Stanford-Binet IV Intelligence Scale: Is its structure supported by LISREL congeneric factor analyses?

Gregory J. Boyle, Bond University

Available at: https://works.bepress.com/defa_gregory/136/
Stanford-Binet Intelligence Scale: Is its Structure Supported by Lisrel Congeneric Factor Analyses?

Gregory J. Boyle
University of Queensland

Address correspondence to Gregory J. Boyle, PhD, Department of Psychology, University of Queensland, St Lucia, Queensland 4067, Australia
Abstract

Boyle (Personality and Individual Differences, 10, 709-715, 1989b) conducted an iterative principal factoring and oblique rotation of the standardisation sample (n = 5,013) for the revised Stanford-Binet Intelligence Scale (Fourth Edition). Nevertheless, because of the exploratory methodology employed, the results were problematic. Keith, Cool, Novak, White and Pottebaum (1988)-(Journal of School Psychology, 26, 253-274) had previously carried out a confirmatory factor analysis and had concluded that the results supported the four Area dimensions (Verbal Reasoning, Quantitative Reasoning, Abstract/Visual Reasoning, and Short-Term Memory) in the new instrument. As congeneric factor analysis via LISREL has not yet been performed, the present paper presents a LISREL reanalysis of the subtest intercorrelations provided in the Stanford-Binet IV Technical Manual. In replicating the confirmatory analysis of Keith et al., it was found that some of the subtests were loaded differently from that previously reported, although each Area dimension was supported strongly by congeneric factor analysis. Even so, the magnitude of the intercorrelations between the four Area dimensions was quite high, suggesting that intellectual abilities are influenced strongly by general ability (g).
Introduction

Boyle (1989b) reported the results of an exploratory factor analysis of the new Stanford-Binet IV Intelligence Scale (SB4) produced by Thorndike, Hagen and Sattler (1986). An iterative principal factoring procedure along the lines suggested by Cattell (1978), Gorsuch (1983), Kline (1987), and Boyle (1988) with number of factors estimated by both the Kaiser-Guttman criterion (cf. Yeomans & Golder, 1981) and psychometric Scree test (Cattell & Vogelmann, 1977; Hakstian, Rogers & Cattell, 1982), and with rotation to direct Oblimin oblique simple structure, was employed. Boyle found that the four major Area dimensions (Verbal Reasoning, Quantitative Reasoning, Abstract/Visual Reasoning, and Short-Term Memory) were all supported adequately. However, the resulting factors may have been influenced unduly by inclusion of the Area dimensions, raising possible problems associated with multicollinearity (cf. Pedhazur, 1982).

Previously, Thorndike et al. (1986) had carried out a 'modified' confirmatory factor analysis of the SB4. However, their findings were inconclusive, and several of the factor loadings on the respective subscales were trivial (generally only factor loadings of at least 0.30 and above are regarded as conceptually meaningful). Subscales failing to meet this minimum requirement (and therefore accounting for <9% of the variance associated with a particular factor) included the Absurdities subtest on the Verbal Reasoning dimension, Bead Memory on the Short-Term Memory factor, Quantitative and Number Series subtests on the Quantitative Reasoning dimension, and three of the subtests on the Abstract/Visual factor (Copying, Matrices, Paper Folding and Cutting).

A principal components analysis, followed by an iterative principal factoring plus Varimax rotation of the subtest intercorrelations had been carried out by Reynolds, Kamphaus and Rosenthal (1987) for each age level. They concluded (p. 15) that, "The factor pattern underlying the Binet is quite fragmented across age and
difficult to make sense of either psychologically or purely empirically". They further concluded (p. 17) that, "interpretation of the new area scores is fraught with problems.... These findings beckon researchers to investigate alternative organizations of the Binet scales that are more congruent with its factor structure".

However, Reynolds et al. (1987) assumed that their particular factor results were reliable and valid. Yet, they employed procedures which are problematic. For instance, they utilised an orthogonal rotation merely on the grounds that previous investigators had used this method in analyses of WISC-R and K-ABC factor structures. Reynolds et al. provided no scientific justification for using orthogonal rotation. They did not investigate the actual degree of obliquity of the derived factors by systematically varying, in say SPSSX, the delta shift parameter, and by examining the resultant hyperplane count to assess the degree of approximation of each factor pattern solution to simple structure. Only in the presence of trivial, non-significant correlations between the obtained factors could use of orthogonal rotation be justified fully (cf. Boyle, 1985, 1989a, for discussion of this issue). Furthermore, they extracted factors on the basis of the somewhat outmoded Kaiser-Guttman criterion with eigenvalues 1.0. Their use of what is essentially the 'little jiffy' method of factor analysis-a method that is fraught with difficulties- undoubtedly accounted for much of their conclusion regarding the inconsistency of Stanford-Binet factor structures across age.

With regard to confirmatory factor analysis of the SB4, Hoffman, Carleton, Bishop-Marbury and Goodwin (1988), as well as Keith et al. (1988) have provided support for the factor structure of the instrument. However, in reviewing this factor analytic research, Glutting (1989) concluded that: "The common element in all this is that no single study has fully substantiated the existence of four Area factors.... Particularly disconcerting is the tendency by the authors of the SB4 to disagree on the number of interpretable factors".
It was in this context that Boyle (1989b) reanalysed the matrix of intercorrelations reported in the SB4 Technical Manual (p. 53), on a combined sample of 5013 Ss (Keith et al., apparently undertook their analysis on 3,354 Ss, although they also claimed to use the entire standardisation sample). In the present study, congeneric factor analyses via LISREL were conducted for each of the four Area dimensions. As well, in view of the criticisms of Glutting, and in order to verify the published results of Keith et al. (1988), a replication of their confirmatory factor analysis was undertaken in an attempt to settle once and for all the controversy surrounding the factorial dimensionality of the SB4 instrument.

Method and Results

Congeneric factor analyses

In order to avoid the statistical problems associated with multicollinearity, only the intercorrelations for the 15 SB4 subtests served as the starting point. In the first instance, the 15 x 15 intercorrelation matrix reported in the Technical Manual (p. 53) was subjected to congeneric factor analyses, whereby the loadings of each of the four latent traits (SB4 Area dimensions) on the respective subtests were examined for their goodness of fit, via the SIMPLIS (Jöreskog & Sörbom, 1987) and LISREL-VI programs (Jöreskog & Sörbom, 1986). The results of the congeneric factor analyses are presented in Table 1.

For the Verbal Reasoning Area, the Goodness of Fit Index (GFI), and Adjusted Goodness of Fit Index (AGFI)-adjusted for the number of degrees of freedom-were both quite high. The two-stage least squares solution provided initial estimates of the GFI, AGFI and Root Mean Square Residual (RMR)-an index of the degree to which the initial correlation matrix is not reproduced by the estimated factor model (Cuttance & Ecob, 1987). The GFI, AGFI and RMR estimates for the corresponding maximum likelihood procedure were 0.977, 0.887, and 0.030.
respectively. The Total Coefficient of Determination was 0.898, indicating that much of the communality in the Verbal Reasoning Area had been accounted for adequately. As expected, the Likelihood Ratio Test (LRT) statistic ($\chi^2$) declined in magnitude from 290.33 to 246.80 (2 df) in going from the two-stage least squares to the maximum likelihood solution, indicating a qualitative improvement in fit of the congeneric factor model. According to Cuttance (1987, p. 260), models with an AGFI of less than 0.8 are inadequate ... most acceptable models would appear to have an AGFI index of greater than 0.9". On these grounds, the Verbal Reasoning Area would appear to be substantially founded, although the fit of the congeneric model was clearly less than optimal.

Perusal of the standardised regression coefficients, as well as the corresponding error variances and multiple R$^2$ estimates, indicated that the Absurdities subtest was contributing most of the 'noise' to the measurement of the Verbal Reasoning dimension. To improve the model fit, it was considered that this subtest might need to be deleted from both the congeneric factor model and the subsequent confirmatory factor analysis of the SB4.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Congeneric factor models of SB4 area dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subscales (x variables)</td>
<td>Standardized LISREL estimates (ML)-(X)</td>
</tr>
<tr>
<td></td>
<td>Parameter</td>
</tr>
<tr>
<td>Verbal</td>
<td>0.91</td>
</tr>
<tr>
<td>reasoning</td>
<td>0.81</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>0.67</td>
</tr>
<tr>
<td>Comprehension</td>
<td>0.16</td>
</tr>
<tr>
<td>Absurdities</td>
<td>0.84</td>
</tr>
<tr>
<td>Absurdities</td>
<td>0.45</td>
</tr>
<tr>
<td>Coefficient of determination for x variables is 0.898</td>
<td></td>
</tr>
<tr>
<td>Goodness of fit statistics: GFI = 0.977; AGFI = 0.887; RMR = 0.030</td>
<td></td>
</tr>
</tbody>
</table>

Abstract/visual reasoning

| Subscales (x variables) | Standardized LISREL estimates (ML)-(X) |
| | Parameter | Standard error | Significance |
| Pattern analysis | 0.79 | 0.01 | <0.001 | 0.63 |
| Copying | 0.61 | 0.01 | <0.001 | 0.37 |
| Matrices | 0.70 | 0.01 | <0.001 | 0.49 |
| Paper folding and cutting | 0.76 | 0.01 | <0.001 | 0.58 |
| Coefficient of determination for x variables is 0.822 |
| Goodness of fit statistics: GFI = 0.998; AGFI = 0.989; RMR = 0.010 |

Quantitative reasoning

| Subscales (x variables) | Standardized LISREL estimates (ML)-(X) |
| | Parameter | Standard error | Significance |
| Quantitative | 0.80 | 0.01 | <0.001 | 0.64 |
| Number series | 0.84 | 0.61 | <0.001 | 0.70 |
| Equation building | 0.76 | 0.01 | <0.001 | 0.58 |

Coefficient of determination for x variables is 0.822
Indeed, removal of the absurdities subtest from the congeneric model resulted in a perfect fit of the model (x² = 0.000; GFI = 1.000; RMR = 0.000). These findings suggest strongly that the SB4 Verbal Reasoning Area is best measured by the Vocabulary, Comprehension, and Verbal Relations subtests alone. This finding is at odds with the scoring procedures currently advocated by the SB4 test authors.

With regard to the Abstract/Visual (Fluid) Area, inclusion of the Pattern Analysis, Copying, Matrices, and the Paper Folding and Cutting subtests in the congeneric factor model resulted in a maximum likelihood GFI of 0.998, an AGFI of 0.989, and a RMR of 0.010, indicating good fit to the model. The Total Coefficient of Determination was 0.822. Perusal of the standardised regression equations indicated that the Copying subtest contributed the most 'noise'. Removal of the Copying subtest resulted in a perfect fit, such that the maximum likelihood GFI was 1.000. The Total Coefficient of Determination was 0.801. Clearly, the Abstract/Visual Area seems adequately demonstrated, although the Copying subtest scores contributed most of the error variance and might best be excluded.

For the Quantitative Reasoning Area, inclusion of the Quantitative, Number Series, and Equation Building subtests in the congeneric factor model resulted in a perfect fit (GFI = 1.000). The Total Coefficient of Determination was 0.847. This
supports the structure of the Quantitative Reasoning Area as proposed by Thorndike et al. (1986).

Finally, the congeneric factor analysis for the Short Term Memory Area based on the Bead Memory, Memory for Sentences, Memory for Digits, and Memory for Objects subtests, also was encouraging. The maximum likelihood GFI was 0.993, the AGFI was 0.963, while the RMR was 0.021, indicating good fit to the congeneric model. The Total Coefficient of Determination was 0.781. Most of the error variance was associated with the Bead Memory and the Memory for Objects subtests, as indicated by the standardised prediction equations. Removal of the Memory for Objects subtest resulted in a perfect fit of the model with a maximum likelihood GFI of 1.000, thereby demonstrating that the STM Area score is probably best based on the Bead Memory, Memory for Sentences, and Memory for Digits subtests alone.

Overall, the results of the congeneric factor analyses for each of the SB4 Area dimensions support the four separate factors, but suggest that the Absurdities, Copying, and Memory for Objects subtests might be deleted from the Verbal Reasoning, Abstract/Visual Reasoning, and Short-Term Memory Areas, respectively, in the subsequent confirmatory factor analysis across all four Area dimensions.

**Confirmatory factor analysis**

The intercorrelation matrix (SB4 Technical Manual) across all 5,013 Ss (median product-moment correlations with some estimated as described in Boyle, 1989b), was subjected to confirmatory factor analysis using SIMPLIS and LISREL VI (cf. Brown, 1986; Mulaik, 1988; Stankov, 1987). Use of the correlation
coefficients combined across the various age samples undoubtedly introduced additional 'noise' into the various analyses, thereby resulting in less than optimal results from the present confirmatory factor analysis.

Initial estimation (based on inclusion of all 15 subtest indicators in the confirmatory factor model) was via the two-stage least squares procedure, the output of which was subjected to maximum likelihood estimation, to obtain more accurate goodness of fit values. Results of the maximum likelihood procedure (see Table 2) were adequate (GFI = 0.910; AGFI = 0.871; RMR = 0.046), although as Cuttance (1987) has indicated, the AGFI for acceptable models preferably should be >0.9. The Total Coefficient of Determination was 0.985, for the SB4 structural model as proposed by Thorndike et al. (1986). The covariances between the exogenous latent traits/Area dimensions (Φ matrix) are presented in Table 3, while the intercorrelations are provided in Table 4. As is evident, both the covariances and correlations are quite high, suggesting significant measurement overlap between the four SB4 Area dimensions.

As was found with the congeneric factor results, examination of the standardised prediction equations for the confirmatory factor model suggested that the Absurdities, Copying, and Memory for Objects subtests contributed considerable 'noise' and might best be removed from the analysis, to obtain a more accurate indication of the adequacy of the four major latent traits (SB4 Area dimensions). When these subtests were removed, the GFI, AGFI and RMR indices exhibited values of 0.937, 0.897 and 0.040 respectively, clearly representing an improved fit to the confirmatory
Table 2

Confirmatory factor analysis of SB4 subscales

<table>
<thead>
<tr>
<th>SB4 subscales</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>DL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal reasoning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vocabulary</td>
<td>0.90</td>
<td></td>
<td>0.80</td>
<td></td>
</tr>
<tr>
<td>Comprehension</td>
<td>0.82</td>
<td></td>
<td>0.68</td>
<td></td>
</tr>
<tr>
<td>Absurdities</td>
<td>0.69</td>
<td></td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td>Verbal relations</td>
<td>0.76</td>
<td></td>
<td>0.58</td>
<td></td>
</tr>
<tr>
<td>Abstract/visual reasoning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pattern analysis</td>
<td></td>
<td>0.73</td>
<td></td>
<td>0.53</td>
</tr>
<tr>
<td>Copying</td>
<td></td>
<td>0.59</td>
<td></td>
<td>0.35</td>
</tr>
<tr>
<td>Matrices</td>
<td></td>
<td>0.77</td>
<td></td>
<td>0.60</td>
</tr>
<tr>
<td>Paper folding and cutting</td>
<td></td>
<td>0.75</td>
<td></td>
<td>0.57</td>
</tr>
<tr>
<td>Quantitative reasoning</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quantitative</td>
<td></td>
<td>0.82</td>
<td></td>
<td>0.67</td>
</tr>
<tr>
<td>Number series</td>
<td></td>
<td>0.85</td>
<td></td>
<td>0.73</td>
</tr>
<tr>
<td>Equation building</td>
<td></td>
<td>0.72</td>
<td></td>
<td>0.52</td>
</tr>
<tr>
<td>Short-term memory</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bead memory</td>
<td></td>
<td></td>
<td>0.70</td>
<td>0.49</td>
</tr>
<tr>
<td>Memory for sentences</td>
<td></td>
<td></td>
<td>0.75</td>
<td>0.56</td>
</tr>
<tr>
<td>Memory for digits</td>
<td></td>
<td></td>
<td>0.67</td>
<td>0.45</td>
</tr>
<tr>
<td>Memory for objects</td>
<td></td>
<td></td>
<td>0.58</td>
<td>0.34</td>
</tr>
</tbody>
</table>

Coefficient of determination = 0.985.
GFI = 0.910; AGFI = 0.871; RMR = 0.046.

Perusal of communality estimates indicates that the Absurdities, Copying, and Memory for objects scales account for the least amount of common factor variance, in accord with the findings from the congeneric factor analysis reported in Table 1.

Table 3

Covariances between exogenous latent traits (Φ matrix)

<table>
<thead>
<tr>
<th></th>
<th>Verbal</th>
<th>Abstract/visual</th>
<th>Quantitative</th>
<th>STM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abstract/visual</td>
<td>0.80</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quantitative</td>
<td>0.78</td>
<td>0.92</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>STM</td>
<td>0.84</td>
<td>0.82</td>
<td>0.84</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Table 4

Estimated correlation matrix

<table>
<thead>
<tr>
<th></th>
<th>Verbal</th>
<th>Abstract/visual</th>
<th>Quantitative</th>
<th>STM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abstract/visual</td>
<td>0.82</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quantitative</td>
<td>0.79</td>
<td>0.89</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>STM</td>
<td>0.82</td>
<td>0.83</td>
<td>0.83</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Area correlations estimated from Two-Stage Least-Squares (TSLS) solution.
factor model. The Total Coefficient of Determination improved slightly to 0.986, as shown in Table 5. Again, the covariances between each of the latent traits (Φ matrix) were rather high (Table 6), as were the intercorrelations (Table 7).

In order to investigate the relationships between the four Area dimensions and superfactor (g), in accord with the work of Eysenck (1979, 1982, 1986a,b,c;
Jensen, 1979, 1987a, b; and Weiss, 1986), the intercorrelations for the four Area dimensions were subjected to a maximum-likelihood factor analysis with extraction of a single factor. For the unmodified SB4, the following prediction equation resulted (where VREAS = Verbal Reasoning; QREAS = Quantitative Reasoning; ABS/VIS = Abstract (Visual Reasoning; and STM = Short-Term Memory, respectively):

$$g = 0.16 \text{VREAS} + 0.31 \text{QREAS} + 0.38 \text{ABS/VIS} + 0.20 \text{STM}$$

For the modified instrument (with deletion of the Absurdities, Copying, and Memory for Objects subtests), the prediction equation which resulted was as follows:

$$g = 0.09 \text{VREAS} + 0.50 \text{QREAS} + 0.31 \text{ABS/VIS} + 0.15 \text{STM}$$

In both instances, the factor loadings (equivalent to the standardised regression coefficients) were highest on the Quantitative and Visual/Abstract Reasoning Areas, and relatively low on the Verbal and Short-Term Memory Areas. This finding suggests the important contribution of fluid abilities (gr), as compared with crystallised abilities (gc) in the composition of the superfactor (g).

Further examination of the relationships between the Area scores and g, using a congeneric model in LISREL, revealed that the fit of the one-factor model was good. For the full SB4 (with all 15 subtests included in the model), the Total Coefficient of Determination was 0.953, the GFI was 0.968, the AGFI was 0.838, while the RMR was only 0.016.

Confirmatory factor analysis was also applied to a three factor-model of the SB4 across all 15 subtests. For the maximum likelihood solution, results revealed a GFI of 0.896, an AGFI of 0.858, and an RMR of 0.048, indicating a
less adequate fit than for the four-factor model proposed by Thorndike et al. (1986).

**Conclusions**

The present congeneric factor analyses of the standardisation sample, and the replicated confirmatory scale factorings suggest that the four SB4 Area dimensions are reasonably well defined, as proposed by Thorndike et al. (1986). These findings confirm those already reported by Keith et al. (1988), and demonstrate the prematurity of criticisms by Glutting (1989), as well as Glutting and Kaplan (1990) regarding the claimed inadequacy of the factor structure of the SB4 instrument. However, it is important to recognise that the four Area dimensions are intercorrelated at a rather high level, thereby suggesting the importance of higher-order organising dimensions of g (cf. Jensen, 1980, 1987a, b; and Horn, 1985, 1986, 1988).

Additionally, three of the subtests in the SB4 appear to be measuring mostly 'noise', rather than the defined Areas. It is therefore suggested in practical work with the SB4, that the Absurdities, Copying, and Memory for Objects subtests be excluded from calculations involving the Verbal Reasoning, Abstract/Visual Reasoning, and Short-Term Memory Areas, respectively. This procedure, not only should provide more accurate assessments of the above Area scores, but simultaneously will reduce the time required to administer the SB4, thereby resulting in more efficient measurement.

**Acknowledgement**

The author is indebted to Dr Ken Rowe, Senior Policy Officer (Research), Victorian Ministry of Education, for his advice regarding the use of the LISREL statistical package.
References


