

Investigating the Theoretical Structure of the Stanford-Binet–Fifth Edition

Christine DiStefano
University of South Carolina

Stefan C. Dombrowski
Rider University

The fifth edition of the Stanford-Binet test went through significant reformulation of its item content, administration format, standardization procedures, and theoretical structure. Additionally, the test was revised to measure five factors important to intelligence across both verbal and nonverbal domains. To better understand these substantial revisions, the underlying factor structure of the instrument was investigated using both exploratory and confirmatory factor analysis procedures across five age groups tested by the publishers. Analyses were conducted using 4,800 cases included in the instrument standardization. Results suggested that the verbal/nonverbal domains were identifiable with subjects younger than 10 years of age whereas a single factor was readily identified with older age groups.

Keywords: *psychological testing; intelligence testing; exploratory factor analysis; confirmatory factor analysis*

The Stanford-Binet Intelligence Scale (SB) is one of the oldest and most widely used individual measures of intellectual ability. Since its development nearly 100 years ago, the SB has ascribed to the concept of general intelligence (*g*). More recent editions of the SB, such as the fourth edition, described factors (i.e., area scores) beyond general intelligence. However, some researchers have criticized the area scores as lacking factor analytic support (Keith, Cool, Novak, White & Pottebaum, 1988; Kline, 1989; Reynolds, Kamphaus, & Rosenthal, 1988).

The most recent version of the Stanford-Binet test, the Stanford-Binet–Fifth Edition (SB5), underwent significant reformulation of not only its item content and administration format but also its standardization procedures and theoretical structure. The test was revised to measure five factors thought to encompass intelligence across verbal and nonverbal domains. The five factors measured by the SB5 include the following: 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*).

Authors' Note: Correspondence concerning this article should be directed to Christine DiStefano, 137 Wardlaw Hall, Department of Educational Studies, University of South Carolina, Columbia, SC 29208; e-mail: distefan@gwm.sc.edu.

The inclusion of intelligence assessment across nonverbal domains is new to the fifth edition. Measurement of nonverbal domains may be useful to psychologists when assessing individuals with limited English, verbal, or communication abilities. Additionally, there exists a research base dating back to World War I that substantiates the nonverbal-verbal dichotomy (Kamphaus, 2001). Thus, the inclusion of nonverbal dimensions indicates attention to these aspects of intelligence and an effort to develop the SB5 as a comprehensive cognitive assessment.

As mentioned, the theoretical underpinning of the SB5 has been revised. The current version of the instrument provides support for using the test information in multiple ways to report and use results relative to three different factor models (Roid, 2003). Particularly, the manual states the SB5 information may be used to measure a unidimensional *g* model, a two-factor verbal and nonverbal model, and a five-factor model where both the verbal and nonverbal scales are meshed together to measure a particular dimension of intelligence. Although this redevelopment represents an improvement over the prior edition, the information supporting the internal structure of the SB5 has been questioned (Dombrowski, DiStefano, & Noonan, 2004). For example, the author of SB5 opted to rely primarily on confirmatory factor analytic procedures using prior knowledge of the theory and information from earlier editions as support for the tested models. However, placing primary reliance on confirmatory factor analytic procedures to assess the underlying theoretical structure of a test that has been substantially revised, both in concept and content, may leave some room for question. One way of managing such major revisions would be to include *both* exploratory factor analytic as well as confirmatory factor analytic procedures to assess and understand the internal structure of a new or substantially revised intellectual assessment instrument (Floyd & Widaman, 1995; Gorsuch, 1983). For instance, the Wechsler Intelligence Scale for Children—Fourth Edition (WISC-IV) was substantially revised and the authors of the WISC-IV used both exploratory and confirmatory procedures to investigate the theoretical structure (Wechsler, 2003).

Considering that the fifth edition of the SB intelligence test represented a significant revision of the instrument, Dombrowski et al. (2004) argued that additional analyses, including testing alternative models and use of exploratory factor analytic techniques, would be helpful to support the models presented in the technical manual. The case for additional analyses is supported by confirmatory fit analytic index information presented in the SB5 manual, which showed adequate model fit (Dombrowski et al., 2004). Thus, based on in-depth revisions of the instrument and a need for additional analyses, it is of interest to reexamine the constructs and factor models measured by the SB5. This study investigated the factor structure of the SB5 using both exploratory and confirmatory factor analytic procedures across five age groups used in the SB5 standardization process. Additionally, an alternative factor structure was proposed and tested. Finally, results are discussed along with implications for clinical interpretations of the SB5.

Method

Description of Instrument

The SB5 is an individually administered measure of intellectual ability designed to assess individuals ranging in age from 2 to 85-plus years. The instrument has been revised to align

with several factors reported in the Cattell-Horn-Carroll (CHC) model of cognitive abilities (Carroll, 1993; Cattell & Horn, 1978). The SB5 contains 10 subscales, 5 of which are considered Verbal and 5 of which are considered Nonverbal (Roid, 2003). Each SB5 subscale is composed of a "testlet," which is a brief minitest at each level of difficulty. Mean and standard deviation information has been rescaled to standard scores of 100 and 15, respectively (Roid, 2003).

The test manual reports that the SB5 subscale information provides for measurement of three areas: (a) general cognitive functioning, or *g*; (b) verbal and nonverbal intelligence; and (c) five CHC factors stratified along a collective verbal/nonverbal dimension. The five SB5 subtests, including the two routing subtests, are designed to measure the following 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*). Additionally, the five scales that assess verbal intelligence measure the verbal dimension; subsequent scales assess the nonverbal domain. Finally, a measure of overall cognitive ability may be reported. The Full Scale IQ (FSIQ) score is derived from the administration of 10 subtests, 5 Verbal and 5 Nonverbal.

Test administration is expedited through adaptive testing via routing subtests and basal/ceiling rules. 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. The inclusion of routing subtests continues the tradition of the SB4 adaptive testing approach; however, this has been expanded to include a routing subtest for both the verbal and nonverbal subtests.

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 (Roid, 2003). Individual subtest reliabilities were slightly lower, ranging from .84 (Verbal Working Memory) to .89 (Verbal Knowledge). Both test-retest and split half (i.e., internal consistency) methods of ascertaining reliability are provided in the technical manual. Reliability figures are appropriately high (i.e., test-retest average median reliability estimate = .86; split half average median reliability estimate = .90), illustrating consistency across time and within form (Crocker & Algina, 1986).

The relationships between the SB5 and other measures of cognitive ability and achievement are reported in the technical manual (Roid, 2003). Concurrent validity evidence is strong with a reported correlation of .90 between the SB5 FSIQ and the SB4 Composite. Convergent validity evidence between the SB5 and other established measures (e.g., academic achievement) is also provided in the technical manual.

Participants and Data

A total of 4,800 individuals were included in the SB5 standardization sample. Using variables identified in U.S. Census Bureau (1998) publications, norm group individuals were stratified according race/ethnicity, sex, parental education level, and geographic region. The stratification variables closely matched census percentages. However, the technical manual reports that children receiving special education services for greater than 50% of their day were excluded from the normative sample. In addition, the technical manual reports that persons were also excluded for conditions such as severe medical conditions, severe emotional/

behavioral disorders (e.g., medically diagnosed autism), limited English proficiency, and sensory or communication deficits (e.g., hearing, speech, vision, orthopedic impairments, or traumatic brain injury). In addition to the 4,800 cases collected for the normative sample, test results on 1,365 children with exceptionalities were collected (e.g., ADHD, autism, orthopedic impairment, serious emotional disturbance, etc.). The types and number of validity cases for special groups is reported in the technical manual (Roid, 2003, p. 59). Overall, the standardization sample appears sound.

The SB5 may be used to assess intelligence across the age spectrum. The test is appropriate for children beginning at the age of 2, and it may also be used with elderly subjects. It is recognized that there may be some difficulty with measuring certain age groups (e.g., very young children). Therefore, test publishers collected and reported information relative to five age groups: 2 to 5 years, 6 to 10 years, 11 to 16 years, 17 to 50 years, and 51 and older. During the norming process, the five groups were examined independently. Reliability and validity evidence is provided in the technical manual for each age group.

Exploratory Factor Analysis (EFA): Statistical Methodology

First, the structure of the SB5 was investigated using EFA techniques. Factor analysis generally refers to a wide array of statistical techniques used to examine relationships between items and latent factors (Comrey & Lee, 1992; Crocker & Algina, 1986; Gorsuch, 1983; Loehlin, 1998). The overriding purpose of EFA is to account for the relationships between observed variables by summarizing the data set into a smaller number of factors or dimensions (Comrey & Lee, 1992; Crocker & Algina, 1986; Gorsuch, 1983; Loehlin, 1998). A truly exploratory factor analysis does not have a hypothesis about the number of common factors required to account for the underlying dimensions of a data set (Crocker & Algina, 1986). However, most EFA research is not truly exploratory because researchers often have an idea of the number of dimensions underlying a scale. EFA methodology is often criticized for its lack of external validation, a researcher's subjectivity in determining a final solution, and the lack of statistical criteria to use when evaluating EFA solutions (Crocker & Algina, 1986; Gorsuch, 1983; Loehlin, 1998).

Traditionally, EFA investigations are conducted prior to confirmatory factor analyses (CFAs) using an independent sample from the same population (Gorsuch, 1983; Schumacker & Lomax, 1996). The SB5 manual noted that EFA was not conducted because (a) the instrument was a revised version of an earlier instrument and (b) the factor structure was known a priori. However, it is reasonable to assume that the substantial revisions made for the fifth edition, both in terms of revised item content and different theoretical organization, may present unique conditions that are not represented in models constructed solely from prior research. Therefore, EFA investigations prior to CFA could hold important information. Gorsuch (1983) suggested that researchers can be confident about the obtained results when they are replicated using various factor extraction methods but that researchers should be wary of results if different analyses (e.g., EFA vs. CFA) produce different results.

EFA Procedures

The technical manual indicates that the SB5 measures five dimensions of intelligence using five subscales, measured along both verbal and nonverbal domains. The set of 10 SB5 scales were used in the analyses. However, the test publishers split the information included

in the 10 subscales, resulting in 20 half scales. Correlations between the 20 half scales were used in the EFA analyses. Analyses were conducted separately for the five different age groups (e.g., 2-5 years, 6-10 years, 11-16 years, 17-50 years, and 51 and older). Exploratory factor analyses were run within each age group using SPSS for Windows (Version 12.0; SPSS, 2003). The EFA information was evaluated for each age group separately and then across age groups to determine if there was empirical evidence to support additional factor solutions to be tested with CFA methods. For all exploratory analyses, a principal axis factor analysis with oblimin rotation was employed.

A critical decision in EFA is to determine the appropriate number of factors to retain (Fabrigar, Wegener, MacCallum, & Strahan, 1999); therefore, a variety of procedures were used. First, a scree plot was printed to gain a sense of the number of factors needed to summarize the SB5 data. Additionally, the default procedure, extraction based on eigenvalues greater than one, was used as a starting point for the analyses.

We note that both the scree plot and eigenvalue greater than one procedures have been criticized for inaccurately identifying the number of factors underlying a data set and subjectivity in interpretation (Gorsuch, 1983). To overcome these shortcomings, two lesser known procedures, parallel analysis and minimum average partial correlation (MAP), were used to determine the correct number of factors. Parallel analysis and MAP methods are statistically based, and both are noted as accurate methods to determine the correct number of factors to retain (O'Connor, 2000; Velicer, Eaton, & Fava, 2000; Zwick & Velicer, 1986).

Parallel analysis (Horn, 1965) compares plots of eigenvalues from the actual data to eigenvalues extracted from a correlation matrix of randomly generated, uncorrelated variables with the same dimensions as the original data set. Factors are retained if the actual eigenvalue is greater than the eigenvalue from the random data set. Velicer's MAP test (Velicer, 1976) requests a number of principal component solutions, ranging from one solution to the number of variables included for analysis. At each step, the principal components are removed from the correlations and a matrix of partial correlations computed. The average partial correlation of the off-diagonal elements is computed, and the minimum average partial correlation across the solutions suggests the number of components to retain (O'Connor, 2000). Both parallel analysis and MAP were conducted using SPSS programs developed by O'Connor (2000).

Each EFA solution was evaluated based upon five criteria. First, percentage of variance explained by the overall set of factors and by each individual factor was assessed. Second, simple structure was considered, where each item should associate strongly with only one factor (Gorsuch, 1983). Items were considered markers of a factor if their loading value was at least .30. Lower item to factor correlations were considered if an item did not associate higher than .30 with another factor. Third, the solution was evaluated for the absence of specific factors, which may indicate that the data set has been "overfactored." Fourth, the residual matrix was examined. Large residual terms imply that there are additional factors still to be extracted. Finally, the factor solution was judged upon its interpretability and match to theory. This criterion is arguably the most important. For a factor solution to be useful, it needs to be substantively important based upon knowledge of intelligence theory.

CFA: Statistical Methodology

As stated earlier, CFA procedures are appropriate to use when a researcher holds prior knowledge of the underlying latent structure (Byrne, 1998). It is also widely known that the a priori models used with CFA studies are driven by a strong theoretical basis (Benson, 1998; Byrne, 1998; Hoyle & Panter, 1993). A series of alternative models were developed to evaluate if one theoretical perspective of the SB test could be supported over another perspective (Benson, 1998; Breckler, 1990; MacCallum & Austin, 2000; Russell, 2002). Although EFA was used prior to CFA, it is recognized that there may be problems with using these two procedures with the same set of data (Kline, 2005), particularly capitalization on chance solutions from one set of data. The results from the EFA were used mainly for guidance for the CFAs. Therefore, EFAs were conducted to give additional support for alternative factor models, and the CFA methods were used to replicate SB5 models presented in the technical manual.

CFA Procedures

For the CFA tests, a series of four alternative models were tested with each of the SB5 age group data sets. Three of the four models tested were constructed based on the results of EFA, theory, and information from the SB5 manual: (a) a unidimensional (*g* or general intelligence) model, where all 20 subscales loaded on one factor; (b) a two-factor model, measuring verbal and nonverbal dimensions, where 10 verbal subscales loaded on the Verbal factor and the remaining 10 nonverbal subscales loaded on the Nonverbal factor; and (c) a five-factor model, where verbal and nonverbal tests that measured the same scale loaded on one dimension. The final CFA model was developed based on EFA results and information from the previous version of the Stanford-Binet test, the SB4. This model consisted of four factors: Knowledge, Abstract Visual Reasoning, Quantitative Reasoning, and Memory. Details of this model are described later in the text.

The LISREL software package (Version 8.54; Jöreskog & Sörbom, 1996), was used for all CFA analyses. For each age group, a covariance matrix of relationships between the 20 SB5 half subscales was used as input. The maximum likelihood estimation technique was used to obtain parameter estimates and fit information. Also, the error covariances between “like” half subscale scores (e.g., Verbal Working Memory Subscale 1, Verbal Working Memory Subscale 2) were allowed to correlate. Error terms may be estimated if the inclusion of the path can be defended theoretically (Byrne, 1998). We felt that it made sense to correlate errors between half scores on the same scale; no other errors were estimated. We note that this setup differs from the procedures conducted by Roid (2003). The SB5 technical manual noted that error covariances were allowed to correlate (Roid, 2003); however, the manual does not provide further discussion.

Selected fit information was used to judge the fit of each individual model as well as to compare across the set of alternative models. These seven fit indices were chosen on the basis of recommendations from Gerbing and Anderson (1993), Hu and Bentler (1999), and Tanaka (1993): (a) chi-square statistic, (b) goodness-of-fit index (GFI), (c) normed fit index (NNFI), (d) root mean square error of approximation (RMSEA), (e) comparative fit index (CFI), (f) standardized root mean residual (SRMR), and (g) expected cross validation index (ECVI). All fit indices are included as part of the LISREL output.

Traditionally, a nonsignificant chi-square value has been used as evidence of good model-data fit, but it is widely known today that the chi-square value is sensitive to model size and

nonnormality (Bollen, 1989). Attention has shifted to using multiple fit indices that cover different aspects of model-data fit (Gerbing & Anderson, 1993; Schumaker & Lomax, 1996; Tanaka, 1993). The GFI is an absolute fit index, and it provides a measure of the amount of variance/covariance in the sample matrix that is predicted by the model implied variance/covariance matrix. Both the NNFI and CFI are incremental fit indices and test the proportionate improvement in fit by comparing the target model to a baseline model with no correlations among observed variables (Bentler & Bonett, 1980). GFI, NNFI, and CFI values approximating .95 were indicative of good fit (Hu & Bentler, 1999). The SRMR is the average of the standardized residuals between the specified and obtained variance-covariance matrices (Bollen, 1989; Jöreskog & Sörbom, 1996). The SRMR value should approximate or be less than .08 (Hu & Bentler, 1999). The RMSEA represents closeness of fit (Browne & Cudeck, 1993). The RMSEA value should approximate or be less than .05 to demonstrate close fit of the model (Browne & Cudeck, 1993). The 90% confidence interval (CI) around the RMSEA point estimate should contain .05 to indicate the possibility of close fit (Browne & Cudeck, 1993). The ECVI is a single sample estimate of how well the current solution would fit in an independently drawn sample, and it can be employed to compare the fit of competing models (Browne & Cudeck, 1993).

Results

EFA Results

For each age group, one- through five-factor solutions were run and interpreted according to the aforementioned criteria. The results were evaluated to determine the optimal EFA solution within age groups and as well as across age groups. Default procedures suggested a five-factor solution underlying each data set. However, these solutions did not recover the model described in the SB5 manual where like subscales load together on distinct factors. For all five age groups, results from the MAP tests stated that one component was sufficient to describe the SB5 data. Parallel analysis tests stated that two components were present for younger age groups (i.e., 2-5 years and 6-10 years) and one component was present for older age groups.

Within each age group, exploratory analyses showed strong support for a unidimensional model that measured general intelligence (*g*). Across the data sets, the single-factor model reported the highest factor loadings (average low loading = .57, average high loading = .77), and, on average, the one-factor model accounted for 46% of the variance. Further support for the unidimensional models was given in the MAP and parallel analysis analyses. Table 1 presents the one-factor models across all age groups.

Of the five age groups, only the groups of young children (2-5 years, 6-10 years) reported evidence of a two-factor model measuring verbal and nonverbal domains. For 2- to 5-year-olds, the Verbal factor included 14 half scales, 6 of which were verbal scales measuring Working Memory, Fluid Reasoning, and Visual/Spatial Ability. The remaining 8 half scales on this factor included both Verbal/Nonverbal Knowledge and Verbal/Nonverbal Quantitative Reasoning. The second factor, Nonverbal included six nonverbal scores measuring Working Memory, Fluid Reasoning, and Visual/Spatial Ability. The verbal and nonverbal dimensions were related, with a correlation of .61 between factors.

Table 1
Exploratory Factor Analysis Results for the Stanford-Binet–
Fifth Edition (SB5), Unidimensional Model

Characteristic	Age Group				
	2 to 5 Years	6 to 10 Years	11 to 16 Years	17 to 50 Years	51 and Older
Sample size	1,400	1,000	1,200	514	686
Subscales					
Verbal (V)–Knowledge 1	.68	.71	.72	.70	.69
V–Knowledge 2	.70	.66	.69	.61	.64
Nonverbal (NV)–Knowledge 1	.66	.60	.67	.75	.74
NV–Knowledge 2	.63	.70	.73	.74	.79
V–Fluid Reasoning 1	.72	.69	.63	.63	.64
V–Fluid Reasoning 2	.65	.72	.70	.74	.74
NV–Fluid Reasoning 1	.53	.64	.62	.63	.63
NV–Fluid Reasoning 2	.53	.64	.61	.67	.63
V–Visual/Spatial Processing 1	.70	.73	.75	.80	.70
V–Visual/Spatial Processing 2	.69	.69	.75	.79	.71
NV–Visual/Spatial Processing 1	.55	.66	.62	.66	.70
NV–Visual/Spatial Processing 2	.58	.66	.56	.59	.65
V–Quantitative Reasoning 1	.67	.69	.71	.78	.77
V–Quantitative Reasoning 2	.65	.66	.76	.83	.79
NV–Quantitative Reasoning 1	.69	.67	.74	.80	.74
NV–Quantitative Reasoning 2	.74	.60	.69	.77	.70
V–Working Memory 1	.70	.61	.60	.70	.62
V–Working Memory 2	.70	.65	.61	.68	.61
NV–Working Memory 1	.63	.59	.63	.70	.68
NV–Working Memory 2	.66	.61	.65	.72	.69
Percentage of variance explained	42.5	43.4	45.5	51.3	48.3

Note: Sample size information provided from the SB5 technical manual (Roid, 2003, p. 111). Numbers (1, 2) denote half scale for a particular dimension.

For the 6- to 11-year-olds, results showed a similar pattern. EFA results identified a Verbal factor that included 10 half scales. Six were verbal scales from the Working Memory, Fluid Reasoning, and Visual/Spatial Ability dimensions; the remaining 4 half scales included Verbal Knowledge and Nonverbal Knowledge. The Nonverbal factor also included 10 half scales. It included the nonverbal counterparts for Working Memory, Fluid Reasoning, and Visual/Spatial Ability and Verbal/Nonverbal Quantitative Reasoning. The relationship between the two factors was high, with a correlation of .76. For the data sets representing ages 11 through adulthood, the two-factor verbal/nonverbal model was not supported. Here, the two factor models reported a dominant factor, which consisted of 18 of the 20 half scales.

A five-factor model, as described in the SB5 manual, was similarly unsupported by exploratory analyses across all age groups. These analyses showed evidence of a dominant factor. Depending on the data set, the dominant factor included between 7 and 10 of the half scales. Interpretations of the extracted five factors did not replicate the five dimensions measured by the SB5.

Some commonalities were uncovered across series of EFAs and across age groups. For example, across the age groups, many factors included both Verbal and Nonverbal Knowledge on a distinct factor and also paired Verbal and Nonverbal Quantitative Reasoning on a separate factor. Additionally, there were virtually no models that reported cross-loading values and the majority of models held even numbers of scales, meaning that scales which measured the same dimension in the same way (i.e., measuring IQ in the verbal domain) were placed on the same factor.

From the collection of EFA results and information concerning the SB4 theoretical structure, a four-factor model was tested with the CFA analyses. This model, although not expressly found in any one data set, was built from a combination of the empirical evidence and substantive theory. The model consists of four factors: Knowledge, Abstract Visual Reasoning, Quantitative Reasoning, and Memory. With this model, the Knowledge factor will pair together eight half scales: Verbal and Nonverbal Knowledge in addition to Verbal Fluid Reasoning and Verbal Visual Spatial Reasoning. Abstract Visual Reasoning will consist of four half scales, measures of both Visual/Spatial Reasoning and Fluid Reasoning across the nonverbal domains. The final two factors, Quantitative Reasoning and Memory, will test factors consisting of the verbal and nonverbal representations of the respective scales. This four-factor model also has strong ties to substantive theory from the previous version of the Stanford-Binet test.

CFA Results

Confirmatory analyses were used to test four different models: (a) unidimensional, (b) two-factor (Verbal/Nonverbal), (c) five factor (by dimension), and (d) a four-factor model that was developed using EFA information and prior theory. CFA runs were conducted for each of the five age groups.

The four CFAs models could be estimated for the youngest age group, 2 to 5 years. CFA results are presented in Table 2. As shown in the table, all four models converged to stable estimates. Fit information illustrated good fit for the set of models tested with the 2- to 5-year age group, with all indices exceeding suggested cutoff values. However, for the models with more than one factor, the correlation between factors was extremely high, revealing redundancy between factors. This one factor model measured a general factor of intelligence. Although fit indices were at the highest levels for the five-factor model, little benefit was gained by including additional factors. Considering both the high correlations between factors as well as the notion of parsimony, the unidimensional model is suggested.

CFA results for the remaining four age groups are presented in Table 3; however, only one- and two-factor solutions are presented. The more complex CFA models (four-factor and five-factor) were not able to be evaluated due to estimation problems. The results of the two-factor model (i.e., measuring verbal and nonverbal dimensions) showed a very high correlation between dimensions, again suggesting that more than one factor might not be necessary. Furthermore, it is posited that the high correlation between factors contributed to the estimation problems. Again, a unidimensional model is suggested based upon the notion of parsimony and the high correlations between factors.

In an attempt to understand the underlying structure of the SB5, additional analyses were conducted. These analyses removed the influence of the general factor to allow for examinations of other factors. Using this process, the two-factor and five-factor models were tested;

Table 2
Confirmatory Factor Analysis Fit Information for the Stanford-Binet-
Fifth Edition (SB5), 2- to 5-Year Age Group

	One Factor, Unidimensional	Two Factor, Verbal/Nonverbal	Four Factor, SB4 Parallel	Five Factor, SB5 Factors
Chi-square	584.62	581.36	485.3	456.07
<i>df</i>	160	159	154	150
GFI	.96	.96	.96	.97
NNFI	.99	.99	.99	.99
CFI	.99	.99	.99	.99
RMSEA	.05	.05	.04	.04
(90% confidence interval)	(.04-.05)	(.04-.05)	(.04-.05)	(.04-.04)
ECVI	.52	.52	.44	.42
(90% confidence interval)	(.47-.58)	(.47-.58)	(.40-.50)	(.38-.47)
SRMR	.04	.04	.03	.03
Range of correlation between factors		.98	.84-.98	.81-.97

Note: GFI = goodness-of-fit index; NNFI = normed fit index; CFI = comparative fit index; RMSEA = root mean square error of approximation; ECVI = expected cross validation index; SRMR = standardized root mean residual.

error terms between half scales measuring the same dimension were allowed to correlate. Across the majority of age groups, the analyses encountered estimation problems and could not be interpreted. Based on this series of analyses, results with the SB5 data sets suggest a one-dimensional model measuring general intelligence best represents all age groups.

Discussion

The purpose of this study was to investigate the latent constructs and theoretical structure of the latest edition of the Stanford Binet intelligence test (SB5). The investigation was suggested based upon extensive changes to the instrument's item content, standardization procedures, and underlying theoretical structure. Furthermore, information presented in the SB5 technical manual prompted questions due to the emphasis on confirmatory modeling procedures presented in the manual for validation support (Dombrowski et al., 2004). In this study, a series of alternative models were developed from EFAs, information provided in the technical manual, and prior research in the area of intelligence theory. Investigations were conducted separately for each of the five age groups cited in the SB5 manual, and solutions were compared within age groups as well as across the ages.

We recognize that the EFA results were conducted "post hoc" and were used as independent samples of data. Nevertheless, interesting results were obtained. For example, the EFA information suggested that the SB5 includes both verbal and nonverbal dimensions for children younger than 10 years of age. For older age groups, results suggested a one-factor model, measuring general intelligence, provided the best fit. These results were supported by information from parallel analysis, MAP tests, and CFA.

Table 3
Confirmatory Factor Analysis Results for the Stanford-Binet–Fifth Edition (SB5),
One- and Two-Factor Solutions

	6 to 10 Years	11 to 16 Years	17 to 50 Years	51 and Older
One-factor model				
Chi-square	336.60	226.64	225.44	263.82
<i>df</i>	160	160	160	160
GFI	.97	.98	.96	.96
NNFI	.99	1.00	1.00	1.00
CFI	1.00	1.00	1.00	1.00
RMSEA	.04	.02	.03	.03
(90% confidence interval)	(.03-.04)	(.01-.02)	(.02-.04)	(.03-.04)
ECVI	.45	.27	.63	.54
(90% confidence interval)	(.40-.51)	(.24-.31)	(.56-.72)	(.48-.62)
SRMR	.03	.02	.03	.03
Two-factor model				
Chi-square	334.04		225.27	263.70
<i>df</i>	159		159	159
GFI	.99		.96	.96
NNFI	1.00		1.00	1.00
CFI	1.00		1.00	1.00
RMSEA	.03		.03	.03
(90% confidence interval)	(.03-.04)		(.02-.04)	(.03-.04)
ECVI	.45		.63	.55
(90% confidence interval)	(.40-.51)		(.57-.72)	(.49-.62)
SRMR	.03		.03	.03
Correlation between factors	.98		1.00	1.00

Note: No estimates were obtained for 11- to 16-year-olds in the two-factor model due to estimation problems. GFI = goodness-of-fit index; NNFI = normed fit index; CFI = comparative fit index; RMSEA = root mean square error of approximation; ECVI = expected cross validation index; SRMR = standardized root mean residual.

For the majority of models tested, confirmatory analyses did not yield acceptable results. With the exception of the youngest age group (2 to 5 years), models testing more than two latent variables encountered estimation problems. Part of the reason for the difficulty was a very high correlation between factors, suggesting that latent variables conceptualized as distinct actually contained a high degree of overlap. Although it is recognized that there will be a relationship between different aspects of intelligence, correlations between factors were very high, ranging from .89 to .98. Relationships of such high magnitudes were not expected. These findings may have implications for models where many distinct variables are of issue, such as the SB5 five-factor model. Scored information that reports on different aspects of intelligence may not report unique information for each of the five domains. Additionally, problems with the SB factor structure have been reported previously (Thorndike, 1990).

Due to discrepancy between the obtained results and information presented in the SB5 technical manual, the test publishers were contacted. We note that the publishers did provide a sample program (two-factor model testing Crystallized and Fluid factors; Roid, 2003, p. 111). However, even with the sample program provided, our analyses revealed that the relationship between factors was reported to be very high (average of .86 across the five age

groups). It is further recognized that the difficulty with the analyses in this study may be due in part to the models selected for testing as well as the constraints imposed upon the models (e.g., error correlations among like subscales). Therefore, additional independent replications of the SB5 theoretical structure are suggested as an avenue for future research. Additionally, it may be of interest to investigate the structure of the SB5 in different populations, especially in situations where the nonverbal dimension may play a prominent role, such as subjects with limited English proficiency or language disorders.

Although the information uncovered through the analyses was not expected, the information provides useful insights into intelligence theory. For example, EFA and CFA results suggested that the SB5 is a solid test of general intelligence (i.e., g) across all age groups (2 to 5 years through 51 and up). Practitioners and researchers should feel comfortable using the SB5 to measure a general intelligence factor (i.e., FSIQ score). However, based on the evidence from this article, information for models beyond one factor may not be as dependable, due to the high overlap between constructs described in the SB5 technical manual as distinct.

With the revisions to the SB5, the publishers also included information to assess all five CHC factors across both verbal and nonverbal dimensions. With the exception of populations younger than the age of 10, our results provided evidence of a single-factor model. The evidence also provides support for a two-factor verbal-nonverbal dichotomy for the preschool (i.e., 2- to 5-year age group) and early childhood (i.e., 6- to 10-year age group) populations. Given that the test materials are child appealing and creatively designed, there is support for use of the verbal-nonverbal dichotomy information with populations that have not fully developed their verbal ability skills. Specifically, the SB5 may be an optimal measure to use with preschool and early school age populations.

Conclusion

The information presented in the SB5 may be particularly useful for measures of general intelligence and as an option for preschool and early childhood assessment. The inclusion of verbal and nonverbal information seems to hold well for young children. For other age groups, the SB5 reports a strong measure of general intelligence. This is supported by the exploratory and confirmatory factor analyses presented here as well as information from the technical manual. Additional independent replications and/or tests of alternative theoretical structures may be useful to conduct to provide greater information about the SB5 instrument and the complex nature of intelligence.

References

- Benson, J. (1998). Developing a strong program of construct validation: A test anxiety example. *Educational Measurement: Issues and Practice*, 17, 10-22.
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-600.
- Bollen, K. A. (1989). *Structural models with latent variables*. New York: Wiley.
- Breckler, S. J. (1990). Applications of covariance structure modeling in psychology: Cause for concern? *Psychological Bulletin*, 107, 260-273.
- 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.

- Byrne, B. A. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS*. Mahwah, NJ: Lawrence Erlbaum.
- 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.
- Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart & Winston.
- Dombrowski, S. C., DiStefano, C., & Noonan, K. (2004). Review of the Stanford-Binet Intelligence Scales: Fifth Edition (SB5). *Communique, 33*(1), 32-34.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*, 272-299.
- Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7*, 286-299.
- Gerbing, D. W., & Anderson, J. C. (1993). Monte Carlo evaluations of goodness-of-fit indices for structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 40-65). Newbury Park, CA: Sage.
- Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179-185.
- Hoyle, R. H., & Panter, A. T. (1993). Writing about structural equation models. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications* (pp. 158-176). Newbury Park, CA: Sage.
- 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., & Sörbom, D. (1996). *LISREL 8: User's reference guide*. Chicago: Scientific Software International.
- Kamphaus, R. W. (2001). *Clinical assessment of child & adolescent intelligence* (2nd ed.). Boston: Allyn & Bacon.
- Keith, T. Z., Cool, V. A., Novak, C. G., White, L. J., & Pottebaum, S. M. (1988). Confirmatory factor analysis of the Stanford-Binet Fourth Edition: Testing the theory-test match. *Journal of School Psychology, 26*, 253-274.
- Kline, R. B. (1989). Is the fourth edition Stanford-Binet a four-factor test? Confirmatory analysis of alternative models for ages 2 through 23. *Journal of Psychoeducational Assessment, 7*, 4-13.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling*. New York: Guilford.
- Loehlin, J. C. (1998). *Latent variable models* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.
- MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology, 51*, 201-226.
- O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers, 32*, 396-402.
- Reynolds, C. R., Kamphaus, R. W., & Rosenthal, B. (1988). Factor analysis of the Stanford-Binet Fourth Edition for ages 2 through 23. *Measurement and Evaluation in Counseling and Development, 21*, 52-63.
- Roid, G. (2003). *Stanford-Binet Intelligence Scale: Fifth Edition*. Chicago: Riverside.
- Russell, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis in Personality and Social Psychology Bulletin. *Personality and Social Psychology Bulletin, 28*, 1629-1646.
- Schumacker, R. E., & Lomax, R. G. (1996). *A beginner's guide to structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum.
- SPSS. (2003). *SPSS: Version 12.0 for Windows*. Chicago: Author.
- Tanaka, J. S. (1993). Multifaceted conceptions of fit in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 10-39). Newbury Park, CA: Sage.
- Thorndike, R. M. (1990). Would the real factors of the Stanford-Binet Fourth Edition please come forward? *Journal of Psychoeducational Assessment, 8*, 412-435.
- U.S. Census Bureau (1998). *Household and family characteristics: Population update*. Retrieved October 2, 2005, from: <http://www.census.gov/prod/3/98pubs/p20-515u.pdf>.
- Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 31*, 321-327.

- Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment: A festschrift to Douglas Jackson at seventy* (pp. 41-71). Norwell, MA: Kluwer Academic.
- Wechsler, D. (2003). *Wechsler Intelligence Scales for Children—Fourth Edition*. San Antonio, TX: Psychological Corporation.
- Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 117, 253-269.