of sufficient difficulty. Imagine trying to measure a six-foot-person with a five-foot ruler (Stanley, 1990). The magnitude of the problem increases with age: the older the child, the more likely he or she is to outstrip the capacity of the measurement tool.

Ceilings vary on different types of tests. Group achievement tests and group ability tests have low ceilings. They are designed to compare students at a particular grade level, so they do not contain items well beyond grade level. For the purpose of these tests, it is enough to know that the child is at the 95th percentile. The 99.9th percentile is as good as it gets. Some individual IQ tests, such as the Raven's Progressive Matrices, also have low ceilings for a different reason. They contain more difficult items, but the norms of the instrument only extend to 135. The highest possible IQ score on the SB-IV appears to be 164, but 148 was the highest recorded score in the standardization sample (R. Thorndike, personal communication, May 8, 1985). The Wechsler scales appear to extend to 160, but the highest score we've seen on the WISC-III and WISC-IV was 155.

When gifted children attain scores on two different IQ tests that are highly discrepant, educators tend to believe the lower score is more accurate. I find this odd, because exactly the opposite perception occurs with developmentally delayed children. If such a child were to attain a

score of 50 on one IQ test and 65 on another, most people I've encountered would assume that the *higher* score is more accurate. Why? Because they can think of myriad reasons why a child might not have performed as well as possible on the test that produced the lower score. It is unlikely that a developmentally delayed child could attain an IQ score higher than his or her abilities. Wouldn't this same logic apply to scores for the gifted? One would hope so.

Educators often think the lower test scores are accurate measures of gifted children's abilities because they have no opportunity to observe what the students are actually capable of doing. Classrooms also have ceiling effects. Gifted children often know more than the teacher is teaching or classroom tests are testing and they have no chance to display their advanced knowledge. The antidote to ceiling effects is *opportunity to demonstrate advanced problem-solving abilities*.

The Talent Searches provide an excellent view of what happens when we remove ceiling effects. In Talent Search programs, middle school students who achieve at the 95th (or 97th) percentile in grade-level reading or mathematics achievement tests are allowed to take 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 anywhere between 200 (lowest score possible) and 800 (highest score possible) on the SAT-Math. 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.

Early Identification

Giftedness, like developmental delay, involves inherent differences in development from birth through maturity. Everyone agrees that we must identify developmentally delayed children as early as possible because it has been shown that *early intervention* is essential to optimal development. It is not as clearly understood that early intervention is also essential for the optimal functioning of developmentally advanced children.

In a study conducted by Gogel, McCumsey and Hewett (1985), nearly half of the 1,039 parents of identified gifted children suspected that their children were gifted before their toddlers were two years old. White and Watts (1973) noted that children who are either unusually rapid or unusually slow in their development show signs of their exceptionality as early as 18 months. Findings from the Fullerton longitudinal study confirmed those of White and Watts, and further indicated that "gifted and nongifted children develop at different levels from infancy through adolescence" (Gottfried, Gottfried, Bathurst & Guerin, 1994, p. 61).

Differences in level of intellectual performance between the gifted and nongifted children emerged on the psychometric testing at 1.5 years and maintained continuity thereafter. However, the earliest difference was found on receptive language skills at age 1 year. Differences in receptive and

expressive language skills were consistently found from infancy onward. (pp. 84-85)

Some educators believe that giftedness cannot be assessed before third grade, but this is due to budgetary constraints rather than to the limitations of testing. Gifted children can be assessed in a valid and reliable manner at the age of four. Gifted four- or five-year-olds are mentally like six- or seven-year-olds, and usually have excellent attention spans, so this is an ideal time for testing. Based on a half-century of her research in testing, Elizabeth Hagen, co-author of the *Cognitive Abilities Test* and the *Stanford-Binet Intelligence Scale, Revision IV*, revealed the following information in an interview:

I don't think four to six is too early to obtain a valid assessment. The correlations between scores obtained at ages four or five and later IQ scores are slightly lower than those obtained at age nine, but not that much lower. The only reservation I would have about testing at that age is being able to locate children who come from somewhat limited backgrounds. (quoted in Silverman, 1986, p. 170)

There is a widespread myth that IQ test scores of preschool and primary-aged children are inflated due to environmental advantage (e.g., parents reading to their children or the children attending excellent preschools). However, the impact of the environment **increases** with age; therefore, the IQ scores of third graders are unquestionably more influenced by the environment than the scores of kindergartners. For girls, in particular, early IQ scores are more reliable than those obtained after they have been socialized into hiding their abilities.

At the Gifted Development Center, we have found that the optimal time to test gifted children is *between the ages of four and nine*. We find that at the age of nine, test scores for gifted children may decline, sometimes as much as 20 points, due to (1) ceiling effects (test items not being sufficiently difficult to measure the full range of abilities); (2) perfectionism (particularly in girls), leading to unwillingness to guess when uncertain; and (3) the increased emphasis on crystallized (learned) knowledge and skills rather than fluid abilities (purer forms of abstract reasoning, considered innate). This does *not* mean that testing is useless after age nine. While the score generated may be an underestimate, we find that children and adults profit from even minimal estimates of their abilities.

Group IQ versus Individual IQ Tests

When psychometric evaluation is being considered, developmentally advanced children, like developmentally delayed children, should be assessed on individual intelligence tests by trained examiners. Group tests are rough *screening* tools only, like vision and hearing screening tests. They indicate the need for further testing by a specialist. Most school districts rely on group IQ tests for selecting students for gifted programs because individual IQ tests are substantially more expensive. However, even the best group IQ tests, such as the *Cognitive Abilities Test* (CogAT), were designed for screening purposes, as co-author Elizabeth Hagen points out:

Although I still believe you should use an individual intelligence test for assessing young children, I would use the *Cognitive Abilities Test* as a screening test to find out what the potential pool is and then use an individual test for final selection ... (Quoted in an interview

with Linda Silverman in *Roeper Review*, 1986, p. 170)

No parent of a disabled child would agree to have his or her child labeled on the basis of a group screening device, or grades, or a teacher's opinion. Yet, this is exactly what we often do with gifted children.

Individual IQ Tests Are Not All Alike

Individual IQ tests also present problems, since the scores they generate for gifted children are not comparable. The newer IQ scales are probably excellent for 95% of the population, but they are inadequate for assessing both the highly gifted and the profoundly delayed. Children in the highly (145-159 IQ), exceptionally (160 – 174 IQ) and profoundly gifted (175+ IQ) ranges have seriously depressed scores on the newer instruments. We continue to believe that the best measure of high levels of giftedness is the *Stanford-Binet Intelligence Scale, Form L-M* (SBL-M) (Silverman & Kearney, 1989; 1992a; 1992b). Since it goes up to Superior Adult III, the SBL-M acts as an above-level measure, similar to the SAT for Talent Search participants. In the words of Julian Stanley (1990), founder of the Talent Searches, “The Binet-type age scale might be considered the original examination suitable for extensive out-of-level testing” (p. 167).

The SBL-M assesses high-level verbal abstract reasoning, as well as mathematical and spatial reasoning; it has very few timed items; and few items require visual-motor abilities. It attracts the attention of younger children better than the Wechsler tests because it moves rapidly from one type of item to another (Vernon, 1987). As ratio-based scoring is used to derive the formula IQ scores beyond the norms of the manual, a greater spread of scores is possible than with the use of deviation-based scores. Therefore, it allows exceptionally and profoundly gifted children to be differentiated from moderately gifted children. Given that the SBL-M is dated, we now assess children first on one of the newer instruments, such as the *Wechsler Intelligence Scale for Children, Fourth Edition* (WISC-IV) or the *Stanford-Binet Intelligence Scale, Fifth Edition* (SB5). Similar to the use of the SAT in the Talent Searches, we use the SBL-M as a supplemental test if the child attains at least two subtest scores at the 99th percentile.

The strongest objection to the use of the SBL-M is its outdated norms. As mentioned earlier, the “Flynn Effect” does not apply in the extremes, and, even if it did, it is estimated to be approximately one-third of an IQ point per year. In a study of 121 cases in four IQ ranges, selected randomly from the Gifted Development Center files, Falk, Moran and Silverman (2003) found that the discrepancy between WISC-III and SBL-M scores was significant well beyond what could be accounted for by the Flynn Effect.

As of March, 2005, the Gifted Development Center had assessed 4,700 clients, and found 882 children with IQ scores on the SBL-M in the exceptionally gifted range (160+) where the other tests abruptly end. Ratio-based formula IQ scores were derived as prescribed on p. 339 of the SBL-M Manual (Terman & Merrill, 1973). In the highest IQ ranges, 218 scored beyond 180 IQ and 56 scored beyond 200 IQ. These children would all obtain scores below 155 on other instruments.

The gender distribution of our sample is particularly noteworthy. Approximately 60% of the

children brought for assessment are male, and 40% female. In the exceptionally gifted range, we find the same gender ratio as the population referred for testing, and the two *highest scoring children were female*. Our sample demonstrates that there are as many brilliant females as males—at least in terms of native ability. If this information were widely known, it would help dispel the 5,000-year-old myth of the “natural inferiority” of women. Using only the newer tests, we would be unable to locate these girls and develop their extraordinary talents. They would quickly go underground and remain there.

Researchers who have studied exceptionally and profoundly gifted children have documented their remarkably different thought processes (e.g., Gross, 1993; Morelock, 1995; Lovecky, 1994). In *Nature's Gambit*, David Feldman and Lyn Goldsmith (1986) describe an “omnibus prodigy” with the highest IQ score on record, whose inexhaustible energy wore out everyone around him, including his parents. Anyone who has studied or lived with a child in these highest IQ ranges can attest to their immense curiosity, their intense absorption in learning, their great intuitive leaps, their profound ethical concerns, their endless energy and their deep sense of social isolation. They are, indeed, the proverbial “horse of a different color.”

One child we tested attained a WISC-III Full Scale IQ score of 138 and an SBL-M score of 223+, a difference of more than 85 points. He graduated high school at 14. A six-year-old tested 155 Full Scale IQ on the WISC-III and 244 on the SBL-M, a difference of 89 points. She completed grades 2 through 7 in two years. Another child achieved a *Wechsler Preschool and Primary Scale of Intelligence* (WPPSI) Full Scale IQ score of 145 and an SBL-M score of 236+, more than a 91-point discrepancy. At the age of eight, he achieved 760 on the SAT-Math test. He could perform several mental operations simultaneously. A seven-year-old exceeded the raw points necessary to attain ceiling scores on six subtests of the WISC-III, generating a Full Scale IQ score of 155. She passed one third of the items at the highest level of the SBL-M. Her IQ score was 262+ and the difference between her SBL-M and WISC-III scores was **more than 107 points!** Her math mentor considered her the most brilliant mathematical mind of our time.

A girl tested by an Eastern examiner scored 124 on the *Kaufman Assessment Battery for Children* (K-ABC), 137 on the WISC-R, and 229+ on the SBL-M. The difference between her K-ABC score and her SBL-M score was **over 105 points!** A child prodigy in writing, her literary gifts confirmed the accuracy of the SBL-M score. Camilla Benbow tested a boy who scored 199 on the SBL-M at the age of seven 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).

The SBL-M remains the only tool that can measure extreme verbal abilities. Unfortunately, due to its age, this valuable instrument may be lost as a means of discovering society's most brilliant minds. What would happen to these children if we relied only on the lower estimates supplied by current tests? Most would be misunderstood, due to their inability to relate to age-peers and age-normed curriculum. Some would be misdiagnosed and placed on medication. Others would languish in grade-level placements, when they desperately need radical acceleration. And a few

would sink into lifelong depression. There would be no way of documenting the extent of their differences and supporting their need for tremendously advanced work. If we had no way of knowing the actual level of their abilities, we would be unable to find them true peers—intellectual equals. If their true abilities were unrecognized and undeveloped, they would be likely to develop patterns of underachievement. There is a higher than expected number of gifted students among dropouts (Seeley, 1998). Motivation and scholarship depend on recognition. It would be debilitating to these individuals, to their families and to our scientific understanding of intelligence, to lose the only tool we have for measuring the highest levels of potential.

Conclusion

Any test can only measure a small portion of a person's competence. Therefore, all tests underestimate children's abilities rather than overestimate them. It is nearly impossible to fake abstract reasoning at an advanced level. When a disabled child achieves two different IQ scores, the higher score is believed to be the best estimate of the child's potential. Gifted children deserve the same attitude.

When an individual IQ test is used for the selection of students for a gifted program, it is recommended that the cut-off score be lowered to 120. The highest index or factor score is the best predictor of success in a gifted program. When a child tops out on a test with a low ceiling, achieving two or more subtests at the 99th percentile, it is recommended that a test with a higher ceiling, such as the SBL-M, be administered as a supplemental test.

Several agencies have found an astonishing number of exceptionally gifted children. Terman (1925) and many other researchers noted that there were more children above 160 in the population than the normal curve would predict (Silverman, 1989). If we are to serve them properly, it behooves us to find them. The adjustment problems of a misdiagnosed child whose actual IQ is 180 are staggering. The further a child is from the norm, the greater the potential for suffering alienation and the greater the need for detection and early intervention.

Nonverbal tests need to be considered if we are to attain equity in the selection of students for gifted programs. Visual-spatial learners are also more likely to do well on nonverbal assessments. Success in traditional gifted programs can often be predicted by nonverbal tests in conjunction with a vocabulary test or other verbal measure. However, as gifted programs start to recognize and serve more children with right hemispheric gifts, nonverbal tests will be sufficient to locate them.

Qualitative Assessment is likely to be the wave of the future. While it can be used in conjunction with IQ tests, it also stands alone as a valid means of identifying gifted children. In addition, it provides a window into the Soul of the child, generating much more valuable information than psychometric assessments and structured interviews. All of these methods need to be carefully considered in selecting students for gifted programs.

References

- Bryant, D. C. (1992). *A comparison of Wechsler Intelligence Scale for Children-Revised (WISC-R) and Wechsler Intelligence Scale for Children-Third Edition (WISC-III) for Gifted Children*. Unpublished doctoral dissertation, The West Virginia Graduate College, Charleston, WV.
- Falk, R. F., Moran, D., & Silverman, L. K. (2003, November). *WISC-III and Stanford-Binet L-M Scores for gifted children*. Paper presented at the 50th annual convention of the National Association for Gifted Children, Indianapolis, IN.
- 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.
- Feldman, D. H., with L. T. Goldsmith. (1986). *Nature's gambit: Child prodigies and the development of human potential*. New York: Basic Books.
- Flanagan, D. P., & Kaufman, A. S. (2004). *Essentials of WISC-IV assessment*. Hoboken, NJ: John Wiley & Sons.
- Flynn, J. R. (1999). Searching for justice: The discovery of IQ gains over time. *American Psychologist*, 54, 5-20.
- Gogel, E. M., McCumsey, J., & Hewett, G. (1985). What parents are saying. *G/C/T*, Issue Number 41, 7-9.
- Gottfried, A. W., Gottfried, A. E., Bathurst, K., & Guerin, D. W. (1994). *Gifted IQ: Early developmental aspects. The Fullerton longitudinal study*. New York: Plenum.
- Gross, M.U.M. (1993). *Exceptionally gifted children*. London: Rutledge.
- Jensen, A. R. (1980). *Bias in mental testing*. New York: The Free Press.
- Kaufman, A. S. (1992). Evaluation of the WISC-III and WPPSI-R for gifted children. *Roeper Review*, 14, 154-158.
- Keith, T. Z., Fine, J. G., Taub, G. E., Reynolds, M. R., & Kranzler, J. H. (2004). *Hierarchical multi-sample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children-Fourth Edition: What does it measure?* (Manuscript submitted for publication)
- Lovecky, D. V. (1994). Exceptionally gifted children: Different minds. *Roeper Review*, 17, 116-120.
- Morelock, M. J. (1995). *The profoundly gifted child in family context*. Unpublished doctoral dissertation, Tufts University, Medford, MA.
- Roid, G. H. (2003). *Stanford-Binet Intelligence Scales, Fifth Edition, Technical Manual*. Itasca, IL: Riverside.
- Seeley, K. R. (1998). Underachieving and talented learners with disabilities. In J. VanTassel-Baska (Ed.), *Excellence in educating gifted & talented learners* (3rd ed., pp. 83-93). Denver: Love.
- Silverman, L. K. (1986). An interview with Elizabeth Hagen: Giftedness, intelligence and the new Stanford-Binet. *Roeper Review*, 8, 168-171.
- Silverman, L. K. (1989, November). *Lost: One IQ point per year for the gifted*. Paper presented at the 36th annual convention of the National Association for Gifted Children, Research and Evaluation Division, Cincinnati, OH. (Available from www.gifteddevelopment.com).
- Silverman, L. K. (1993). The gifted individual. In L. K. Silverman (Ed.), *Counseling the gifted & talented* (pp. 3-28). Denver: Love.
- Silverman, L. K., & Kearney, K. (1989). Parents of the extraordinarily gifted. *Advanced Development*,

1, 41-56.

- Silverman, L. K., & Kearney, K. (1992a). The case for the Stanford-Binet L-M as a supplemental test. *Roeper Review*, 15, 34-37.
- Silverman, L. K., & Kearney, K. (1992b, November). *Don't throw away the old Binet*. Presented at the 39th annual convention of the National Association for Gifted Children, Los Angeles, CA. [Appeared in part in *Understanding Our Gifted*, 4(4), 1, 8-10.]
- Stanley, J. C. (1990). Leta Hollingworth's contributions to above-level testing of the gifted. *Roeper Review*, 12(3), 166-171.
- 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. (1973). *The Stanford-Binet Intelligence Scale: Manual for the Third Revision, Form L-M*. Boston: Houghton Mifflin.
- Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *The Stanford-Binet Intelligence Scale: Fourth edition. Technical manual*. New York: Riverside.
- Vernon, P. E. The demise of the Stanford-Binet scale. *Canadian Psychology/Psychologie Canadienne*, 28(3), 251-258.
- Wechsler, D. (2003). *The WISC-IV technical and interpretive manual*. San Antonio, TX: Psychological Corporation.
- White, B. L., & Watts, J. C. (1973). *Experience and environment* (Vol. 1). Englewood Cliffs, NJ: Prentice-Hall.

BIO: Linda Kreger Silverman, Ph.D., is a licensed psychologist and Director of The Gifted Development Center and the Institute for the Study of Advanced Development, in Denver, Colorado. Founder and consulting editor of the first journal on adult giftedness, *Advanced Development*, she also edited the popular textbook, *Counseling the Gifted and Talented* (Love, 1993), adopted at over 50 colleges and universities. Her latest book is *Upside-Down Brilliance: The Visual-Spatial Learner* (DeLeon, 2002). For nine years she served on the faculty of the University of Denver. The Gifted Development Center began in June, 1979, and has assessed more than 4,700 children from all over the globe.