Stanford-Binet Intelligence Scales: Fifth Edition (SB5)

By Gale H. Roid (Riverside Publishing, 2003)

Age range: 2 to 85+
Administration time: 45 to 75 minutes (full test); 15 to 20 minutes (abbreviated)
Cost: $858 (Complete kit with canvass bag) ; $1,014 (Complete kit with canvass bag & scoring software)

Reviewed by Stefan C. Dombrowski, Christine DiStefano, & Kelly Noonan

The Stanford-Binet Intelligence Scale was last revised in 1986. The latest revision resulted in changes to the test layout, norming standards and the theoretical structure underlying the instrument. The revised instrument is creatively designed, well-organized and aligned with Cattell-Horn-Carroll theory. The current review offers an investigation of the theoretical and psychometric foundation of the test as well as information concerning test administration, scoring and reporting.

Description of the SB5

The Stanford-Binet 5th Edition (SB5) is an individually administered assessment of intellectual ability. The SB5 has been revised to align with the Cattell-Horn-Carroll (CHC) model of cognitive abilities and includes 10 subscales (Roid, 2003). The material has been organized to provide measurement of three areas: 1) general cognitive functioning, or 'g'; 2) verbal and nonverbal intelligence; and 3) five CHC factors stratified along a verbal/nonverbal dimension. The SB5 subtests are comprised of "testlets," brief mini-tests at each level of difficulty that form the scale for the subtests. Testlets typically have six items or a total of six points at each of the ability levels for a given subtest. In turn, subtests, including the two routing subtests, are designed to measure five CHC factors: Fluid Reasoning (Fluid Intelligence or Gf), Knowledge (Crystallized Knowledge or Gc), Quantitative Reasoning (Quantitative Knowledge or Gq), Visual-Spatial Processing (Visual Processing or Gv) and Working Memory (Short-Term Memory or Gsm).

The technical manual indicates that the inclusion of the nonverbal-verbal dichotomy was based on evidence for the dichotomy from recently constructed tests (e.g., Universal Nonverbal Intelligence Test) and "previous methods of verbal and nonverbal assessment" (p. 14). The manual provides instructions for using the nonverbal-verbal subscale information to measure and report a subject's ability across these two dimensions (p. 14). Finally, a measure of overall cognitive ability may be reported. The Full Scale IQ (FSIQ) score is derived from the administration of 10 subtests (five Verbal and five Nonverbal). The first two subtests (contained in Item Book 1) are routing subtests, and are used to determine start points for the remaining Nonverbal (Item Book 2) and Verbal (Item Book 3) subtests. The two routing subtests contained in Book 1 may also be used as a brief measure of intellectual ability.

Theoretical Overview

The SB5's theoretical structure has been creatively designed and packaged — it attempts to blend the hierarchical notion of 'g', the nonverbal-verbal dichotomy customary in neuropsychological theory, and several of Cattell-Horn-Carroll's cognitive abilities (see Carroll, 1993; Cattell & Horn, 1978). In the fifth edition, the theoretical structure differs substantially from the fourth edition of the Stanford Binet.

Although there are notable structural differences between the fourth and fifth editions, the technical manual did not report the use of exploratory factor analysis (EFA) to assist in instrument development. While such a major revision in theory may have warranted its use, EFA was not conducted in favor of confirmatory factor analysis (CFA). The rationale provided was that CFA is a more appropriate test since the SB5 factor structure was established *a priori* guided by contemporary intelligence theory

(e.g., Cattell-Horn-Carroll). Two points are made on that matter. First, many researchers still find EFA to be a useful complement to CFA, especially in the early stages of instrument development, as with the major revisions to the new edition of the SB5 (Floyd & Widaman, 1995). (In fact, the WISC-IV underwent a significant revision, and included EFA analyses in its technical manual.) Second, rather than just confirming one structure, CFA is often used in an "exploratory" manner, to revise a tested model based upon suggestions from the data (Jöreskog, 1993). Both EFA and CFA used in an exploratory manner should be replicated on independent samples of data to ensure that the findings are not sample specific.

Although exploratory factor analytic studies might have aided in instrument development and clarifying the SB5 factor structure for the CHC and verbal-nonverbal dichotomies, they were not reported. Instead, the manual devotes over a page (p. 108) to describing why independent reviews might not be able to replicate the described factor structure using traditional EFA methods. Further inspection illustrated that EFA was conducted, but only to show that the test was an acceptable measure of overall cognitive ability (p. 106).

Independent EFA reviews using principal axis factoring with oblique rotation were conducted using information provided by the test developers (i.e., correlations between 20 half scales). It is noted that exploratory analyses were conducted to investigate the underlying structure given the correlational information and are not a substitute for theory driven analyses. The information neither provided a clear picture of the CHC five-factor structure nor the two-factor verbal-nonverbal dichotomy. This may suggest problems with interpretation of the SB5 at the Verbal-Nonverbal level as well as at the five-factor CHC level. However, when verbal and nonverbal dimensions were separated, the five-factor CHC structure fell out clearly with EFA methodology. This may illustrate that the nonverbal subscales are in need of additional testing and refinement to ensure that these dimensions can be uniquely and accurately measured in the company of verbal dimensions. Replicated EFA analyses indicated that the SB5 is an excellent measure of a single factor, presumably 'g.'

The manual provides CFA fit information illustrating support for the SB5's five- factor CHC model over competing IQ theoretical perspectives. Multiple fit indices are reported to provide information concerning the fit of the model across different areas (Tananka, 1993). First, while the manual provides the information to support the 5 factors, many of the indices provided illustrate adequate fit of the five factor model. For the five-factor solution, RMSEA values for the tested models are at fair levels (.076-.088 range) and NNFI and CFI values are above, but close to, the generally accepted benchmark of .90, (NNFI ranges: .89–.93; CFI ranges: .91–.94) (Schumaker & Lomax, 1996).

Second, the manual provides information concerning chi-square difference tests between competing theoretical perspectives. It is important to test different theoretical conceptualizations of IQ to lend support for the CHC five-factor model (Benson, 1998). However, the models described in the text sound like different conceptualizations of IQ as illustrated by differing numbers of factors. In this situation, information from the AIC (Akaike Information Criteria) or the ECVI (Expected Cross-validation index) are better suited to compare models testing different theoretical perspectives with differing numbers of latent variables (Schumaker & Lomax, 1996) than the nested model tests provided in the manual (p. 113).

While the manual suggests that the goodness of fit for the five-factor solution is adequate, it is further noted that fit information for the two-factor verbal-nonverbal solution was not presented. This information would give test administrators greater comfort when using information from this theoretical structure to make decisions. Based on the EFA analyses conducted and the CFA support presented in the technical manual, independent reviews of the SB5 underlying factor structure are suggested.

As discussed, only five of the CHC factors have been measured by the SB5. The Technical Manual reports that the SB5 excluded the CHC factors of auditory processing (Ga), processing speed (Gs) and long-term retrieval (Glr) because of concerns about encumbering the test with additional materials (e.g., timers and response booklets), stimuli (e.g., tape recorder) and administration length. However, one wonders if the cognitive ability areas of Ga and Gs were excluded because of difficulty assessing nonverbal dimensions of auditory processing and verbal dimensions of processing speed? The SB5 might have been well served by incorporating the assessment of auditory processing and processing speed as supplemental subtests.

Technical and Psychometric Properties

Standardization. The technical manual reports that the SB5 underwent a rigorous standardization process consistent with the high level required by APA/NCME/AERA (1999) test standards. The SB5 has a mean of 100 and a standard deviation of 15 (in contrast to the prior edition standard deviation of 16). Furthermore, individual subtests are now scaled with a mean of 10 and a standard deviation of 3. The standard errors of measurement of the Full Scale IQ, the Verbal scale, the Nonverbal scale and the Abbreviated composite are 2.30, 3.05, 3.26 and 4.55 respectively. To facilitate assessment at the lower and higher range of cognitive abilities, greater breadth of item content was incorporated into the present edition.

A total of 4,800 individuals were included in the norming process. Each of the early years (ages 2 through 4) was divided into half-year groupings (doubling the sample size at the 2 to 4 year period) to account for the instability and rapid cognitive change of the very youngest age group. Using variables identified in the U.S. Census (2001), norm group individuals were stratified according race/ethnicity, sex, parental education level and geographic region. Experts knowledgeable about the following areas conducted reviews for bias: gender, ethnicity/race/linguistics (African-American, American Indian and Alaskan Native, Asian, Hispanic and Caucasian), religion (Buddhist, Christian, Jewish, Hindu and Muslim) and individuals with disabilities (hearing impaired and a general disability category). Moreover, statistical measures of Differential Item Functioning (DIF), with Rasch-based methods and Mantel-Haenszel statistics (Holland & Thayer, 1988), were calculated for each item. The data for these analyses were not included in the manual. Both expert bias reviews and statistical analyses were used to refine, select, or delete items.

Reliability. The technical manual reports strong evidence for reliability, with average internal consistency reliabilities in the range of .91 (Abbreviated) to .98 (FSIQ) at the full scale and index levels. Individual subtest reliabilities were slightly lower, ranging from 0.84 (Verbal Working Memory) to .89 (Verbal Knowledge). Both test-retest and split half (i.e., internal consistency) methods of ascertaining reliability were used. Reliability figures reported in the technical manual are appropriately high.

Validity: The relationship between the SB5 and other measures of cognitive ability and achievement are reported in the technical manual. Concurrent validity evidence is strong with a reported correlation of .90 between the SB5 FSIQ and the SB4 Composite. Concurrent validity evidence between the SB5 and other established measures is also provided in the manual. The correlations between the SB5 FSIQ and the Full Scale IQ composites on the Wechsler Preschool and Primary Intelligence — Revised (WPSSI-R), the Wechsler Intelligence Scale for Children, 3rd Edition (WISC-III) and Wechsler Adult Intelligence Scale, 3rd Edition (WAIS-III) were .83, .84 and .82, respectively.

The technical manual reports criterion related validity of .67 (Reading), .79 (Mathematics), .53 (Written Language), .77 (Oral Language) and .80 (Total Achievement) between the SB5 FSIQ and respective Wechsler Individual Achievement Test, 2nd Edition (WIAT-II) composites. The technical manual also reports a similar correlation range between the SB5 and Woodcock-Johnson Test of Achievement Clusters (3rd Edition): Basic Reading Skills (.50), Reading Comprehension (.80), Basic Math Skills (.76), Math Calculation Skills (.74), Math Reasoning (.80), Written Expression (.70) and Academic Applications (.84).

Administration and Scoring

The nonverbal and verbal subtests of the SB5 contain test items that are developmentally appropriate, engaging and creative. This serves to maintain the interest and motivation of examinees from the preschool age on up. The directions for administration and scoring of Verbal subtests are clear and easy to follow. For instance, Verbal Knowledge (Vocabulary) requires examinees to define a word that is presented in both visual and oral form. The examiner is then required to match the examinees response to scoring criteria, and circle whether the score was a 0, 1 or 2 point response. The administration and scoring of the Vocabulary subtest is typical of that required of most other verbal subtests and does not require verbatim recording or significant examiner effort to administer and score.

While the verbal subtests are easy to administer and score, some examiners might find that the SB5 Nonverbal subtests have an initially steep administration learning curve. There are a significant number of manipulatives that require a higher level of examiner dexterity. This becomes most evident when assessing younger children. At the 2 to 5 year-old age level, the nonverbal subtests will require examiners to demonstrate facility with manipulating the very child friendly materials, support the child's behavior and score the items of the subtest to determine whether a discontinue threshold has been met. The difficulty associated with administering and scoring the SB5 nonverbal subtests at younger age levels is both a function of working with preschool aged children and a function of the complexity of administering SB5 nonverbal subtests.

However, this initially steep administration learning curve is more than offset by the creatively designed, engaging and developmentally appropriate nonverbal test items. SB5 nonverbal subtests seem to be extremely appealing to children. For instance, test items from nonverbal level 1 testlet (Object series/matrices) are presented using three-dimensional geometric figures, rather than a two-dimensional test booklet. This has the dual effect of increasing the appeal of the test for younger examinees and increasing the difficulty of administration for examiners. Like the early level subtests of fluid reasoning, the nonverbal working memory and visual spatial-processing subtests require significant examiner fine motor dexterity to maintain the interest, attention and motivation of examinees. Examiners must simultaneously demonstrate facility with manipulating (i.e., taking out, returning and taking out) test materials, supporting the child's behavior and scoring the items of the testlet to determine whether a discontinue threshold has been met.

Test administration is expedited through adaptive testing via routing subtests and basal/ceiling rules. The inclusion of routing subtests continues the tradition of the SB4 adaptive testing approach. This has been expanded, however, to include a routing subtest for both the verbal and nonverbal subtests. (In prior editions, the Stanford-Binet had only a verbal routing [Vocabulary] subtest.) The administration of the SB5 is different from its predecessors as well as most other tests of cognitive ability. Test administration occurs according to ability level across all subtests (e.g., Gf, Gc, Gsm, Gvs, Gq). Thus, administration does not progress by starting and then completing a specific subtest. Rather, all testlets starting at a particular ability level are administered until a discontinue point at a subsequent ability level has been met for that subtest. This means that an examinee may reach a discontinue point at various ability levels depending upon subtest (e.g., ability level two for Gsm, ability level 3 for Gv and level 4 for Gf). Examiners will require additional time to become familiar with this style of administration. However, test administration in this fashion may serve to increase measurement precision, shorten administration time, and lessen examinee frustration with the assessment process.

Administration of the SB5 would be further facilitated with the inclusion of tabs to identify testlets. At present, the testlets are structured according to a color code (e.g., green, orange, etc.) and this requires the examiner to spend time literally thumbing through the pages of the test booklet to arrive at the appropriate testlet level. The basal and ceiling rules are initially difficult to understand, in part because of how they are phrased. For instance, the discontinue threshold is considered met following two or fewer points in a level.

Recording of responses to SB5 test items is relatively easy and consistent with the response recording criteria of other standardized intelligence measures. There may be some difficulty, however, with attempting to score nonverbal working memory test items at higher levels. Examiners must have a sufficiently high level of visual discrimination and working memory to note examinee performance and determine whether the examinee has tapped the green blocks in the correct order. The SB5 allows for both hand scoring and computerized scoring of the protocol. Hand scoring is relatively quick and easy.

Compared with the SB4, scores on the SB5 are now furnished on a more traditional scale (e.g., mean of 100; standard deviation of 15; subtest mean of 10 and standard deviation of 3). This scoring standard facilitates interpretation of data and comparison to other measures that use this scoring rubric.

Summary

The SB5 has been substantially revised and aligned with CHC theory. However, there are some questions lingering concerning the underlying factor structure reported in the technical manual which might make interpretation beyond the FSIQ level somewhat uncertain. Independent reviews of the factor structure are suggested. Younger examinees will find the SB5 engaging due to its child-friendly and developmentally appropriate materials. This also serves to increase the difficulty of administration of nonverbal items. The determination of basal and ceiling items will be initially challenging, and the administration of nonverbal testlets requires significant practice. Once beyond the administration learning curve for early level nonverbal testlets, the SB5 should be viewed as a viable test of overall cognitive ability. ◆

References

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1999). *Standards for educational and psychological testing.* Washington, DC: Author.

Benson, J. (1998). Developing a strong program of construct validation: A test anxiety example. *Educational Measurement: Issues and Practice, 17,* 10-22.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage Publications.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies.* New York: Cambridge University Press.

Cattell, R. B., & Horn, J. L. (1978). A check on the theory of fluid and crystallized intelligence with description of new subtest designs. *Journal of Educational Measurement, 15,* 139-164.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment 7,* 286-299.

Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I Braun (Eds.), *Test validity* (pp. 129-145). Hillsdale, NJ: Erlbaum.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.

Jöreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 294-316). Newbury Park, CA: Sage.

Roid, G. (2003). *Stanford-Binet Intelligence Scale: Fifth Edition.* Chicago: Riverside.

Schumaker, R. E., & Lomax, R. G. (1996). *A beginner's guide to structural equation modeling.* Mahwah, NJ: Erlbaum.

The Psychological Corporation. (2001). *Wechsler Individual Achievement test — second edition.* San Antonio, TX: Author.

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

Thorndike, R. M. (1990). Would the real factors of the Stanford-Binet Fourth Edition please come forward? *Journal of Psychoeducational Assessment, 8,* 412-435.

Wechsler, D. (1989). *Wechsler Preschool and Primary Scale of Intelligence, Revised.* San Antonio, TX: Psychological Corporation.

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

Wechsler, D. (2003). *Wechsler Intelligence Scales for Children-Fourth Edition.* San Antonio, TX: Psychological Corporation.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement.* Itasca, IL: Riverside Publishing.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Cognitive Abilities.* Itasca, IL: Riverside Publishing

 Stefan C. Dombrowski, PhD, is an assistant professor in the school psychology program at Rider University. Christine DiStefano, PhD, is an assistant professor of educational measurement and statistics at Louisiana State University. Kelly Noonan, PhD, is an assistant professor in the economics department at Rider University and specializes in health economics.