

OVERVIEW OF THE SB5 AND ITS HISTORY

INTRODUCTION

New editions of nationally standardized tests provide modern wording, illustrations, enhanced measurement procedures, updated theory and research, and new standardizations, enhancing the validity of test interpretations. Such enhancements come at a price, however, in time, effort, and costs to the developers and consumers. After a 7-year revision project, the Stanford-Binet Intelligence Scale, Fifth Edition (SB5) (Roid, 2003b) was published with enhanced features, norms, and procedures. Initial reviews and receptions by users have shown that the revision project was successful in many respects, to be described in the chapters that follow.

The SB5 combines the point-scale format of the fourth edition (SB4) by Thorndike, Hagen, and Sattler (1986) with the age-level format found in previous editions such as the classic Forms L, M, and L-M (Terman & Merrill, 1937, 1960). Examiners begin a standard test administration by giving the two routing subtests: Object Series/Matrices (nonverbal) and Vocabulary (verbal). Estimates of ability in the nonverbal and verbal domains are obtained from raw scores on each routing test and used to tailor the remaining assessment to the examinee's functional ability. Simple conversion tables show the examiner which functional levels (ranging from easy to hard, Levels 1 through 6) of the nonverbal and verbal scales to continue testing. By adding the non-

DON'T FORGET

The SB5 combines the point-scale format used in the SB4 and many intelligence tests with the functional levels design of the Terman and Merrill (1960) Form L-M. The purpose of this combination is to tailor the test to the ability level of the examinee—a critical feature that improves measurement for individuals with disabilities and for highly gifted children.

verbal routing test and an entire one-half of the SB5 in the nonverbal domain, the new edition provides excellent features for testing individuals with limited English or with communication difficulties. The nonverbal section requires a low language demand—that is, minimal receptive language and mostly nonverbal responses (pointing, moving pieces, etc.) by the examinee. Also, the two-stage testing

procedure, with routing subtests first and functional-level sections next, provides highly precise estimates of cognitive ability in a relatively short period of time.

Thus, the SB5 is a wide-ranging, individually administered test battery. Norms were designed for ages 2 through 85+ years and the subtests cover five cognitive factors—Fluid Reasoning, Knowledge (crystallized ability), Quantitative Reasoning, Visual-Spatial Processing, and Working Memory—in both the verbal and nonverbal domains. Importantly, the SB5 is the first intellectual battery to cover five cognitive factors in both the nonverbal and verbal domains. Five nonverbal subtests and five verbal subtests measure each of the factors.

Many new features have been added to the SB5 and features of previous editions enhanced, as shown in Rapid Reference 1.1. Many of the new features were designed to enhance the usefulness of the SB5 for assessments with preschool children, individuals with mental retardation, and individuals with intellectual giftedness. In addition, many features were added to make the test easy for examiners to administer and score.

Rapid Reference 1.1

New Features of the SB5 Compared to Previous Editions

- Brightly colored toys, blocks, and pictures to enhance preschool assessment.
- New composite scores (IQ and Factor Index) with a mean of 100 and standard deviation of 15, and subtests with a mean of 10 and standard deviation of 3.
- Extended low-end items for early childhood assessment and high-end items for giftedness assessment, including an innovative Extended IQ measuring IQ down to 10 and up to 225.
- An abbreviated IQ based on the two routing subtests.
- Retention of a few classic Binet items such as Picture Absurdities; new scales such as block tapping, recall of the last words in a series of sentences (measuring working memory processes; Baddeley, 1986), and formation of designs using an expanded set of form-board pieces as a replacement for block designs.
- Change-Sensitive Scores, providing criterion-referenced interpretation of scores based on item response theory.
- Linkage to the Woodcock-Johnson III Tests of Achievement for assessment of learning disabilities.
- Interpretation of differences between nonverbal and verbal abilities within each factor to identify strengths and weaknesses in the individual's profile of abilities.

HISTORY

A number of the features of the design of the SB5 become apparent when the history of the Stanford-Binet and its various editions is reviewed. Developments began in France at the turn of the 20th century. Alfred Binet and Theodore Simon (1905) developed a useful tool to assess general intelligence, which is widely cited as the first major break-

through in intelligence testing. As a member of a French governmental commission working on mental retardation, Binet developed a practical test, sensitive to different levels of cognitive development, that could be given during a clinical interview. Binet's early work is summarized in Rapid Reference 1.2.

Terman's 1916 Stanford Revision

Researchers in the United States, such as Goddard (1908) and Terman (1911), quickly saw the theoretical and practical value of Binet's work and began to adapt the work to the American context. Lewis Terman, of Stanford University, worked with Child (Terman & Child, 1912) to develop a preliminary revision of the Binet-Simon scale. Within a few years, Terman (1916) had extended the scale and collected data on more than 2,300 children and adolescents. The improved scale was published as the *Stanford Revision and Extension of the Binet-Simon Scale* by a division of Houghton-Mifflin Company called Riverside Press. Alternative versions of the Binet-Simon scales that presented intelligence as a singular dimension were distributed in the United States (e.g., the Goddard, 1910 version). However, Terman's 1916 revision retained Binet's concept of intelligence as a complex mixture of abilities, and is the only revision that has stood the test of time, remaining in publication to the present day. The standardization that Terman accomplished was quite rigorous for the early 1900s and increased the scale's technical quality. Also, the scale had thorough directions for examiners and used the ratio of mental age to chronological age first introduced by Stern (1914). This intelligence quotient (IQ) became the new standard for the assessment of intelligence.

Revisions of the Terman Scales in 1937, 1960, and 1972

Within 20 years of its release in 1916, the Stanford Revision emerged as the most widely used test of intellectual ability in America. The scale

Rapid Reference 1.2

History of the Early Work of Binet

Alfred Binet (1857–1911) authored almost 300 books, articles, and reviews during his career. His work began with intelligence testing, when Binet collaborated with Victor Henri (1872–1940) to outline a project for the development of a series of mental tasks to measure individual differences (Binet & Henri, 1895). The tasks were designed to differentiate a number of complex mental faculties, including memory, imagery, imagination, attention, comprehension, aesthetic sentiment, moral sentiment, muscular strength, motor ability, and hand-eye coordination.

The 1905 Binet-Simon Scale

Binet was named a member of a government educational commission and took the lead in devising a useful and reliable diagnostic system for identifying children with mental retardation. Binet's project culminated in the publication of the first practical intelligence test (Binet & Simon, 1905) with physician Theodore Simon (1873–1961).

Binet sought to make the 1905 scale efficient and practical: "We have aimed to make all our tests simple, rapid, convenient, precise, heterogeneous, holding the subject in continued contact with the experimenter, and bearing principally upon the faculty of judgment" (from the Kite translation, Binet & Simon, 1916). The scale consisted of 30 items, which were scored on a pass-fail basis. The items presented various word problems, paper-cutting tasks, repeating sentences and digits, and comparing blocks to put them in order by weight (Wolf, 1969). Combinations of mental and physical strategies were required in solving each item or mastering each task. The complex nature of these practical items is generally recognized as the major breakthrough that allowed intelligence to be assessed during a clinical interview. Prior to Binet's use of complex items and tasks, mental abilities were measured in isolated fashion with laboratory equipment and included visual perception, reaction time, hearing acuity, and other physical measurements (e.g., Galton, 1883, as described in Johnson, McClearn, Yuen, Nagoshi, Ahern, & Cole, 1985).

(continued)

The 1905 scale included several important innovations that would be used in subsequent measures of intelligence. Items were ranked in order of difficulty and accompanied by careful instructions for administration. Binet and Simon also utilized the concept of age-graded norms (Wolf, 1973). The use of age-graded items allowed the scale to estimate mental age by the pattern of correct answers.

The 1905 Binet-Simon Scale was revised in 1908 (Binet & Simon, 1908) and again in 1911. By the completion of the 1911 edition, Binet had extended the scales through adulthood and balanced them with five items at each age level. The scales included procedures for assessing language (e.g., receptive naming and expressive naming in response to visual material, repeating sentences, and defining familiar objects), auditory processing (e.g., word rhyming), and visual processing (e.g., rapid discrimination of lines, and drawing the unfolded design of a folded paper with cut portions). Also included were tasks to assess learning and memory (e.g., repeating prose passages, repeating phrases and sentences of increasing length, drawing two designs from memory, recalling the names of pictured objects, and repeating numbers), and judgment and problem solving (e.g., answering problems of social and practical comprehension, giving distinctions between abstract terms).

Source: Roid (2003f). Reproduced from the *Stanford-Binet Fifth Edition, Technical Manual* by Gale H. Roid, Ph.D., with permission of the publisher. Copyright © 2003 by The Riverside Publishing Company. All rights reserved.

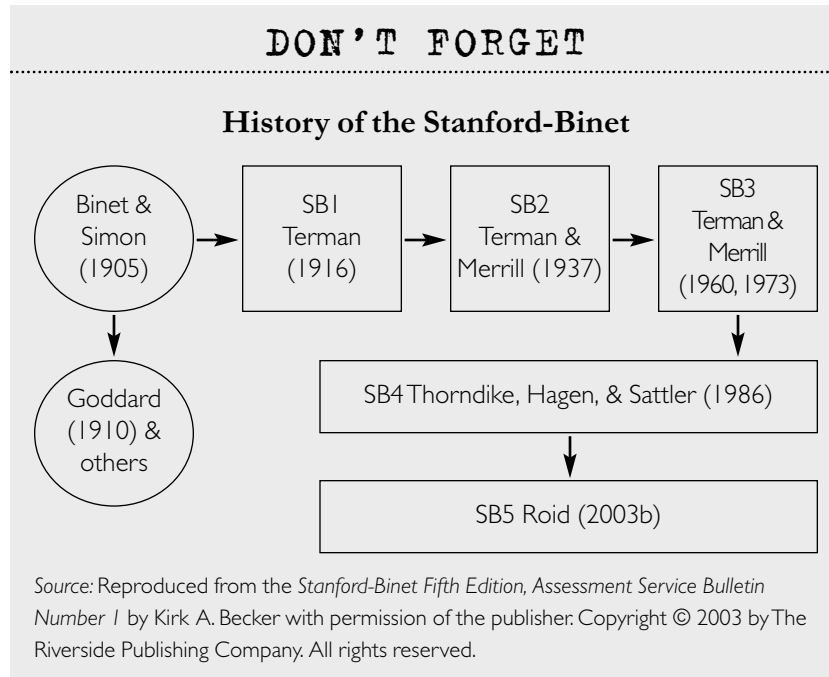
had several language translations and was used internationally. In subsequent years, Terman continued to experiment with easier and more difficult items to extend the measurement scale downward and upward and to increase the age range by collecting more standardization cases. As was the practice in test publishing at that time, an alternative form of the scale was also developed. Funding came from Stanford University grants and help came from colleague Maud Merrill, who served as codirector of the revision project. Interestingly, the new edition took 7 years to create—the same length of time devoted to the new fifth edition (Roid, 2003f)—and was called the *New Revised Stanford-Binet Tests of Intelligence* (Terman & Merrill, 1937).

The 1937 revision was standardized on 3,200 examinees aged 1 year

6 months to 18 years. Terman made efforts to include a broader representation of geographic regions and socioeconomic levels in the normative sample. Two alternative forms, Form L and Form M, were included. Improvements over the 1916 edition included greater coverage of nonverbal abilities, less emphasis on recall memory, extended range of the scale at the lower and upper ends, and more objectified scoring methods. The descriptions of the norming methods employed on Forms L and M and use of the *Hollerith sorter* (an early data-processing machine that sorted coded punch cards) for statistical analyses are quite impressive and interesting as examples of early psychometric methods (Terman & Merrill, 1937, p. 22).

As happens with any widely used test of ability or achievement, certain items of the 1937 edition became dated by the 1950s, and some of the toys used with the test had become difficult to obtain after World War II. Therefore, Terman and Merrill began a further revision of the scale based on the accumulated information and data collected since 1937. According to the preface in Terman and Merrill (1973, p. vi), “plans for the third revision had been formulated and were well on their way” when Terman died in 1956. Thus, the *Stanford-Binet Intelligence Scale, Third Revision* (Terman & Merrill, 1960) was published. Several new features were included in the third revision (now called the third *edition*), including use of the deviation IQ (standardized normative mean of 100 and standard deviation of 16) rather than the ratio IQ, and production of a combined Form L-M consisting of the most discriminating 142 items from the 1937 forms. Extensive item analyses had been conducted on data from nearly 4,500 subjects from the 1930s to 1950s and any items showing substantial change in difficulty were eliminated or adjusted in sequence. No newly created items were added to Form L-M. Six items were placed at each age level of the combined form (reduced from its previous eight items per level) and some directions were clarified.

After Maud Merrill retired, Robert L. Thorndike of Columbia University was asked to lead a project to collect new norms for the third edi-



tion. Thus, the same edition was reprinted with the new normative tables—an update of Form L-M (Terman & Merrill, 1973). Because the *Cognitive Abilities Test* (CogAT; Thorndike & Hagen, 1994) was being standardized at the same time as the 1972 renorming of the Stanford-Binet, Thorndike selected subjects and some siblings of subjects tested on the CogAT to compose the new norm sample. The stratification variables used on the sample (e.g., age, geographic region, ethnicity, community size) were similar to those used today, as were the levels of ability on the verbal portion of the CogAT. The items in the test remained essentially the same as on the 1960 revision, with two minor exceptions.

The 1986 Edition by Thorndike, Hagen, and Sattler

With the SB4 (Thorndike et al., 1986), the test took on a new appearance and structure. The SB4 was based on a four-factor, hierarchical

model with general ability (*g*) as the overarching summary score. The four cognitive factors were Verbal Reasoning, Abstract/Visual Reasoning, Quantitative Reasoning, and Short-Term Memory. The most significant change from previous editions, however, was the use of point scales for all subtests rather than the developmental age levels used in previous forms. Vocabulary was still retained as a routing test, allowing the test to be tailored to the examinee's verbal ability. Also, many classic Stanford-Binet tasks were retained, including Absurdities, Vocabulary, Matrices, Quantitative Reasoning, and Memory for Sentences—tasks also included in the SB5. Composite scores for each cognitive factor and profile scores for each subtest allowed a comprehensive examination of strengths and weaknesses among abilities within general intelligence.

CAUTION

As with all published measures of intellectual abilities, the SB5 measures a fixed number (five) of all the multiple dimensions of cognitive abilities and skills that have been researched over the last 100 years. Examiners should be cautious in claiming that all cognitive abilities have been assessed in any one examination or any one test. Also, be cautious in describing the Full Scale IQ as an errorless, fixed entity that summarizes the full potential of the individual.

THEORY AND STRUCTURE OF THE SB5: THE 2003 FIFTH EDITION

Based on the important research of Carroll (1993), the SB5 was constructed on a five-factor hierarchical cognitive model. The five factors were derived from the combined models of Carroll, Cattell (1943), and Horn (1965). The combination of models, now called the Cattell-Horn-Carroll (CHC) theory, normally lists 8 to 10 factors. Many of the supplemental factors, such as processing speed, auditory processing, and long-term retrieval, require specialized timing or test apparatus (e.g.,

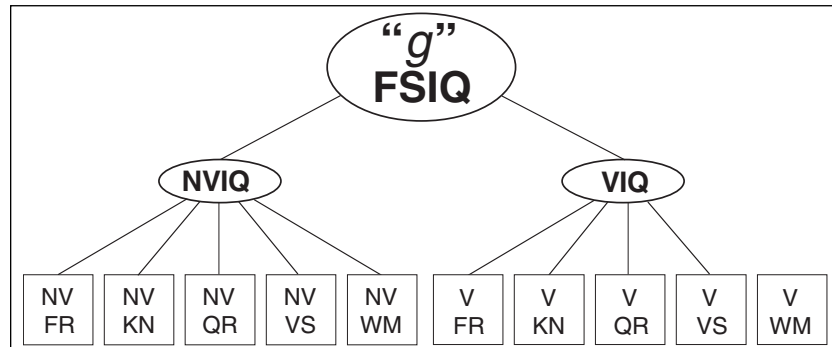


Figure 1.1 The Hierarchical Structure of the SB5 Scoring System

Notes: FSIQ = Full Scale IQ; *g* = general ability; NVIQ = Nonverbal IQ; VIQ = Verbal IQ; FR = Fluid Reasoning; KN = Knowledge; QR = Quantitative Reasoning; VS = Visual-Spatial Processing; WM = Working Memory.

tape recorders). However, the five cognitive factors of the SB5 (see Fig. 1.1) were selected based on research on school achievement and on expert ratings of the importance of these factors in the assessment of reasoning, especially in giftedness assessment. Also, the memory factor was shifted from an emphasis on short-term memory only, as in the fourth edition of the Stanford-Binet, to an emphasis on Working Memory. Therefore, the overall model shown in Figure 1.1 is a hierarchical *g* model with five factors emphasizing reasoning abilities that can be easily administered within a one-hour assessment. The important change in emphasis in the fifth edition is the duplication of all five cognitive factors in both the nonverbal and verbal domains.

The structure of the SB5 is also shown in Figure 1.1. The verbal and nonverbal domains include five subtests each, for a total of 10 profile scores. The nonverbal subtests require a small degree of receptive language and allow for pointing responses, the movement of puzzle-like pieces, and manipulation of toys to indicate correct answers. The verbal subtests require facility with words and printed material (reading or speaking). A description of each of the 10 subtests is provided in Rapid Reference 1.3.

Rapid Reference 1.3

Subtests of the SB5

Subtest	Description (activities at various levels)
<i>Nonverbal</i>	
Fluid Reasoning	Object Series/Matrices (a point scale used for routing). Includes new sequential reasoning items and classic matrices.
Knowledge	Procedural Knowledge (a new type of item involving gestures), followed by Picture Absurdities (a classic subtest in the Stanford-Binet tradition).
Quantitative Reasoning	Nonverbal Quantitative Reasoning items, tapping number concepts, problem solving, and figural-geometric/measurement-estimation problems.
Visual-Spatial Reasoning	Form Board (classic items for the lower levels), followed by the new Form Patterns (making designs from an expanded set of form-board pieces).
Working Memory	Delayed Response (e.g., hiding an object under a cup) at the low levels followed by Block Span (the new block-tapping procedure).
<i>Verbal</i>	
Fluid Reasoning	Early Reasoning items (e.g., picture reasoning) followed by classic Verbal Absurdities and Verbal Analogies.
Knowledge	Vocabulary (a point scale used for routing). Includes toys, identification of body parts, Child Card, and classic word definitions.
Quantitative Reasoning	Verbal Quantitative Reasoning items, tapping number concepts, problem solving, and figural-geometric/measurement-estimation problems.
Visual-Spatial Reasoning	Innovative new Position and Direction (verbal-spatial problems requiring explanation of directions, identifying spatial relations in pictures, understanding complex statements of spatial orientation).
Working Memory	Classic Memory for Sentences followed by an innovative Last Word procedure (requiring memory of the last word of series of questions).

Psychometric Properties

Extensive studies of reliability, validity, and fairness were conducted as part of the SB5 standardization. The SB5's main technical features are briefly outlined here. The normative sample for the SB5 included 4,800 subjects aged 2 to 96 years. The highest age grouping employed in the norm tables was 85+. The composition of the normative sample closely approximated the stratification percentages reported by the U.S. Census Bureau (2001). Stratification variables included gender, geographic region, ethnicity (African-, Asian-, and Anglo/Caucasian-American; Hispanic; Native American; and Other), and socioeconomic level (years of education completed, or parent's educational attainment). Additionally, subjects were tested ($N = 1,365$) from officially documented special groups such as individuals with mental retardation, learning disabilities, attention deficit, and speech or hearing impairments.

Internal-consistency reliability ranged from .95 to .98 for IQ scores and from .90 to .92 for the five Factor Index scores. For the 10 subtests, average reliabilities (across age groups) ranged from .84 to .89, providing a strong basis for profile interpretation (see Rapid Reference 1.4). Split-half reliability formulas were used for subtests and composite reliabilities for IQ and Factor Index scores.

Test-retest and interexaminer reliability studies were also conducted and showed the stability and consistency of SB5 scoring.

Evidence for content-, criterion-, and construct-related validity of the SB5 was detailed in Roid (2003f), including extensive studies of concurrent, predictive, and factorial validity. Also, good evidence of consequential validity and fairness of predicting achievement were reported in Roid (2003d). Examples of validity, including the correlations with other assessment batteries, are shown in Rapid Reference 1.5. The correlations shown are quite substantial and similar in magnitude to the concurrent correlations observed for other major intelligence batteries.

Rapid Reference 1.4

Average Reliability of SB5 Scores

	Split-Half Reliability	Standard Error of Measurement
Full Scale IQ	.98	2.30
Nonverbal IQ	.95	3.26
Verbal IQ	.96	3.05
Abbreviated IQ	.91	4.55
<i>Favorite Index Scores</i>		
Fluid Reasoning	.90	4.85
Knowledge	.92	4.36
Quantitative Reasoning	.92	4.33
Visual-Spatial Processing	.92	4.41
Working Memory	.92	4.62
<i>Nonverbal Subtests</i>		
Fluid Reasoning	.86	1.18
Knowledge	.85	1.18
Quantitative Reasoning	.86	1.14
Visual-Spatial Processing	.87	1.12
Working Memory	.88	1.07
<i>Verbal Subtests</i>		
Fluid Reasoning	.86	1.18
Knowledge	.89	1.01
Quantitative Reasoning	.87	1.09
Visual-Spatial Processing	.88	1.07
Working Memory	.84	1.22

Source: Roid (2003f).

*Rapid Reference 1.5***Correlations of SB5 Full Scale IQ (FSIQ)
With Other Tests**

Test	Correlation	N
Stanford-Binet Intelligence Scale, Fourth Edition (SB4), Composite	.90	104
Stanford-Binet Form L-M, IQ	.85	80
Wechsler Intelligence Scale for Children– Third Edition (WISC-III), FSIQ	.84	66
Wechsler Adult Intelligence Scale– Third Edition (WAIS-III), FSIQ	.82	87
Wechsler Preschool and Primary Scale of Intelligence–Revised (WPPSI-R), FSIQ	.83	71
Woodcock-Johnson III Tests of Cognitive Abilities, five factors	.90	145
Woodcock-Johnson III Tests of Achievement		
Broad Reading Skills	.66	472
Broad Math	.76	472
Wechsler Individual Achievement Test–Second Edition		
Reading	.67	80
Mathematics	.79	80

Source: Reproduced from the *Stanford-Binet Fifth Edition, Technical Manual* by Gale H. Roid, Ph.D., with permission of the publisher. Copyright © 2003 by The Riverside Publishing Company. All rights reserved.

The substantial predictive correlations between the SB5 and two major achievement batteries—the Woodcock-Johnson III (WJ III) and the Wechsler Individual Achievement Test–Second Edition (WIAT-II)—provide a strong basis for comparing intellectual and achievement scores of individuals.

Extensive studies of the factor structure of the SB5 were conducted, including confirmatory factor analyses using LISREL 8.3 (Joreskog & Sorbom, 1999). The factor analyses were calculated for five successive age groups (2–5, 6–10, 11–16, 17–50, and 51+) comparing factor models with one, two, three, four, and five factors. Split-half scores (scores for odd- and even-numbered items in each of the 10 subtests) were employed to provide more stable estimates of each factor in the maximum-likelihood analyses. The five-factor models showed superior fit, including the non-normed fit index (NNFI), ranging from .89 to .93; the comparative fit index (CFI), ranging from .91 to .93; and the root mean square error of approximation (RMSEA), ranging from .076 to .088. A second series of confirmatory analyses was conducted with LISREL using conventional full-length subtests across two batteries: the SB5 and the WJ III (Woodcock, McGrew, & Mather, 2001b). Again, the five-factor model showed the best fit, with alignment of the fluid, knowledge, quantitative, visual, and memory factors across the SB5 and WJ III as predicted.

COMPREHENSIVE REFERENCES ON THE TEST

Ethical guidelines and details of test administration, scoring, and interpretation of the SB5 are covered in the *Examiner's Manual* (Roid, 2003c). Details of the development and standardization; characteristics of the normative sample; and extensive studies of reliability, fairness, and validity are presented in the *Technical Manual* (Roid, 2003f). Additional statistical analyses, useful tables, and case studies are presented in the *Interpretive Manual* (Roid, 2003d) and in Roid and Pomplun (in press). Rapid Reference 1.6 provides basic information about the SB5 and its publisher.

Rapid Reference 1.6

Description of Stanford-Binet Intelligence Scale, Fifth Edition (SB5)

Author: Gale H. Roid

Publication Date: 2003

What the test measures: Nonverbal, verbal, and general intellectual abilities

Age Range: 2 to 85+ years

Administration Time: Entire battery 45 to 75 min, depending on the examinee. Nonverbal or verbal sections separately, 30 min each. Abbreviated battery (two subtests), 15–20 min.

Qualification of Examiners: Graduate or professional-level training in psychological assessment.

Publisher: Riverside Publishing
420 Spring Lake Drive
Itasca, IL 60143
(800) 767-8420



TEST YOURSELF



1. **How many routing subtests are there on the SB5, and what are their names?**
 - (a) One—Vocabulary
 - (b) One—Object Series/Matrices
 - (c) Two—Vocabulary and Nonverbal Visual-Spatial
 - (d) Two—Vocabulary and Object Series/Matrices
2. **What is the main advantage of the two-stage testing in the SB5 (routing subtests followed by functional levels)?**
 - (a) Reduced emphasis on speed of response
 - (b) Greater measurement precision in a shorter time
 - (c) Higher predictive validity
 - (d) Ease of test administration for the examiner

3. Name the author(s) and the publication date of the first version of the Stanford-Binet published in America.

- (a) Binet & Henri, 1895
- (b) Stern, 1910
- (c) Terman, 1916
- (d) Terman & Merrill, 1937

4. Which edition of the Stanford-Binet replaced the original intelligence quotient (ratio IQ) with the deviation IQ?

- (a) Second, 1937
- (b) Third, 1960
- (c) Fourth, 1986
- (d) Fifth, 2003

5. How many cognitive factors, domains, and subtests does the SB5 contain?

- (a) Four factors; one general domain; 8 subtests
- (b) Four factors; quantitative and verbal domains; 12 subtests
- (c) Five factors; nonverbal and verbal domains; 8 subtests
- (d) Five factors; nonverbal and verbal domains; 10 subtests

6. What is the range of subtest average reliability (split-half) on the SB5?

- (a) .74 to .79
- (b) .84 to .89
- (c) .90 to .92
- (d) .95 to .98

7. What are the age range and approximate administration time of the full-battery SB5?

- (a) 2 to 21 years, 30 to 45 min
- (b) 6 to 16 years, 50 to 75 min
- (c) 18 to 74 years, 60 min
- (d) 2 to 85+ years, 45 to 75 min

(continued)

8. Which edition of the Stanford-Binet introduced a cognitive model with hierarchical g and four primary cognitive factors?

- (a) Second, 1937
- (b) Third, 1960
- (c) Fourth, 1986
- (d) Fifth, 2003

9. What was the size of the normative sample of the SB5 and the age of the oldest subject tested?

- (a) 4,800, 96 years
- (b) 3,560, 85 years
- (c) 2,200, 74 years
- (d) 1,365, 69 years

10. Which is an example of a classic type of Stanford-Binet item that was retained in the fifth edition?

- (a) Picture Absurdities
- (b) Form Patterns
- (c) Last Word
- (d) Block Span

Answers: 1. d; 2. b; 3. c; 4. b; 5. d; 6. b; 7. d; 8. c; 9. a; 10. a