Educational Research (ISSN: 2141-5161) Vol. 5(2) pp. 42-51, February, 2014
DOI: http:/dx.doi.org/10.14303/er.2011.481
Available online@ http://www.interesjournals.org/ER

Full Length Research Paper

Scaling Three Versions of the Stanford-Binet Intelligence Test: Examining Ceiling Effects for Identifying Giftedness

***Tandra Tyler-Wood[1], Gerald Knezek[1], Rhonda Christensen[1], Cesareo Morales[1], Peter Dunn-Rankin[2]**

[1]University of North Texas, [2]University of Hawaii

Abstract

Three versions of the Stanford-Binet Intelligence Scale were administered to 42 children aged 8-14 years who had been referred to the researchers as gifted. Twenty-five subjects were retained in the analysis pool by virtue of scoring 118 or higher on one of the three versions of the Stanford-Binet. Researchers analyzed the scores on the three test versions to better understand the nature of the differences. The L-M version was found to have a greater range (higher ceiling) than the Fourth Edition or the Fifth Edition of the Stanford-Binet. The children averaged 136.84 (SD = 15.68) on the L-M version, 129.52 (SD = 9.73) on the Fourth Edition, and 124.24 (SD = 9.66) on the Fifth Edition. A re-norming process is recommended to establish appropriate criteria for gifted placement.

Keywords: Stanford-Binet Intelligence Scale, giftedness, ceiling effects.

INTRODUCTION

“The *Stanford-Binet Intelligence Scale* has a long and rich tradition, which began in 1916 when Lewis M. Terman completed his American revision of the 1908 Binet-Simon Scale. At that time it was called the Stanford Revision of the Binet-Simon Scale. Through various editions in 1937, 1960, and 1986, the Stanford-Binet has become widely known as a standard measure for intellectual abilities” (Roid, 2003a, p. 5). Although the Stanford-Binet is a widely-accepted standard, concerns have been voiced over the past two decades that newer versions have ceilings too low to allow exceptional children to earn high scores (Robinson, Dale AND Landesman,1990;Tyler-Wood AND Carri, 1990; Silverman, L.K, 1989) (Silverman AND Kinney, 1992). Examination of this ceiling effect across the L-M (1972), Fourth (1986) and Fifth (1998) Editions is the primary purpose of this paper. Brief descriptions of each version follow.

Stanford-Binet, Form L-M

The L-M version of the Stanford-Binet (Terman AND Merrill, 1972) is a wide-range individual test assessing intelligence from age two through the superior adult level. It requires subjects to solve problems, give definitions, memorize new material, and use some visual-motor skills at various age levels criteria. The test yields one comprehensive IQ score with no subtest scores provided. The L-M provides an array of age related test items from the Age II Level to the Superior Adult III Level. Although a vocabulary test is included at every other age level, there is no consistency among age levels as to the type of items that are presented. For example, a picture completion item at the Year V Level provides a partially completed stick figure for the examinee to complete. This type of item is not represented at any other age level on the test.

The examiner must establish a basal age for the child and continue testing until a ceiling is reached. Test items are grouped according to age level (not subject area) allowing an individual access to higher items even when easier items may be missed. Individuals must miss all items administered at an age level to meet the criteria for test cessation. For each correct item answered, a specified number of months is added to obtain a mental age score. Examiners obtain an IQ by locating the examinees chronological age compared to mental age in a table. The table to determine IQ when using the SB-LM was last revised in 1972.

Stanford-Binet Intelligence Scale, Fourth Edition

The Fourth Edition of the Stanford-Binet, released in 1986, was designed with a larger, more diverse,

*Corresponding Author E-mail: Tandra.Tyler-Wood@unt.edu

representative sample to minimize gender and racial inequities in earlier versions.

All test subjects take an initial vocabulary test, which along with the subject's age, determines the number and level of subtests to be administered. The SB-IV assesses subject areas such as Vocabulary, Comprehension, Absurdities, Verbal Relations, Pattern Analysis, Copying, Matrices, Paper Folding and Cutting, Quantitative, Bead Memory, Memory for Sentences, Memory for Digits and Memory for Objects. Total testing time is 45-90 minutes, depending on the subject's age and the number of subtests given. Raw scores are based on the number of items answered, and are converted into a standard age score corresponding to age group, similar to an IQ (Intelligence Quotient) score, though with a standard deviation of 16. Since standard age scores represent an average of a variety of skill areas, a trained psychologist will evaluate and interpret an individual's performance on the scale's subtests to discover strengths and weaknesses and offer recommendations based upon these findings (Thorndike, Hagen AND Sattler, 1986). The SB-IV provides a global IQ score called a Test Composite Score. The Test Composite Score is comprised of four Standard Age Scores (broad factors) which include Verbal Reasoning, Abstract/Visual Reasoning, Quantitative Reasoning, and Short-Term Memory.

Stanford-Binet Intelligence Scale, Fifth Edition

The Fifth Edition maintains many of the same subtests and items of previous editions, and includes a global factor (Full Scale IQ Score) and several broad factors at the second level like the Fourth Edition. The Fifth Edition has five factors, (Fluid Reasoning, Knowledge, Quantitative Reasoning, Visual-Spatial Processing, and Working Memory), as opposed to the four in the Fourth Edition. Many toys and objects were returned from earlier editions, as they are helpful with early childhood assessment. Unique to the Fifth Edition is the use of a nonverbal mode of testing covering all five cognitive factors. The range of the scales has been extended to more accurately measure both higher and lower areas of functioning. Changes in the Item and Record Forms are reported to have made the scale more useful for clinical, forensic, school, and vocational applications and interpretations. This Edition also allows for evaluation of the abilities of elderly examinees (Roid, 2003b).

Differences in Versions of the Stanford-Binet

The Stanford-Binet L-M has received much criticism over the years, but it has remained the most stable instrument for predicting school success and assessing intellectual giftedness (Silverman and Kearney,1992). The Fourth Edition and L-M differ in content and emphasis: “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). According to Silverman and Kearney (1992):

Both the old Binet and the new scale are needed for different purposes. The SB Fourth Edition is not a revision of the old scale; it is essentially a new test (Keith, Cool, Novak, White AND Pottebaum, 1988; Rothlisberg, 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. (Silverman and Kearney, 1992, p 35)

Considerable research supports the premise of Silverman and Kearney (1992) that the L-M and Fourth Edition of the Stanford-Binet are different instruments, particularly for gifted students. Thorndike, Hagen and Sattler (1986) compared 82 gifted children’s L-M and Fourth Edition scores and found the L-M scores to be an average of 13.5 points higher. However, Hartwig, Sapp, and Clayton (1987) found no significant differences between the L-M and Fourth Edition composite scores for a sample of 30 non-exceptional children. Livesay (1986) compared L-M and Fourth Edition scores for a group of 120 six-year-old gifted children and found a significant (eight point) difference, with L-M scores higher than Fourth Edition scores. Livesay stated that the Fourth Edition, if used without adjustment of the criterion for consideration as gifted, would result in a lower proportion of children identified as gifted. For Livesay’s sample of 82 gifted children, the correlation between L-M total and Fourth Edition composite was only .27. The correlation between L-M total and Fourth Edition composite for a non-exceptional sample of 139 children was .81. Silverman and Kearney (1992) indicate that the L-M allows a student with a strength in one or two domains to proceed to the Superior Adult III Level on the basis of those strengths alone, allowing a gifted student to demonstrate strengths in specific areas of expertise and consequently achieving a higher overall global score.

Ruf (2003), in the Fifth Edition Technical Manual reports differences in mean scores and correlations across the L-M, Fourth Edition and Fifth Edition (Roid, 2003b). Flynn (1987) found that scores on intelligence tests tend to increase over time. There is some question as to whether the Flynn Effect applies to extremes in IQ measurement (Silverman AND Kearney, 1992). In addition, Teasdale AND Owen (2008) report that the Flynn Effect documented primarily in the 1980s does not appear to be a factor in studies conducted at a later date, particularly studies conducted in the 21st century.

Ruf (2003) indicates that even with taking the Flynn

Effect into account, scores on the Fifth Edition appear generally equivalent to scores on the previous editions of the Stanford-Binet, with one exception – scores on the L-M seem to increase faster than corresponding Fifth Edition as scores approach the gifted extreme. This difference in scores is attributed to the method used to calculate IQ. More recent tests have moved away from the use of the modified ratio IQ scores, which were used to calculate IQ on the L-M, to the use of standard scores which are used to calculate IQ scores on the Fourth and Fifth Edition. Standard Score IQs tend to have an upper limit in the 150-160 point range. But ratio scores go well past that level. Although the Fifth Edition does offer an extended IQ score, such scores are exceedingly rare because they must conform to the percentile frequency requirement of a standard score, not a ratio score.

Few studies have been found in the literature that contain a field-based analysis of all three versions of the test. This scarcity of studies was the impetus for the currently

METHODOLOGY

Subjects

Teachers of gifted children and school district personnel from several school districts in the North Texas and North Florida regions of the USA were asked to identify students currently served in a gifted program between the ages of 8 and 14 who would be willing to participate in a research study involving approximately 3 to 5 hours of testing. As compensation for participation, each student participant was awarded a fifty-dollar gift certificate. Forty-five students were referred and tested. Students who did not score at least a 118 on one of the three tests were eliminated after the initial testing phase. Data from 25 of the original 45 referred students were retained for analysis. Descriptive data are shown in Table 1. Fifty-two percent of the subjects retained were female (n = 13). Participants ranged in ages from 100 to 166 months with a mean value of 131.8 months (SD = 21.95). Although currently receiving services in the gifted program, no student had been administered any version of the Stanford-Binet. All students had previously been administered some form of group intelligence test.

Instruments

Each participant took three versions of the Stanford-Binet. Versions that were administered included: Form L-M (1972), Fourth Edition (1986), and Fifth Edition (1994). The three versions of the Stanford-Binet were administered to participants in random order to eliminate possibilities of a “practice effect” producing differences in test scores. To minimize the effect of examiner experience, each student remained with the same examiner for all three tests.

Procedure

Three advanced graduate students in the Educational Diagnostician Program at a large, comprehensive research university in the Dallas-Ft. Worth Metropolitan area of Texas administered the tests under the supervision of university faculty members, in three sessions. Prior to administering the various versions of the Stanford-Binet, each graduate student was observed by a university faculty member who teaches advanced assessment courses to make certain that administration of the Stanford-Binet complied with procedures delineated in the manual. During training for the administration of the various versions of the Stanford-Binet, scoring of each test was reviewed by two separate faculty members. Differences in scoring were resolved through arbitration among the university faculty and test administrator.

RESULTS

Comparability of Versions of the Stanford-Binet Test

Rank-sum scaling (Dunn-Rankin, Knezek, Wallace, AND Zhang, 2004) of the three tests confirmed that all three were significantly different from each other in the composite IQ scores assigned to this group of subjects. As shown in Table 2, the rank totals resulting from assigning each subject a rank of 1 for the test with the lowest numeric score, a 2 for the test with the second-lowest score, and 3 for the test with the highest score for that individual, resulted in rank totals across the 25 subjects of 34 for SB-5, 49 for SB-4, and 68 for L-M. Also shown in Table 2, rank totals were converted to scale scores on a 0-100 scale. These scales are graphically displayed in Figure 1. The rank-sum differences among the three tests (see Table 2) range from 15 for SB-4 vs. SB-5 to 34 for L-M vs. SB-5, with SB-4 lying in between L-M and SB-5 at a rank sum difference of 19 from L-M. All rank sum differences are well beyond the critical value of 10 required to reach significance at the $p < .001$ level (Dunn-Rankin, et al., 2004).

Magnitude of Differences in Test Versions

Parametric statistical techniques were employed to assess the magnitude of the differences between versions of the Stanford-Binet. Paired t-tests verified that

Table 1. Data for 25 Subjects on Three Versions of the Stanford-Binet Intelligence Test

Subject	GENDER	Age/Mon.	SB-L-M	SB-4	SB-5	CLUSTER
1	F	146.00	146	125	122	3.00
2	F	135.00	126	128	125	3.00
3	F	164.00	120	118	102	2.00
4	M	164.00	119	114	110	2.00
5	M	107.00	131	125	127	3.00
6	M	118.00	168	144	136	1.00
7	F	100.00	136	133	126	3.00
8	F	124.00	131	121	128	3.00
9	F	105.00	133	124	132	3.00
10	M	128.00	149	135	123	1.00
11	F	109.00	123	115	118	2.00
12	F	142.00	132	134	118	3.00
13	M	116.00	140	138	118	3.00
14	F	118.00	124	145	119	3.00
15	M	143.00	140	126	112	3.00
16	M	118.00	115	118	116	2.00
17	M	103.00	116	121	123	2.00
18	M	130.00	120	116	114	2.00
19	M	152.00	154	132	137	1.00
20	F	166.00	149	132	132	1.00
21	M	152.00	161	143	139	1.00
22	F	116.00	145	143	140	1.00
23	M	165.00	163	142	136	1.00
24	F	160.00	158	134	125	1.00
25	F	114.00	147	132	128	1.00

Table 2. Rank Totals and Rank Sum Differences for Three Versions of the Stanford-Binet IQ Test.

Item	Rank Total	Scale score
Min	25	0
1	68	86
2	49	48
3	34	18
Max	75	10

Rank Sum Differences

		L-M	SB-4	SB-5
L-M	1	0		
SB-4	2	19	0	
SB-5	3	34	15	0

```
0    10    20    30     40    50   60    70    80    90    100

+----+---*+----+----+---*+----+----+----+--*-+----+
          SB-5              SB-4                    L-M
```

Figure 1. Display of scale scores for three versions of the Stanford-Binet IQ test

group mean scores for all versions were significantly ($p < .01$) different from each other. Effect sizes for test versions were:

- .56 for L-M vs. SB-4 with L-M 7.32 higher ($t = 3.28$, 24 df, $p < .001$);
- .97 for L-M vs. SB-5 with L-M 12.6 higher ($t = 6.03$,

24 df, p < .0005); and

- .54 for SB-4 vs. SB-5 with SB-4 5.28 higher (t = 3.14, 24 df, p < .004).

These are moderate-to-large effect sizes according to the guidelines provided by Cohen (1988).

Alignment of Versions with Discrimination Index Based on Stanford-Binet L-M

The 25 subjects in this study were assigned to two categories (above average = 1 and exceptional = 2) for the purposes of assessing the discriminating ability of the different versions of the test. A dividing point of IQ 140 was used to classify students as either above average or gifted for the L-M version of the Stanford-Binet. Thirteen students were placed in group 1 (exceptional, IQ at or above 140) and 12 in group 2 (above average, IQ at or below 139). One-way analysis of variance (ANOVA) was used to determine how well the three tests separated the "above average" from "exceptional" groups. As shown in Table 4, all three versions of the test were acceptable in separating the exceptional students from those who were above average. However, Fourth Edition had a much higher F-ratio (f = 13.1, p < .001) than Fifth Edition (f = 7.0, p < .014), indicating it was somewhat better than Fifth Edition for separating the upper versus lower portions of this group of subjects, if the criterion is based on a cut point of 140 or higher on the L-M version of the Stanford-Binet. The discrimination ability of the L-M version (f = 68.03), by virtue of serving as the basis of high-low dichotomization, also can beviewed as an upper limit against which the other f-ratios can be judged.

A visual presentation of the same information contained in Table 4, from a slightly different perspective of the data in Figure 4, illustrates that the Fourth Edition and Fifth Edition under-represent the intellectual ability of children to a greater extent, that is,the higher a child scores on the L-M version of the Stanford-Binet, the more likely the child's IQ score will be compressed towards the mean on the Fourth and Fifth Editions. As shown in Figure 4, for the 13 students in this sample who scored lower than 140 on the L-M version, the overall (mean) values for the group are not greatly different, whether one views the L-M, Fourth Edition, or Fifth Edition average values. However, if one looks at the comparable trends for the 12 students who scored 140 or higher on the L-M version, then the picture is very different. It makes little difference which version is chosen if one examines the lower half of this group, but it makes a very large practical difference if one examines the upper group. The typical 140+ student on the L-M version would most likely have difficulty achieving the 140+ designation on the Fourth Edition or Fifth Edition.

Scaling Analysis of Subjects

A hierarchical cluster analysis procedure based on Ward's method (Dunn-Rankin, et al., 2004) (SPSS, 2004) was carried out to examine the logical grouping of the subjects. A dendogram illustrating the associations among the subjects is shown in Figure 5. Examination of the dendogram in Figure 5 and changes in the amount of variance accounted for during step-wise combinations of clusters resulted in the identification of three profiles of test score patterns:

- Cluster I (median scores of 154-135-135 for SB-L-M, SB-Fourth Edition, SB-Fifth Edition): A group of nine students with high scores on all forms but especially high scores on the L-M version. The greatest differential between the L-M and other forms is 19 points.
- Cluster II (119-117-115): A group of six students with the lowest scores and little differentiation by test version.
- Cluster III (132.5-127-123.5): A group of ten students with moderately high scores and a trend line across test versions resembling Cluster II.

These trends are graphically displayed in Figure 6. This differential analysis graphically demonstrates the tendency for some non-exceptional students to score as well or better on the newer forms of the Stanford-Binet Intelligence Scale, compared to the L-M. The problem appears to reside in the category of the exceptional students (Cluster 1), where L-M scores are higher than either Fourth or Fifth Edition.

DISCUSSION

Our findings of low congruence between the IQ scores assessed on the Stanford-Binet L-M versus either the Fourth Edition or the Fifth Edition for exceptional (>=140 IQ) children appear to be consistent with findings from previous studies. As mentioned in an earlier section of this paper, Thorndike, Hagen and Sattler (1986) found that the correlation between L-M and Fourth Edition for a sample of 139 non-exceptional children was .81 while the correlation between L-M and Fourth Edition for a sample of 82 gifted children was .27. Ruf (2003) indicates that indeed the earlier L-M version of the Stanford-Binet differs greatly from the more recent Fourth and Fifth Editions. The L-M is based on ratio scores and the Fourth and Fifth Edition are based on standard scores. Obtaining an extremely high score on the Fourth or Fifth Edition is not likely because there is less of an opportunity to achieve a high score when the examiner must adhere to numbers representing a normal distribution. Although the current sample of 25 subjects divided into 13 non-exceptional and 12 exceptional is

Table 3. Final Coordinates and Goodness of Fit for Unidimensional Solution, Dimensionality 1 Coordinates

	Dimension
	1
L-M	.772
SB-4F	-.155
SB-5Total	-.616

Normalized Raw Stress	.0001
Stress-I	.0086
Stress_II	.0227
S-Sress	.0002
Dispersion Accounted for	.9999
Tucker's Coefficient of Congruence	1.0000

Table 4. Analysis of Variance for Above Average (Group 1) and the Exceptional Students (Group 2) based on cut point of 140 or higher on the L-M version of the Stanford-Binet IQ Test.

		N	Mean	Std. Deviation	Minimum	Maximum
L-M	1.00	13	125.08	6.9	115	136
	2.00	12	151.67	9.1	140	168
	Total	25	137.84	15.7	115	168
SB-4	1.00	13	124.00	9.0	114	145
	2.00	12	135.50	6.6	125	144
	Total	25	129.52	9.7	114	145
SB-5	1.00	13	119.85	8.2	102	132
	2.00	12	129.00	9.1	112	140
	Total	25	124.24	9.7	102	140

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
L-M	Between Groups	4411.8	1	4411.8	68.029	.000
	Within Groups	1491.6	23	64.9		
	Total	5903.4	24			
SB-4F	Between Groups	825.2	1	825.2	13.117	.001
	Within Groups	1447.0	23	62.9		
	Total	2272.2	24			
SB-5Total	Between Groups	522.9	1	522.9	7.009	.014
	Within Groups	1715.7	23	74.6		
	Total	2238.6	24			

much smaller, correlation trends are similar to those of previously reported. Kendall's tau-b for L-M with Fourth Edition was .42 ($p < .05$) for the non-exceptional students in our sample while it was slightly lower at .41 (NS) for the exceptional students. Nonparametric correlations between L-M and Fifth Edition show the same trend but even larger differences in that for non-exceptional students tau = .54 ($p < .05$) for L-M with Fifth Edition but for exceptional students tau = .45.

Differences in mean scores by test version for the 25 subjects in this study align well with findings from a previous study and extend findings to the Fifth Edition. Livesay (1986) found that for 120 six-year-old gifted children the average L-M to Fourth Edition decrement

Table 5.Two Dimensional Coordinates for 25 Subjects Based on Multidimensional Preference Mapping of Scores on Three Versions of the Stanford-Binet Intelligence Scale

Subject	Dimension 1 Coordinate	Dimension 2 Coordinate
1	0.992	-0.123
2	0.049	0.999
3	0.760	0.650
4	0.976	0.219
5	0.841	-0.541
6	1.000	0.004
7	0.872	0.490
8	0.549	-0.836
9	0.377	-0.926
10	0.971	0.238
11	0.814	-0.580
12	0.603	0.798
13	0.748	0.664
14	-0.102	0.995
15	0.960	0.281
16	-0.580	0.815
17	-0.999	-0.042
18	0.996	0.095
19	0.897	-0.441
20	0.972	-0.236
21	0.998	-0.067
22	0.921	0.389
23	1.000	-0.026
24	1.000	0.028
25	0.999	-0.038

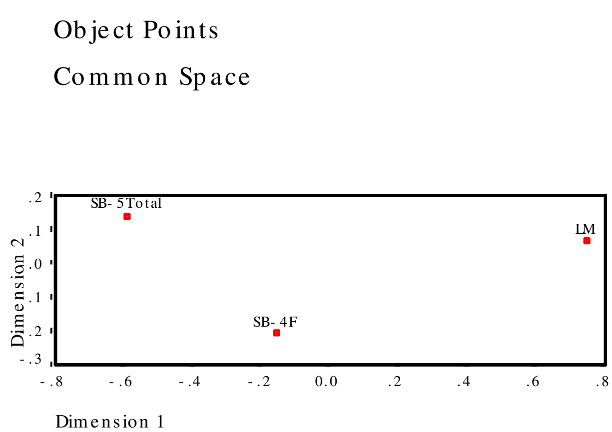

Figure 2. One-dimensional solution from Multidimensional Scaling analysis

was eight points. This difference closely corresponds to the 7.3 point average decrement found between L-M and Fourth Edition in 9 to 14 year subjects in the current study, and adds credibility to the 12.6 point average decrement found between the L-M and Fifth Edition. The bias that was identified in Fourth Edition for exceptional children in 1986 appears to be present and perhaps greater in Fifth Edition.

CONCLUSION

The Stanford-Binet Intelligence Scale, Form L-M has historically been used to locate highly gifted students. The Stanford-Binet Fourth Edition was developed to compensate for the cultural bias, difficulty in scoring, and subjective bias in interpreting the results many associated with the Stanford-Binet L-M. However, the

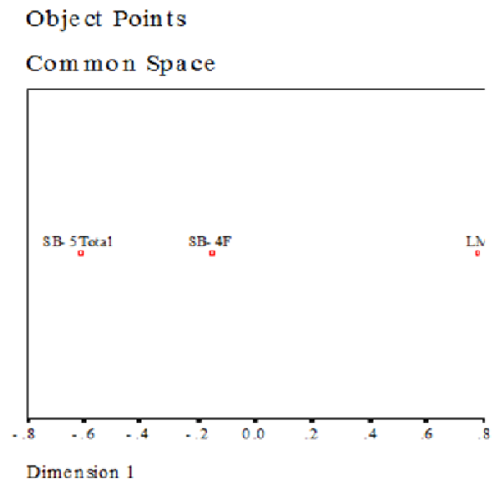

Figure 3. Two-dimensional solution from Multidimensional Scaling analysis

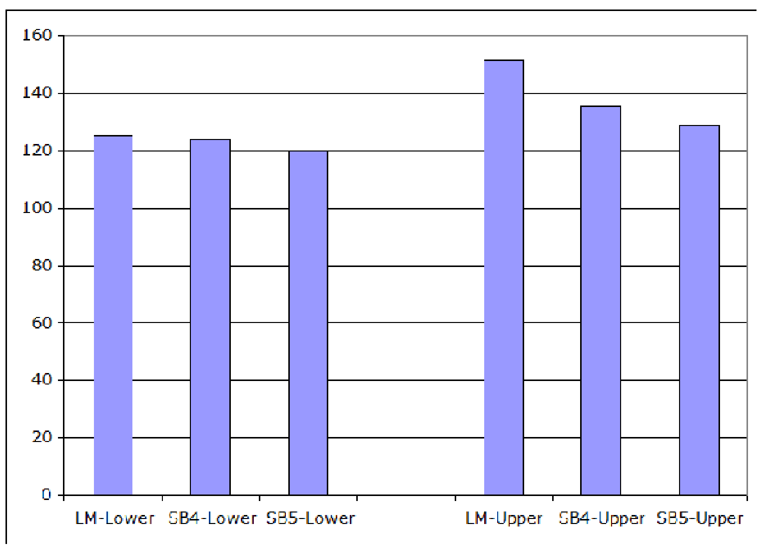

Figure 4. Mean values by test version for 13 students completing three versions of the Stanford-Binet who scored lower than 140 on the L-M version versus 12 students who scored 140 or higher.

Fourth Edition had many other weaknesses such as limiting the ability of many gifted students to obtain a high score. The Stanford-Binet Fifth Edition was developed in part to address the weaknesses of Fourth Edition. Unfortunately, the new test may actually be biased even more against the highly gifted. In our study, children who averaged 136.84 (SD = 15.68) on the L-M version averaged 129.52 (SD=9.73) on the Fourth Edition and 124.24 (SD=9.66) on the Fifth Edition. These differences align closely with those found by researchers who have completed similar studies (Thorndike, Hagen, AND Sattler, 1986; Livesay, 1986) in the past. In addition, Ruf (2003) found that gifted students with a mean score of 129.17 (SD=10.89) on the Fifth Edition obtained a mean score of 153.89 (SD=21.79) on the L-M. In the technical manual of the Fifth Edition, Roid actually indicates that students in the gifted range who score an IQ of 145 on the L-M are predicted to score an IQ of only 122-123 on the Fifth Edition. The L-M version of the Stanford-Binet appears to continue to have a higher ceiling for exceptional children than the newer and more widely-used Stanford-Binet Fourth Edition and Stanford-Binet Fifth Edition.

Based on the data gathered for this study, students

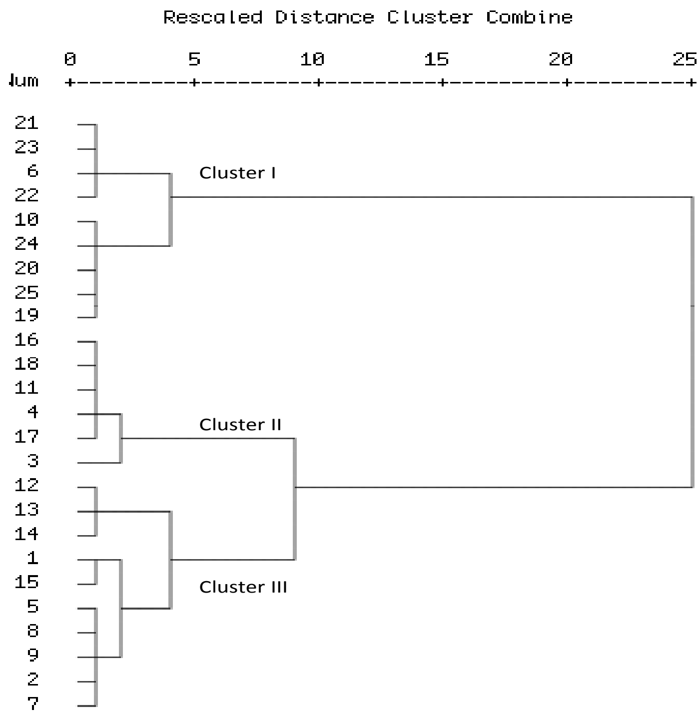

Figure 5. Horizontal dendogram of 25 subjects clustered according to responses on 3 versions of the Stanford-Binet Intelligence Test

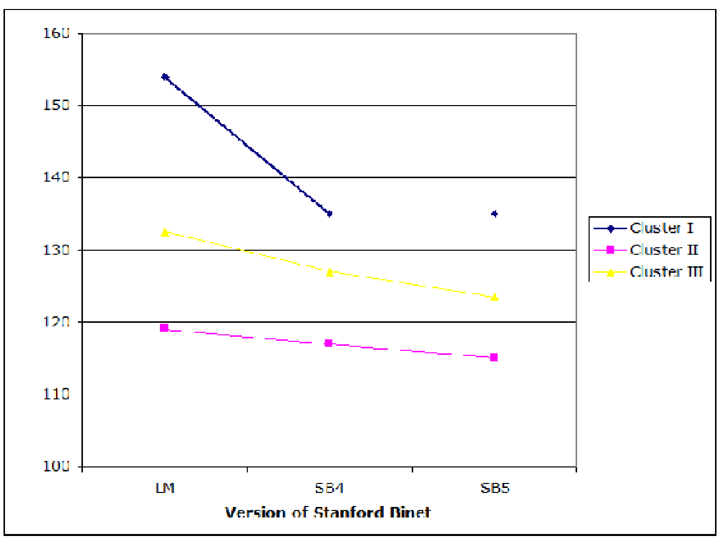

Figure 6. Graphical display of the median scores for each cluster of subjects on three versions of the Stanford-Binet Intelligence Test.

who score 140 or above on the L-M version of the Stanford-Binet will be unlikely to reach that mark on either the Fourth Edition or Fifth Edition. A practical consequence of this is that many children who should be identified as highly gifted are not being receiving that classification and consequently may not gain access to programs which support students who have extremely high intellectual abilities. A possible short-term solution to this ceiling effect is to have students re-tested on the L-M version whenever the referring party suspects that the exceptional abilities are not being adequately identified by the Fourth Edition or Fifth Edition. Another possibility would be to lower the cut-off scores for gifted programs for scores obtained on the Fifth Edition. A preferable long-term solution would be to have all three versions of the Stanford-Binet renormed together, by having a large pool of subjects complete all three. Version-specific, numeric cutoff points reflecting equivalent intelligence quotients throughout the low-to-high continuum, could then be set in place.

REFERENCES

Cohen J (1988). *Statistical power analysis for the behavioral sciences*, *(2nd ed.)* New York: Academic Press.

Dunn-Rankin P, Knezek G, Wallace S, Zhang S (2004) *Scaling Methods*, *(2nd ed.)*. Mahwah, NJ: Erlbaum.

Hartwig S, Sapp G, Clayton G (1987). Comparison of the Stanford-Binet Intelligence Scale, Form L-M and the Stanford-Binet Intelligence Scale, Fourth Edition, Psychological Reports, 60:1215-1218.

Keith TZ, Cool VA, Novak CG, White LJ, Pottebaum SM (1988). Confirmatory factor analysis of the Stanford-Binet Fourth Edition: Testing the theory-test match. *J. School Psychol. 26*:253-274.

Livesay K (1986). Comparisons of the Stanford-Binet: Fourth Edition to the Stanford-Binet: LM and WISC-R with gifted referrals. Paper presented to the Annual Conference of the Florida Association of School Psychologists, Jacksonville.

Riverside Publishing Company (2005). *Stanford-Binet Intelligence Scales*. Retrieved February 3, 2005 from http://www.riverpub.com/products/clinical/sbis/home.html

Robinson N, Dale PS, Landesman S (1990). Validity of Stanford-Binet IV with linguistically precocious toddler. Intelligence, *14*:173-186. Roid, G.H. (2003). *Facts in brief on the SBR*. Riverside Publishing Co.

Roid GH (2003). Stanford-Binet Intelligence Scales, Fifth Edition, Technical Manual. Itasca, IL: Riverside Publishing.

Rothlisberg BA (1987). Comparing the Stanford-Binet, Fourth Edition, to the WISC-R: A concurrent validity study. *The Journal of School Psychology, 25*(2):193-196.

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

Ruf DL (2003). Use of the SB5 in the Assessment of High Abilities (Stanford-Binet Intelligence Scales, Fifth Edition Assessment Service Bulletin No. 3). Ilasca, IL: Riverside Publishing.

Silverman LK (1989). Lost: IQ point per year for the gifted. Paper presented at the National Association for Gifted Children 36th Annual Convention. Cincinnati, OH.

Silverman LK, Kearney K (1992). The case of the Stanford-Binet L-M as a supplemental test. Roeper Review 15(1):34-38.

Stanley JC (1990). Leta Hollingsworth's contribution to above-level testing of the gifted. *Roeper* Rev. 12(3):166-171.

Statistical Package for the Social Sciences (SPSS), Inc. (2004). *SPSS reference guide*. Chicago: SPSS Inc.

Teasdale TW, Owen DR. (2008). Secular declines in cognitive test scores: A reversal of the Flynn Effect. Intelligence, 36(2):121-126.

Terman LM (1916). The measurement of intelligence. Boston: Houghton Mifflin.

Terman LM, Merrill MA (1972). Manual for the Stanford-Binet, Form L-M. Boston: Houghton Mifflin Company.

Thorndike RL, Hagen EP, Sattler JM (1986). The Stanford-Binet Intelligence Scale: Fourth Edition. Technical manual. New York: Riverside Publishing.

Tyler-Wood T, Carri L (1990). The effectiveness of various measures of cognitive ability for identifying gifted students. Roeper Review 14(2):63-64.

How to cite this article: Tyler-Wood T, Knezek G, Christensen R, Morales C, Dunn-Rankin P (2014). Scaling Three Versions of the Stanford-Binet Intelligence Test: Examining Ceiling Effects for Identifying Giftedness. Educ. Res. 5(2):42-51