Stanford BINET Intelligence Scales

Stanford-Binet Intelligence Scales, Fifth Edition
Assessment Service Bulletin Number 1

History of the Stanford-Binet Intelligence Scales: Content and Psychometrics

Kirk A. Becker

The 2003 publication of the Stanford-Binet Intelligence Scales, Fifth Edition *represents the latest in a series of innovations in the assessment of intelligence and abilities. Using qualitative and quantitative methods, this bulletin examines the similarities and differences between the different editions of the Stanford-Binet published over the past century. It discusses the development and integration of age-scale and point-scale formats for subtests, the theoretical structure of the test (single versus hierarchical, use of Nonverbal and Verbal domains, and links to the Cattell-Horn-Carroll theory), and changes in item content related to this theoretical structure. The new edition provides greater differentiation in the measurement of abilities, the precursors of which were nevertheless present in earlier editions.*

Riverside Publishing
A HOUGHTON MIFFLIN COMPANY

Printed in the United States of America.

Reference Citation

- To cite this document, use:

  Becker, K. A. (2003). *History of the Stanford-Binet intelligence scales: Content and psychometrics.* (Stanford-Binet Intelligence Scales, Fifth Edition Assessment Service Bulletin No. 1). Itasca, IL: Riverside Publishing.

For technical information, please call 1.800.323.9540, visit our website at www.stanford-binet.com, or e-mail us at rpcwebmaster@hmco.com

1 2 3 4 5 6 – XBS – 06 05 04 03

History of the Stanford-Binet Intelligence Scales: Content and Psychometrics

Overview

The study of the history of major psychological assessment instruments is hardly an active field. Most professionals view assessment instruments simply as practical tools of little interest beyond their immediate applied use. Nevertheless, when authors begin a revision of such an instrument, particularly when the instrument has already undergone a number of revisions, they are wise to examine the history of that assessment to provide continuity of measurement, to improve on the features, and to overcome limitations of earlier versions. In addition, understanding the history of a test may help the clinician compare the scores on the newest edition of the test to the scores on an earlier edition of the test with which they are familiar.

The history of assessment instruments may also be of interest to historians of education, psychology, and science. Such interest would probably focus on an original approach to assessment and the subsequent development of that approach. Truly original assessment tools are rare. Such instruments, and their attendant theories, may have a sufficiently large impact on the field as to essentially spur a revolution in theory and application, along the lines of Kuhn's (1970) discussion of paradigm shifts in the history of science. A new assessment paradigm would begin with revolutionary fervor but would cool over time into "normal science." The history of intelligence measurement, and particularly Binet's major contribution, almost certainly qualifies as an original approach to assessment. As with other scientific revolutions described by Kuhn, Binet's initial contribution led to a period of revolutionary science followed by a drawn-out period of normal science, during which the original brilliant insight was improved gradually over time with steady and practical enhancements. Kuhn noted that although revolutionary science tends to grab the attention and the spotlight, normal science is what tends to be responsible for real progress, in both basic science and its applications. One can probably say the same for the history of intelligence tests and intelligence testing.

This bulletin provides a history of the Stanford-Binet, beginning with its precursors in the work of Alfred Binet and Theodore Simon at the turn of the 20th century. It focuses on its five American editions, from the first version, which was published in 1916, to the most recent version, the *Stanford-Binet Intelligence Scales, Fifth Edition* (SB5) (Roid, 2003a), published in 2003. Although the purpose of this bulletin is to describe the similarities and differences across the five American editions, readers with a broader interest in the history of assessment instruments may also find the discussion of interest, although it is not the primary goal of this bulletin to provide the full history of intelligence testing in terms of a Kuhnian scientific revolution.

History of the Stanford-Binet

At a conference in Rome in April 1905, Dr. Henri Beaunis read a paper prepared by Alfred Binet and Theodore Simon that announced the development of an objective measure capable of diagnosing different degrees of mental retardation (Wolf, 1973). This announcement was followed 2 months later by the publication of the *Binet-Simon Intelligence Test* in *L'Anée Psychologique* (Binet & Simon, 1905). The original form of this test was expanded and revised, leading to new versions in 1908 and 1911. The new forms were the result of extensive research and testing involving "normal" as well as mentally retarded examinees.

In 1916, Lewis Terman authored *The Measurement of Intelligence: An Explanation of and a Complete Guide for the Use of the Stanford Revision and Extension of the Binet-Simon Intelligence Scale* (Terman, 1916). This manual presented translations and adaptations of the French items, as well as new items that Terman had developed and tested between 1904 and 1915. Although there were other translations of the Binet-Simon available around this time (Binet & Simon, 1916; Kuhlmann, 1912; Melville, 1917; Herring, 1922), Terman's normative studies and his methodical approach are credited with the success of the Stanford-Binet (Minton, 1988).

Over the two decades following the initial publication of the Stanford-Binet, Terman continued his research and development of the test. Working with Maud Merrill, first his student and later a fellow professor and research collaborator at Stanford University, Terman created two parallel forms of the Stanford-Binet. These forms used many of the items from the original Stanford revision and added a substantial number of new items. With regard to this revision, Terman and Merrill wrote that they had "provided two scales instead of one, have extended them so as to afford a more adequate sampling of abilities at the upper and lower levels, have defined still more meticulously the procedures for administration and scoring, and have based the standardization upon larger and more representative populations" (Terman & Merrill, 1937, p. ix). These parallel forms were published as Form L (for Lewis) and Form M (for Maud) of the Stanford-Binet.

In the 1950s, Merrill took the lead in revising the Stanford-Binet, selecting the best items from Forms L and M to include in a new version of the test. The two forms from 1937 were combined to create the Form L-M. This form was published in 1960 (Terman & Merrill, 1960) and was later renormed in 1973 (Terman & Merrill, 1973). This form added alternate items at all levels, but otherwise, the format remained similar to the 1937 forms.

The *Stanford-Binet Intelligence Scale: Fourth Edition* (Thorndike, Hagen, & Sattler, 1986) moved from the age-scale format introduced by Binet to a point-scale format (the section on "Test Structure" that follows provides details on the characteristics of age and point scales). Many of the items and item-types from the prior editions were included in the Fourth Edition, and extended scales were created using the same types of items and activities. In the Absurdities test, for example, four classic items were used in addition to 28 new items. Also, several completely new subtests, such as Matrices and Equation Building, were created. Besides the new and expanded tests, the Fourth Edition provided several factors (Verbal Reasoning, Abstract/Visual Reasoning, Quantitative Reasoning, and Short-Term Memory) in addition to IQ. Although the prior versions had items that related to these four areas (McNemar, 1942; Sattler, 1965), the published test had never offered scores for these factors. The Fourth Edition also formalized the practice of multi-stage testing, in which performance on the Vocabulary scale determines the starting point for subsequent tests. While some examiners used the vocabulary test for routing on earlier editions of the test, this was not official practice.

In 2003, the Fifth Edition (Roid, 2003a) was published. This edition attempts to carry on the tradition of the prior editions while taking advantage of current research in measurement and cognitive abilities. Like the Fourth Edition, the SB5 includes multiple factors. These factors are modified from those on the Fourth Edition, but represent abilities assessed by all former versions of the test. The use of routing subtests continues, with a nonverbal routing test added to complement vocabulary. The Fifth Edition reintroduces the age-scale format for the body of the test, presenting a variety of items at each level of the test. The age-scale is intended to provide a variety of content to keep examinees involved in the testing experience and to allow for the introduction of developmentally distinct items across levels.

Test Structure

The Stanford-Binet is one of the first examples of an adaptive test (Reckase, 1989). Examiners use the information they have about an examinee to determine where to begin testing and administer only those items that are appropriate for that examinee. This format reduces the time required to obtain reliable information from a test and decreases the frustration examinees experience when presented with items that are too hard or too easy. The use of multiple possible starting points, along with basal and ceiling rules, limits the time required to administer the test and maximizes the information obtained from each item.

One element of test structure that appears throughout the history of the Stanford-Binet is that of point scales and age scales. A point scale is the currently widespread arrangement of tests into subtests, with all items of a given type administered together. Age scales, long a part of the Stanford-Binet format, may not be familiar to the current generation of examiners. Initially, this format was used to provide a direct translation of the child's performance to mental age. Psychometric as well as developmental information was used to place items on the test. Examinees experienced a variety of items that changed both quantitatively and qualitatively. Definitions, for example, went from concrete words to abstract words to the comparison of abstract words. The argument has often been made that this format is more engaging and provides a richer opportunity for the examiner to observe the examinee's performance. Although this perspective guided the development of the Fifth Edition, it is not without its detractors. With the publication of the 1916 edition, Robert Yerkes began a series of debates with Lewis Terman on the appropriateness of the age scale (Yerkes, 1917). While Terman's methods prevailed at the time, the structure of many current tests (e.g., Wechsler, 1991) shows the popularity of the point-scale subtest.

Throughout most of its history, the Stanford-Binet maintained a hybrid structure, combining point-scale and age-scale formats. Terman presented two parallel vocabulary scales in 1916, and every version of the Stanford-Binet except Form M has included a vocabulary scale. In 1986, the Fourth Edition provided a standard method for using Vocabulary as a routing test to determine where to begin testing. Although this was the first formal mention of routing in a Stanford-Binet manual, using vocabulary for this purpose had been the unofficial practice of many examiners for decades. The Fifth Edition includes a nonverbal routing test in addition to Vocabulary, and it uses performance on these subtests to route to the Nonverbal and Verbal age scales. Table 1 provides a summary of the structure and features of the different editions of the Stanford-Binet.

Table 1

Test Structure of the Stanford-Binet: 1916 to 2003

Edition	Structure	Abilities Measured
1916	■ Parallel vocabulary tests ■ Single age scale	■ General intelligence
1937	■ Form L vocabulary test ■ Parallel age scales	■ General intelligence
1960/1973	■ Vocabulary test ■ Single age scale	■ General intelligence
1986	■ Vocabulary routing test ■ Subtest point scales	■ General intelligence ■ Verbal Reasoning ■ Abstract/Visual Reasoning ■ Quantitative Reasoning ■ Short-Term Memory
2003	■ Hybrid structure ■ Verbal routing test ■ Nonverbal routing test ■ Verbal and nonverbal age scales	■ General intelligence ■ Knowledge ■ Fluid Reasoning ■ Quantitative Reasoning ■ Visual-Spatial Processing ■ Working Memory ■ Nonverbal IQ ■ Verbal IQ

Verbal and Nonverbal Content

The verbal content of the Stanford-Binet has been of concern since the first publication in 1916, in which verbal items predominated. Subsequent revisions have attempted to better balance the verbal and nonverbal items, and McNemar reports that attempts were made to create a parallel nonverbal form in 1937 (McNemar, 1942). During the 1986 revision, the authors tried to create a nonverbal routing test using Matrices; however, sufficient low-end items were not created. While several nonverbal forms were recommended for different editions of the Stanford-Binet (e.g., McNemar, 1942; Delaney & Hopkins, 1987; Glaub & Kamphaus, 1991), the test never had a published nonverbal form prior to 2003. The Fifth Edition provides an equal balance of verbal and nonverbal content within each factor. Norms are provided for Verbal IQ (VIQ) and Nonverbal IQ (NVIQ) scores as well as for Full Scale IQ (FSIQ), and the correlations between these scores range from .94 to .97 across the age range of the test. Tables 2 and 3 present the nonverbal content of all editions of the Stanford-Binet. Table 4 compares the VIQ and NVIQ of the Fifth Edition with FSIQ.

Table 2

Nonverbal Items on the Stanford-Binet: 1916 to 1973

Level	1916	Form L	Form M	Form L-M
2-0	n/a	2	2	3
2-6	n/a	1	2	1
3-0	0	5	3	6
3-6	n/a	2	5	5
4-0	3	2	2	1
4-6	n/a	2	3	2
5	3	4	4	6
6	0	5	2	2
7	2	2	1	2
8	1	0	0	0
9	1	2	1	1
10	2	1	0	0
11	n/a	1	1	1
12	1	0	2	2
13	n/a	3	1	3
14	0	1	1	0
AA	0	0	1	1
SAI	1	0	0	0
SAII	n/a	0	0	0
SAIII	n/a	1	0	0
Total NV items	14	34	31	36
Total items	90	129	129	142
Total percent	16%	26%	24%	25%

Note. The 1916 edition did not have all of the levels that later editions had, so the levels not applicable to that edition are marked n/a.

Abbreviations: AA = average adult, SAI = superior adult I, SAII = superior adult II, SAIII = superior adult III, NV = nonverbal

Brief/Abbreviated Forms

Testing time and examinee fatigue have always been important issues for psychological assessment practitioners. Recognizing that testing time might need to be reduced, Lewis Terman included instructions for a minimal level of lenience in establishing basal and ceiling performance (Terman, 1916). An examinee, he reasoned, could miss one item at an age level and still have a basal. Later, in 1937, instructions were included with the Stanford-Binet for the administration

Table 3

Nonverbal Content of the Fourth and Fifth Editions of the Stanford-Binet

SB IV	SB5
Bead Memory	Nonverbal Fluid Reasoning (Object Series/Matrices)
Pattern Analysis	Nonverbal Knowledge (Procedural Knowledge, Picture Absurdities)
Absurdities	Nonverbal Quantitative Reasoning (Quantitative Reasoning)
Copying	Nonverbal Visual-Spatial Processing (Form Board, Form Patterns)
Memory for Objects	Nonverbal Working Memory (Delayed Response, Block Span)
Matrices	
Paper Folding and Cutting	

Table 4

Correlations of Verbal IQ (VIQ) and Nonverbal IQ (NVIQ) With Full Scale IQ (FSIQ)

	2-0 to 5-11	6-0 to 10-11	11-0 to 16-11	17-0 to 50-11	51 to 85+	2 to 85+
VIQ/FSIQ	.96	.96	.96	.97	.97	.96
NVIQ/FSIQ	.94	.95	.96	.97	.97	.96

Note. From Roid, 2003a.

of an abbreviated test battery (Terman & Merrill, 1937). With little loss of reliability, designated items could be omitted from the test to conserve time. Remaining items administered in this fashion were given more weight in the calculation of IQ. This feature was retained in the 1960/1973 edition where, for instance, an examiner could omit Memory for Sentences and Copying a Bead Chain from Memory at age 13. The Fourth Edition also allowed for an abbreviated administration through its flexible selection of subtests. Different combinations of subtests were recommended for different purposes, including a four-test screening battery and a six-test screening battery. While these screeners do not allow for a good measure of an individual's pattern of abilities, they provide a quick measure of IQ (Thorndike, Hagen, & Sattler, 1986). A similar method is used in the Fifth Edition, with Vocabulary and Object-Series/Matrices providing the Abbreviated Battery IQ (ABIQ). The ABIQ, like the FSIQ, is equally weighted in terms of verbal and nonverbal content. Table 5 summarizes the different methods for abbreviated test administration over time.

Table 5

Summary of Stanford-Binet Abbreviated Forms

Edition	Abbreviated Form
1916	Leniency in basal/ceiling rules to save time
1937	Selected items omitted
1960/1973	Selected items omitted
1986	4- and 6-subtest abbreviated batteries recommended
2003	2-subtest Abbreviated Battery IQ available

Reviews of the Stanford-Binet

Along with the extensive research literature on the Stanford-Binet, reviews of the test have been available since before the first Mental Measurement Yearbook was published (Buros, 1938; Pratt, 1917). The content of these reviews has been distilled where possible in Table 6 to offer a summary of the advantages and possible limitations of the different versions of the test. The 1916 edition of the Stanford-Binet provided continuity with the Binet-Simon test, while at the same time expanding the range of items and providing a large research sample. The test offered only limited nonverbal content, provided only sparse instructions on scoring some items, and was not an adequate measure of adult intelligence. The 1937 editions again extended the range of items, provided a set of toys and objects to engage young children, and added more nonverbal content, predominantly at the lower end of the scale. This edition also improved the psychometric characteristics of the test by introducing a parallel form and more

Table 6

Features and Possible Limitations of the Stanford-Binet Over Time

Year	Advantages	Limitations
1916	■ Contains alternate items at most age levels ■ Shares items to maintain continuity with earlier versions ■ Emphasizes abstraction and novel problem solving ■ Extends range of items relative to Binet-Simon ■ Based on extensive research literature ■ Extensive standardization performed	■ Inadequately measures adult mental capacity ■ Has inadequate scoring and administrative procedures at some points ■ Measures only single factor (*g*) ■ Has nonuniform IQ standard deviation ■ Has single test form ■ Is verbally loaded
1937	■ Contains alternate items at most levels ■ Shares items to maintain continuity with earlier versions ■ Extends range of items ■ Based on extensive research literature ■ Contains more performance tests at earlier age levels ■ Contains more representative norms ■ Includes parallel form ■ Uses toys to make test more engaging for young children ■ Verbal items allow subjects to display fluency, imagination, unusual or advanced concepts, and complex linguistic usage	■ Some items have ambiguous scoring rules ■ Form M lacks vocabulary ■ Has longer administration time than 1916 version ■ Measures only single factor (*g*) ■ Has nonuniform IQ standard deviation ■ IQs not comparable across ages ■ Sample had higher SES and higher percentage of urban children than general population ■ Has unequal coverage of different abilities at different levels ■ Is verbally loaded
1960/1973	■ Administers several varied tests to each examinee to keep children interested ■ Retains best items from Forms L and M ■ Has better layout than previous versions ■ Manual presents clear scoring rules ■ Contains alternate items at each age level ■ Shares items to maintain continuity with earlier versions ■ Eliminates items that are no longer appropriate ■ Based on extensive research literature ■ Presents stimulus material in spiral-bound book ■ Has uniform IQ standard deviation ■ Uses toys to make test more engaging for young children	■ Has inadequate ceiling for adolescents and highly gifted examinees ■ Measures only single factor (*g*) ■ Separates scoring standards from items ■ Is verbally loaded
1986	■ Contains both a general composite score and several factor scores ■ Shares items to maintain continuity with earlier versions ■ Easel format with directions, scoring criteria, and stimuli makes administration easier ■ Emphasizes abstraction and novel problem solving; emphasizes verbal reasoning less compared with prior versions ■ Technical Manual reports extensive validity studies ■ Has flexible administration procedures ■ Contains higher ceilings for advanced adolescents than Form L-M ■ Number of basic concepts in preschool level tests compares favorably with other tests for that age range ■ Contains understandable age-level instructions for young children ■ Uses adaptive testing (routing) to economize on administration time and reduce examinee frustration ■ Uses explicit theoretical framework as guide for item development and alignment of subtests within modeled hierarchy ■ Has wider age range than prior versions (2-0 through 23) ■ Creatively extends many classic item types	■ Less gamelike than earlier versions; yields less information from styles and strategies due to decreased examiner/examinee interaction ■ Contains no toys ■ Norming sample overrepresents managerial/professional and college-educated adults and their children ■ Has possible lack of comparability in the content of area scores at different ages due to variability of subtests used in their computation ■ Has psychometric rather than developmental emphasis ■ Has standard deviation of 16 rather than 15 for composite scores; *M* = 50, *SD* = 8 for subtests ■ Contains subjectivity (examiner preference) when determining subtests used to compute composite score ■ Unable to diagnose mild retardation before age 4 and moderate retardation before age 5

Table 6 (Continued)

Features and Possible Limitations of the Stanford-Binet Over Time

Year	Advantages
2003	■ More gamelike than earlier versions with colorful artwork, toys, and manipulatives ■ Matches norms to 2000 U.S. Census ■ Contains nonverbal as well as verbal routing test ■ Contains both a general composite score and several factor scores ■ Shares items to maintain continuity with earlier versions ■ Covers age range of 2-0 through 85+ ■ Change-sensitive scores allow for evaluation of extreme performance ■ Has easel format with directions, scoring criteria, and stimuli, for easy administration ■ Has equal balance of verbal and nonverbal content in all factors ■ Contains Nonverbal IQ ■ Has standard deviation of 15 for composite scores, allowing easy comparison with other tests; *M* = 10, *SD* = 3 for subtests ■ Uses adaptive testing (routing) to economize on administration time and reduce examinee frustration ■ Uses explicit theoretical framework as guide for item development and alignment of subtests within modeled hierarchy ■ Extends low-end items, allowing earlier identification of individuals with delays or cognitive difficulties ■ Extends high-end items to measure gifted adolescents and adults

Note. Because the 2003 measure was just published, a discussion of possible limitations does not yet exist in the literature. (See discussion in text.)

representative norms. While better than the 1916 edition, Forms L and M were still criticized for the quality of the scoring rules, the paucity of nonverbal content at the upper levels of the test, and the nonuniform standard deviation of IQ that led to different interpretations of IQ at different ages. This last point was corrected with the publication of the Form L-M, which provided tables to correct for this. Not only were the best items from the parallel 1937 forms used to create Form L-M, but also the ambiguities in scoring were cleared up, and the stimulus materials were presented in a much more convenient bound format. Some reviewers suggested that the format could be further improved by providing the scoring standards alongside the items in a single book. The test was also criticized for remaining heavily weighted with verbal materials, having what was perceived as an inadequate ceiling for adolescents and highly gifted examinees, and continuing to provide only a single measure of general intelligence.

The Fourth Edition attempted to address many concerns that had been raised with prior versions of the test, while maintaining the same types of tasks and items. In particular, this version of the test offered several factor scores based on an explicit theoretical framework. This test introduced the easel format to the Stanford-Binet, providing both administration and scoring information as well as stimulus material in one place. The test featured higher ceilings for adolescents and over five times as many nonverbal items as the previous edition. Although the test provided for flexible administration, such a degree of flexibility can also lead to unneeded complexity. The test introduced creative extensions of classic items, but it lacked many of the toys and other interesting stimuli from the earlier editions. The continued use of a standard deviation of 16 was also criticized, as was the sample weighting done to approximate the characteristics of the general population.

The Fifth Edition is too new at this point to provide a list of possible limitations collected from the literature. Attempts were made to address the limitations of prior versions, while maintaining the advantages. Artwork and manipulatives have been improved, and toys and gamelike materials were included. The SB5 does not allow for as many administrative options as the Fourth Edition; this makes for a more straightforward testing session. The Fifth Edition covers the widest age range of any Stanford-Binet (2 through 85+ years) and addresses the criticism about verbal content, norms, and the standard deviation. (The Fifth Edition uses a standard deviation of 15 for its IQ scales.)

Score/Test Comparability

With any newly revised test, questions arise about the relationship between scores and interpretations of the new version relative to the research and use of the prior editions. With the research tradition of the Stanford-Binet (as of 2003, over 2,200 articles), it is especially important to provide these comparisons. The comparability of new and old forms, as well as the validity evidence reported in the Technical Manual for the Fifth Edition (Roid, 2003b), provide the initial evidence for appropriate use of the test until independent research is published. Table 7 shows the correlations between general ability on the last three editions of the Stanford-Binet, while Table 8 shows the correlations between factors on the Fourth and Fifth editions.

Table 7

Correlations of FSIQ for the SB Form L-M, SB IV, and SB5

	Test Correlations (Corrected for Restriction or Expansion of Variance)		Test Correlations (Uncorrected)	
	SB5	**SB IV**	**SB5**	**SB IV**
SB IV	0.90		0.83	
Form L-M	0.85	0.80	0.80	0.81

Note. N = 104 for SB5 and SB IV, *N* = 80 for SB5 and SB L-M, *N* = 139 for SB IV and SB Form L-M

Table 8

SB IV and SB5 Factor Correlations

	Factor Correlations (Corrected for Restriction or Expansion of Variance)				Factor Correlations (Uncorrected)					
	SB IV Scores				SB IV Scores					
SB5 Scores	**VR**	**A/VR**	**QR**	**STM**	**VR**	**A/VR**	**QR**	**STM**	***M***	***SD***
KN	.73				.57				108.64	9.85
FR		.54				.48			106.96	12.93
VS		.69				.64			107.90	13.20
QR			.77				.73		107.37	13.34
WM				.64				.55	104.33	11.76
M	110.71	111.96	112.11	106.66			110.71	111.96	112.11	106.66
SD	12.71	14.24	14.90	13.27			12.71	14.24	14.90	13.27

Note. N = 104. SB IV scores include Verbal Reasoning (VR), Abstract/Visual Reasoning (A/VR), Quantitative Reasoning (QR), and Short-Term Memory (STM). SB5 scores include Knowledge (KN), Fluid Reasoning (FR), Visual-Spatial Processing (VS), Quantitative Reasoning (QR), and Working Memory (WM).

Test Comparability

In addition to the correlations between the SB5, the SB IV, and Form L-M, estimated equating tables were developed to show the relationship between scores on these editions of the Stanford-Binet. The SB5 was administered in counterbalanced order to two samples; one sample was also administered the SB IV and the other Form L-M. Angoff's (1984) design II.A.1 was used to equate the different forms of the test. Tables 9 and 10 show the results of these analyses, with the range of SB5 scores expected for given IQs from the SB IV and Form L-M.

Table 9

Estimated Equating Table: Expected SB5 Full Scale IQ Ranges for Selected SB IV Composite SAS Scores

SB IV Composite SAS Score	SB5 FSIQ
55	52–60
70	67–73
85	82–85
100	96–99
115	106–113
130	122–128
145	135–143

Abbreviation: FSIQ = SB5 Full Scale IQ Score.

Table 10

Estimated Equating Table: Expected SB5 Full Scale IQ Ranges for Selected SB Form L-M IQ Scores

SB Form L-M IQ Scores	SB5 FSIQ
55	63–74
70	75–82
85	86–91
100	96–100
115	105–110
130	114–121
145	122–133

Shared Test Content

Comparisons between the exact content shared in the 1916 through 1973 versions of the Stanford-Binet are somewhat difficult to quantify due to the occurrence of items at multiple levels (with different passing criteria). Moreover, many items on the earlier forms of the test were grouped into testlets, and in some cases only parts of a testlet are shared across forms. To present a parsimonious description of shared items over time, the percent of content shared with other forms is presented for the SB IV and SB5, both of which present individual items rather than groups of items. Interested readers are referred to the Form L-M manual for a complete content map of the 1937 and 1960/1973 forms of the test (Terman & Merrill, 1973). For both the Fourth and Fifth Editions, all test items were placed in a spreadsheet. Each item was checked against the content of all prior editions, and matching items were marked. Once all items were coded, the total number of items in each edition, as well as the total number of shared items overall, were counted. Table 11 shows the results of this process.

Table 11

Percent of SB IV and SB5 Items Appearing in Other Versions of the Stanford-Binet

	SB5	SB IV	L-M	L	M	1916	Binet	Total
SB5								
Shared Items		42	27	19	20	11	2	69*
Total Items		284	284	284	284	284	284	284
Percent		14.8%	9.5%	6.7%	7.0%	3.9%	0.7%	24.3%
SB IV								
Shared Items	42		32	28	14	14	0	75*
Total Items	460		460	460	460	460	460	460
Percent	9.1%		7.0%	6.1%	3.0%	3.0%	0.0%	16.3%

Note. *The total number of shared items is the number of unique items coming from prior editions. All items shared with the SB Form L-M, for example, would also have appeared on Form L and/or Form M.

Test Structure

While early versions of the Stanford-Binet yielded only a general intelligence score, the items included a diverse mix of mental abilities. For example, items for age 4 of the 1916 edition contained visual-spatial (copying a square), quantitative (counting 4 pennies), knowledge (comprehension), and memory (repeating 4 digits) items. Although the types of items at any level of the 1916 to 1973 versions varied, almost all items on these forms relate to knowledge, fluid reasoning, visual-spatial processing, quantitative reasoning, or short-term/working memory. The Fourth Edition separated the diverse types of items found on prior editions into subtests and factors, grouping Visual-Spatial Processing and Fluid Reasoning together into an Abstract/Visual Reasoning factor. The Fifth Edition also maintains these five predominant areas of mental ability with a factor structure that keeps Fluid Reasoning items separate from Visual-Spatial Processing items. The Fifth Edition also places more emphasis on working memory compared with short-term memory. While working memory was present to some extent on earlier editions, it was limited to Memory for Digits Reversed, with all other memory tasks involving short-term memory.

Table 12 presents a count of the types of items (using the Fifth Edition categories) that appeared on prior versions of the Stanford-Binet. These classifications are based on a content analysis of prior editions as well as on empirical studies (e.g., Woodcock, 1990). Of course, content analysis does not necessarily ensure that an item will load on a particular factor in a factor analysis, or that all practitioners will agree on the classifications given.

Additionally, a number of items and subtests fall across factors (for example, Knowledge and Visual-Spatial Processing contribute to Mutilated Pictures in the Form L-M). For the purpose of Table 12, items with multiple possible classifications are counted in both classes.

Table 12

Item Content on the 1916 to 1973 Editions of the Stanford-Binet

	SB5	SB IV	Form L-M	Form L	Form M	1916
Knowledge	26%	27%	40%	41%	38%	38%
Fluid Reasoning	17%	14%*	20%	17%	21%	17%
Visual-Spatial Processing	18%	17%	18%	16%	15%	11%
Quantitative Reasoning	21%	17%	6%	6%	6%	8%
Working Memory	12%	2%	4%	3%	4%	9%
Short-Term Memory	6%	22%	9%	13%	11%	11%
Other**	0%	0%	4%	3%	5%	6%

Note. *Verbal Relations is counted in Fluid Reasoning as well as Knowledge. Number Series is counted in Fluid Reasoning as well as Quantitative Reasoning.

**These items cover such abilities as auditory processing, long-term retrieval, and judgements of weights.

Conclusions

The *Stanford-Binet Intelligence Scales, Fifth Edition* represents the latest in a series of enhancements derived from the tradition of intelligence testing originated in 1905 by Alfred Binet and Theodore Simon. However, as Kuhn (1970) points out, the real progress in a science, and certainly the real progress in the fruits of a science, lies not in the revolutionary period but rather in the periods of normal science that follow it. According to Kuhn, “the results gained in normal research are significant because they add to the scope and precision with which the paradigm can be applied” (p. 36). The SB5 incorporates many insights implicitly designed into the early editions of the measure as implemented by Binet, Simon, Terman, and Merrill, but presents them in a way that provides vast practical improvements in the areas of content coverage and psychometric characteristics. In this way, the revolutionary work of the earlier authors has shaped the more recent enhancements and advancement of the test under Thorndike, Hagen, and Sattler, and most recently, Roid.

References

Angoff, W. H. (1984). *Scales, norms, and equivalent scores*. Princeton, NJ: Educational Testing Service.

Binet, A., & Simon, T. (1905). Méthodes nouvelles pour le diagnostic du niveau intellectual des anormaux. *L'Année psychologique, 11*, 191–336.

Binet, A., & Simon, T. (1916). *The development of intelligence in children* (E. Kit, Trans.). Baltimore, MD: Williams & Wilkins.

Buros, O. K. (1938). *The 1938 mental measurements yearbook*. New Brunswick, NJ: Rutgers University Press.

Delaney, E. A., & Hopkins, T. F. (1987). *Examiner's handbook: An expanded guide for fourth edition users*. Itasca, IL: Riverside Publishing.

Glaub, V. E., & Kamphaus, R. W. (1991). Construction of a nonverbal adaptation of the Stanford-Binet Fourth Edition. *Educational & Psychological Measurement, 51*, 231–241.

Herring, J. P. (1922). *Herring revision of the Binet-Simon tests and verbal and abstract elements in intelligence examinations*. New York: World Book Co.

Kuhlmann, F. (1912). *A revision of the Binet-Simon system for measuring the intelligence of children*. Faribault, MN: Minnesota School for Feeble-Minded and Colony for Epileptics.

Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press.

McNemar, Q. (1942). *The revision of the Stanford-Binet Scale: An analysis of the standardization data*. New York: Houghton Mifflin Company.

Melville, N. J. (1917). *Testing juvenile mentality*. Philadelphia: J. B. Lippincott Company.

Minton, H. L. (1988). *Lewis M. Terman: Pioneer in psychological testing*. New York: New York University Press.

Pratt, C. C. (1917). Book review: The measurement of intelligence. *Journal of Applied Psychology, 1*, 191–192.

Reckase, M. D. (1989). Adaptive testing: The evolution of a good idea. *Educational Measurement: Issues & Practice, 8*, 11–15.

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

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

Sattler, J. M. (1965). Analysis of functions of the 1960 Stanford-Binet Intelligence Scale, Form L-M. *Journal of Clinical Psychology, 21*, 173–179.

Terman, L. M. (1916). *The measurement of intelligence: An explanation of and a complete guide for the use of the Stanford revision and extension of the Binet-Simon Intelligence Scale*. Boston: Houghton Mifflin.

Terman, L. M., & Merrill, M. A. (1937). *Measuring intelligence*. Boston: Houghton Mifflin.

Terman, L. M., & Merrill, M. A. (1960). *Stanford-Binet Intelligence Scale: Manual for the Third Revision Form L-M*. Boston: Houghton Mifflin.

Terman, L. M., & Merrill, M. A. (1973). *Stanford-Binet Intelligence Scale: Manual for the Third Revision Form L-M* (1972 Norm Tables by R. L. Thorndike). Boston: Houghton Mifflin.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Stanford-Binet Intelligence Scale: Fourth Edition*. Itasca, IL: Riverside Publishing.

Wechsler, D. (1991). *The Wechsler Intelligence Scale for Children, Third Edition*. San Antonio, TX: Psychological Corporation.

Wolf, T. H. (1973). *Alfred Binet*. Chicago: University of Chicago Press.

Woodcock, R. W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment, 8*, 231–258.

Yerkes, R. M. (1917). The Binet versus the point scale method of measuring intelligence. *Journal of Applied Psychology, 1*, 111–122.

Riverside Publishing

A HOUGHTON
MIFFLIN COMPANY

425 Spring Lake Drive
Itasca, IL 60143-2079

1.800.323.9540
www.stanford-binet.com

9-96306