THE CASE FOR THE STANFORD-BINET L-M AS A SUPPLEMENTAL TEST

Linda Kreger Silverman, Ph.D.
Katheryn Kearney, M.A. Ed.

Linda K. Silverman, Ph.D., is a licensed psychologist, Director of the Gifted Child Development Center in Denver, Colorado, Editor of *Understanding Our Gifted* and *Advanced Development,* and a member of Editorial Advisory Board of Roeper Review. **Katheryn Kearney,** M.A. Ed., is Founder of the Hollingworth Center for Highly Gifted Children in South Casco, Maine, Editor of Highly Gifted Children, a doctoral candidate at Teachers College, Columbia University, and a Contributing Editor of Roeper Review.

Once upon a time when people were about 5 feet tall (as can be seen by examining suits of armor), bathtubs were almost 6 feet long (remember the ones with the feet?). Today, many men and women are 6 feet tall, so why is it that we now have 4 1/2 foot bathtubs? How did we manage to exchange those wonderful old tubs for the confining modern versions that have no room for tall people? Probably for the same reasons that we chose to lop off the top and bottom ends of the continuum in modern IQ tests. It is more economical to mass produce and sell a product that fits the mean than it is to cater to the extremes. That is the sell out that has occurred in intelligence testing which requires us to retain an antiquated model—a tub with feet. Time marches on, but it is clear that newer is not always better.

The population is apparently gaining in intelligence (Flynn, 1987), but modern intelligence tests have less power to measure the high end of that intelligence. The original Stanford-Binet Intelligence Scale was structured to measure development from very early childhood to adulthood. It has a higher ceiling (and lower floor) than any of the current tests, including its successor, the Stanford-Binet IV. And the most important factor in assessing gifted students is the height of the ceiling. As Julian Stanley (1990) indicates, you can't get an accurate assessment on individuals 6 or 7 feet tall with a 5 foot ruler. But that is all we have today—a number of 5 foot rulers with items of insufficient difficulty to discriminate within the gifted range. Elizabeth Hagen, one of the constructors of the Stanford-Binet IV, explains why:

> *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)*

Newer tests, like the Binet-IV, WPPSI-R or WISC-III, are inadequate for highly gifted children, because when an item can be solved "only" by children enrolled in a gifted class, it is removed from the test. The gifted population was simply not a priority of the test constructors (Silverman & Kearney, 1992).

New Norms Discriminate Against the Gifted

One of the main reasons Nancy Robinson has provided for replacing the old Binet with the new one is the "obvious superiority of more recent norms." This is also the major concern of most school psychologists—even those who continue to use the WISC-R with norms only two years newer than those of the Binet L-M. Flynn (1984, 1987) has demonstrated that the general population is increasing in intelligence at the rate of approximately one-third of an IQ point per year (other researchers estimate one-fourth of an IQ point). As summarized in our companion article, "Every couple of decades, the norms have to be adjusted about one-half a standard deviation." We have no quarrel with this. If the norms in the gifted range were adjusted 8 points every 20 years, that would be fine. But that is not the case. A careful inspection of the 1960, 1972 and 1986 norms on the Binet scales reveals that the scores have been adjusted at least 26 points in those 26 years (depending on age and range of IQ), not just 8 or 10 (Silverman, 1989). This is blatantly unfair, and cannot be attributed solely to the increase in general intelligence.

What happened to that entire standard deviation of intelligence beyond the 10 point anticipated increment? It seems to have been lost as a result of forcefitting a trimodal distribution into the normal curve—a unimodal distribution (Jensen, 1980;

Laycock, 1979). Deviation IQs, generated by all modern tests, rest on the assumption that intelligence is normally distributed according to the Gaussian curve. However, several investigators have discovered unexpectedly frequent scores at the upper end of the IQ distribution that parallel the higher frequencies found at the lower end of the continuum (Dunlap, 1967; Gallagher & Moss, 1963; Jensen, 1980; McGuffog, Feiring, & Lewis, 1987; Robinson, 1981; Stott & Ball, 1965; Terman, 1925). In large-scale studies in several countries, J. C. Raven (1959) and J. Raven (1983) found distributions with the Progressive Matrices Intelligence Test which sharply depart from the Gaussian curve.

> *J. C. Raven (1959)...argued that it might be the assumption about what reality should be like, and not the obtained distributions of test scores, which was wrong. For this reason he resisted the use of deviation IQs as the means of expressing test scores. (J. Raven, 1983, p. 174)*

Even Wechsler (1944) noted "the wide-spread but mistaken belief that measures [of IQ] distribute themselves according to the normal curve" (p. 126). According to Jensen (1980), the normal curve theoretically and empirically fits the distribution of scores for 98 percent of the population, from 60 to 150 IQ, but does not apply at the extremes.

In pressing scores at the extremes to fit the normal curve, each new set of norms discriminates against highly gifted students. Even those in the moderately gifted range are not immune from this discrimination. The technical manual of the Fourth Edition of the Stanford-Binet (Thorndike, Hagen & Sattler, 1986) reports only two studies of the gifted: one with a sample of 82 who took both the L-M and the Fourth Edition, and one with only 19 students who took the Fourth Edition and the WISC-R. In the first sample, the mean score on Form L-M was 135.3, whereas the composite score for the same group on the Fourth Edition was 121.8 (p. 70). This was a difference of 14 points, almost a full standard deviation. For children in the average range of ability who took both tests, the change in score between the L-M and the Fourth Edition was only 2 or 3 points. Similar discrepancies between the two tests were found by Robinson, Dale and Landesman (1990) in a sample of 26 children 30 months of age (137.7 on the L-M vs. 124.7 on the Binet IV).

In the second "gifted" sample reported in the technical manual, the Fourth Edition yielded a mean composite score of 116.3 compared to a mean WISC-R Full Scale score of 117.7 (Thorndike, Sattler & Hagen, 1986, p. 72). Ironically, the "very superior" (gifted) range is defined in the Stanford-Binet IV manual as 132 IQ, even though the only two samples of "gifted" students had mean composite scores of 116 and 122!! The logical inconsistency went unnoted, and 132 was still maintained as the entry level for many gifted programs.

The new WISC-III and WPPSI-R suffer from the same myopia. The 38 children who had been identified as gifted on other measures achieved a mean Verbal IQ (V) of 128, mean Performance IQ (P) of 125, and mean Full Scale IQ (FS) score of 129 on the WISC-III (Wechsler, 1991), low enough to just miss the cut-off score for most gifted programs! On the WPPSI-R, the formerly gifted are now achieving V and FS means of 122 and P of 115 (Wechsler, 1989). Since the Stanford-Binet IV and the WISC-R yield comparable results in the gifted range (116 vs. 117), and WISC-III scores are at least 5 points lower than the WISC-R for children in the gifted range, we can estimate that in the last 31 years, there has been a mysterious loss of 31 IQ points for moderately gifted students, only a third of which can be attributed to the increase in general intelligence of the American population.

New Conceptions of Giftedness

Other major concerns expressed about the Stanford-Binet L-M are its lack of factorial structure, reliance on a global score and a presumed lack of ethnic representation in the normative sample. Factor analytic studies of the Binet L-M are reported in the manual, along with the socio-economic and ethnic makeup of the 1972 normative sample (Terman & Merrill, 1973). (Lack of ethnicity in the sample only applied to the 1960 norms.) Item analyses can generate diagnostic information about strengths and weaknesses comparable to that obtained from subtest and factor scores; see, for example, Valett's (1964) analysis and Sattler's (1974) "Binetgram" (p. 135).

While the newer conceptions of intelligence appear to be fairer to children with diverse abilities, most modern intelligence tests require the child to be good in all domains in order to obtain a score in the gifted range. The Binet IV purports to avoid this problem by allowing practitioners to select parts of the test, but in practice, composite IQ scores are typically used for placement purposes. And there is some question about the viability of each practitioner selecting the appropriate parts of the test to administer.

> *For example, a school psychologist might obtain an overall IQ of 125 with his or her selected battery, and recommend against accepting a pupil for a gifted program. The parents might consult a private psychologist whose battery gave an IQ of 135 and who would recommend acceptance as gifted. (Vernon, 1987, p. 252)*

The Binet L-M allows a student with strengths in one or two domains to proceed to the Superior Adult III level on the basis of those strengths alone. Therefore, a child with very strong verbal abstract reasoning abilities, who has difficulty with short term memory or motor speed or spatial abilities, can demonstrate the full strength of those abilities. On the newer tests, overall performance (e.g., a composite score) is depressed by any one of these factors, while the amount of verbal reasoning that can be demonstrated is limited by lack of items of sufficient complexity. "The L-M provides a richer sampling of verbal abilities. In the NS [Binet IV] there are no rhymes, analogies, sentence completions, ideational fluency, interpreting proverbs, comprehension of paragraphs, or verbal reasoning problems" (Vernon, 1987, p. 255). In a comparative study of highly gifted children (N = 21), the 5 students with ratio IQs ranging from 182 to 193 on the Stanford-Binet L-M attained mean scores on the WISC-III of 143 V, 127 P and 139 FS—a loss of over 46 IQ points, making them indistinguishable from moderately gifted children (Silverman & Atkinson, in preparation).

It is curious that we are moving away from a focus on abstract verbal reasoning abilities in the measurement of intelligence in favor of greater emphasis on visual-spatial abilities and speed of performance.

> *The biggest negatives for gifted assessment are the new emphasis on problem-solving speed on the WPPSI-R [and] the substantially increased stress on performance time in the WISC-III compared to the WISC-R... The speed factor will penalize gifted children who are as reflective as they are bright, or who*

tend to go slow for other non-cognitive reasons such as a mild coordination problem. (Kaufman, 1992, p. 158)

These new emphases also add barriers to finding gifted girls. On the original Stanford-Binet Intelligence Scale, an untimed test of verbal abstract reasoning abilities, girls surpassed the boys at every age level until adolescence (Terman, 1916). But from the time Wechsler first studied men who could not qualify on the Army Alpha to serve in the armed forces (Matarazzo, 1981), we have moved steadily toward more spatial conceptions of intelligence. "Intelligence tests that deemphasize verbal skills and emphasize performance, that is, spatial-visual activities, may...be biased against girls who receive much less practice than boys at puzzles and assembly" (Kerr, 1991, p. 408). With all the efforts on behalf of gender equity over the last 75 years, we seem to have moved toward more sexist methods of measuring intelligence. New theories of intelligence offer no solution to these problems as they have "little or nothing to say about gender" (Kerr, 1991, p. 402).

An alternative conception of giftedness provides a rationale for determining just how different gifted children are from the norm.

*Giftedness is **Asynchronous Development** in which advanced cognitive abilities and heightened intensity combine to create inner experiences and awareness that are qualitatively different from the norm. This asynchrony increases with higher intellectual capacity. The uniqueness of the gifted renders them particularly vulnerable and requires modifications in parenting, teaching and counseling in order for them to develop optimally. (The Columbus Group, 1991)*

From this definition, it can be seen that the degree of difference in children's verbal abstract reasoning abilities has major ramifications in terms of their experience and ability to fit in with society. Hal Robinson (1981) points out 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). Without the tools to find such children, the children themselves remain doubly at risk. Children of 180 IQ have qualitatively different experiences from children of 140 IQ (Hollingworth, 1942), and they deserve the right to be identified on instruments that can differentiate between 140 and 180. Only the Stanford-Binet (Form L-M) can do this; discarding it increases the chances that these children will be misunderstood and that their needs will not be met.

It is safe to say that if any other special population of gifted children (or any other group of children, for that matter) was at risk in similar ways, we would use whatever effective tools were available in order to identify them and provide appropriate services for them. For this particular population, an older tool may well be more effective than newer ones (Silverman & Kearney, 1989).

Strengths of the Binet L-M

The Stanford-Binet L-M is the only instrument designed to differentiate highly gifted from moderately gifted children; it has a higher ceiling than other tests and serves as the prototype for above level tests (Stanley, 1990); it is *untimed*, which makes it a fairer assessment of children with motor delays, reflective children, gifted children with learning disabilities, and gifted girls; it is more engaging for preschoolers (Canter, 1990; Vernon, 1987); its emphasis on verbal abilities enables the identification of more African-American gifted children than group tests (Kearney & LeBlanc, in press) or the so-called "culture-fair" nonverbal tests (Ehrlich, 1978); it is an excellent assessment of abstract reasoning—the *sine qua non* of giftedness (Snyderman & Rothman, 1988); it provides a mental age which can be used to derive deviation and ratio IQ scores as well as to help parents and teachers understand why these children prefer older playmates and need advanced coursework; it has a strong research base, high predictive validity, and a long history of successful use with both retarded and gifted populations (Wilson, 1992); and it has a lower floor so that it can better differentiate the developmental levels of preschool and retarded children (Canter, 1990; Glutting, 1989; Wilson, 1992). If the old Binet is preserved, revamped, and renormed, it will most likely be due to the pressure of advocates and professionals concerned with the developmentally disabled, since all of the newer tests have too high a floor as well as too low a ceiling. Advocates for the gifted should join forces with advocates of the disabled to preserve the L-M.

Although the Stanford-Binet (L-M) has received a good share of criticism over the years, it has remained the most stable instrument for predicting school success and assessing intellectual giftedness. The success of the test appears to have been due to Terman's skill at 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). In an article describing the endurance of the Stanford-Binet, Thorndike (1975) wrote: "[It] has been for the past 60 years the workhorse of psychometric appraisal of cognitive development, the standard against which other tests of cognitive ability have been evaluated" (p. 4).

After a careful analysis of the Stanford-Binet IV, Philip Vernon (1987) makes the following set of recommendations:

I believe that a strong case can still be made for retaining the L-M, with its apparently haphazard arrangement of items... children below about 6 years have great difficulty with WPPSI and WISC in maintaining the same set throughout all the items in a particular subtest. In contrast, the shortness of the Binet items and their great variations in content help the tester to catch and hold the child's attention. (p. 253) There are two special groups for whom the L-M is often preferable to the Wechsler scale: The potentially gifted who are being considered for special classes or enrichment programs, and severely retarded...children or adults. Neither the four verbal subtests in WISC or WAIS nor the four NS [Binet IV] verbal subtests give as much opportunity as the L-M for gifted children to display their fluency, imagination, unusual or advanced concepts, and complex linguistic usage. (p. 256)Psychologists who wish to continue using the third edition (Form L-M) with 2 - 6 year-olds, or with likely gifted children, should do so (p. 257).

In case psychologists are concerned that it might be somehow "unethical" to use a dated test, Stanley E. Jones, Director of the Office of Ethics for the American Psychological Association, writes:

It would not be my reading that Principle 2.9 would prohibit the use of any test for a purpose that can be defended. It does make it the responsibility of the psychologist to provide such a defense when using tests

which are not obviously current. (personal communication to Sylvia Rimm, November 25, 1991)

Recommendations for the Continued Use of the Binet L-M

Given the difficulties with current instruments, it is recommended that (1) entrance requirements for gifted programs be lowered to 120 to take into account the deflated scores for gifted students on newer norms; (2) the Stanford-Binet Form L-M should be used to assess preschool-aged and developmentally disabled children; (3) children should be retested on the Stanford-Binet L-M when they obtain ceiling scores on any current instrument (e.g., 17, 18 or 19 on 3 or more subtests on the WPPSI-R, WISC-R or WISC-III; one factor score above 148 on the Stanford-Binet IV*); (4) children who have difficulties with motor speed or processing speed should be allowed to take the Stanford-Binet L-M, which is essentially an untimed test; (5) if the old norms pose a problem, scores on the L-M could be adjusted by 8-10 points to take into account increases in general intelligence reflected in newer norms.

The Stanford-Binet L-M is strongly recommended as a supplemental test to obtain further information about highly gifted children, and to tie that information to the 70-year research history regarding the extraordinarily gifted, which used this test and its predecessors extensively for identification. Using standard formulas derived from 1960 norms (Pinneau, 1961), scores should be extrapolated for any child who scores beyond the norms in the manual, in order to obtain a rough estimate of the child's ability. Since a number of highly gifted children have dramatic weaknesses that may artificially depress IQ scores, parents should request administration of the L-M as a supplemental test whenever they suspect that the newer assessments have underestimated their children's abilities.

Finally, it is essential that researchers do what the test constructors and publishers did not do. It is 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" (Robinson, 1981, p. 73). Nor, apparently, is this simply a result of using older norms rather than newer ones, since Terman noted this phenomenon with his original test as early as 1925, finding many more children above 170 IQ in his longitudinal study than statistically should have been there. It is absolutely essential that researchers test a substantial number of children who score above 160 or 170 IQ on the Stanford-Binet Form L-M with both the Stanford-Binet Fourth Edition and the WISC-III. Then, rather than comparing total IQ scores, the subtest score patterns should be analyzed to see whether or not typical "highly gifted" score patterns exist which would make identification of extremely gifted children possible using the newer tests.

Both the old Binet and the new scale are needed for different purposes. The Stanford-Binet IV is not a revision of the old scale; it is essentially a new test (Keith, Cool, Novak, White & Pottebaum, 1988; Rothlisburg, 1987; Thorndike, 1990). The two tests may not be comparable. They may be useful for different populations at different times for different reasons and they may be testing different things. Each has its place and both scales should be revamped and renormed.

REFERENCES

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

Columbus Group (1991, July). Unpublished transcript of the meeting of the Columbus Group. Columbus, OH.

Dunlap, J. M. (1967). The education of children with high mental ability. In W. M. Cruickshank & G. O. Johnson (Eds.), *Education of exceptional children and youth* (2nd ed.) (pp. 143-193). Englewood Cliffs, NJ: Prentice-Hall.

Ehrlich, V. Z. (1978). *The Astor Program for gifted children: Prekindergarten through grade three*. New York: Teachers College, Columbia University.

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

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101*, 171-191.

Gallagher, J., & Moss, J. (1963). New concepts of intelligence and their effect on exceptional children. *Exceptional Children, 30*(1), 1-5.

Glutting, J. (1989). Introduction to the structure and application of the Stanford-Binet Intelligence Scale—Fourth Edition. *Journal of School Psychology, 27*, 69-80.

Hollingworth, L. S. (1942). *Children above 180 IQ Stanford-Binet: Origin and development*. Yonkers-on-Hudson, NY: World Book.

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.

Kearney, K., & LeBlanc, J. (in press). Forgotten pioneers in the study of gifted African-Americans, *Roeper Review*.

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

Kerr, B. A. (1991). *A handbook for counseling the gifted and talented*. Alexandria, VA: American Association for Counseling and Development.

Laycock, F. (1979). *Gifted children*. Glenview, IL: Scott, Foresman.

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

McGuffog, C., Feiring, C., & Lewis, M. (1987). The diverse profile of the extremely gifted child. *Roeper Review, 10*(2), 82-88.

Pinneau, S. R. (1961). *Changes in intelligence quotient: Infancy to maturity*. Boston: Houghton Mifflin.

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

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

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.

Robinson, N., Dale, P. S., & Landesman, S. (1990). Validity of Stanford-Binet IV with linguistically precocious toddler. *Intelligence, 14*, 173-186.

Rothlisburg, B. A. (1987). Comparing the Stanford-Binet, Fourth Edition, to the WISC-R: A concurrent validity study. *The Journal of School Psychology, 25*, 193-196.

Sattler, J. M. (1974). *Assessment of children's intelligence*. Philadelphia: W. B. Saunders.

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

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, October). *Lost: One IQ point per year for the gifted*. Paper presented at the National Association for Gifted Children 36th Annual Convention, Cincinnati, OH.

Silverman, L. K., & Atkinson, D. R. (in preparation). Comparisons of the highly gifted on the WISC-III and the Stanford-Binet Form L-M.

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

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

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

Stanley, J. C. (1990). Leta Hollingworth's contributions to above-level testing of the gifted *Roeper Review, 13*, 166-171.

Stott, L., & Ball, R. (1965). Infant and preschool mental tests: Review and evaluation. *Monograph of the Society for Research in Child Development, 30*, 4-42.

Terman, L. M. (1916). *The measurement of intelligence*. Boston: Houghton Mifflin.

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. (1975). Mr. Binet's test 70 years later. *Educational Researcher, 4*, 3-4.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *The Stanford-Binet Intelligence Scale: Fourth edition. Technical manual*. New York: Riverside Publishing.

Thorndike, R. L., with Lohman, D. F. (1990). *A century of ability testing*. Chicago: Riverside Publishing.

Valett, R. E. (1964). A clinical profile for the Stanford-Binet. *Journal of School Psychology, 2*, 49-54.

Vernon, P. E. (1987). The demise of the Stanford-Binet scale. *Canadian Psychology/Psychologie Canadienne, 28*(3), 251-258.

Wechsler, D. (1944). *Measurement of adult intelligence* (3rd ed.). Baltimore: Williams & Wilkins.

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 III. Manual*. San Antonio, TX: The Psychological Corporation.

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.

*Note: Professional judgment will be needed to determine ceiling level scores on the Stanford-Binet Fourth Edition since these vary by age. Scores above 148 on the Fourth Edition are of questionable validity since they were derived through statistical extrapolation due to lack of representation of highly gifted children in the normative sample (R. L. Thorndike, personal communication, October, 1985).