DOCUMENT RESUME

ED 302 577 TM 012 624

AUTHOR Keith, Timothy Z.; And Others
TITLE Hierarchical Confirmatory Analysis of the Stanford-Binet Fourth Edition: Testing the Theory--Test Match.
PUB DATE 9 Apr 88
NOTE 66p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 5-9, 1988).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Adolescents; Adults; Age Differences; Age Groups; Children; Cognitive Measurement; Factor Analysis; *Intelligence Tests; Memory; Psychological Testing; Test Theory; *Test Validity; *Theory Practice Relationship
IDENTIFIERS Confirmatory Factor Analysis; First Order Cognitive Analysis; Hierarchical Analysis; Reasoning Tests; *Stanford Binet Intelligence Scale

ABSTRACT

The Stanford-Binet Intelligence Scale: Fourth Edition is a conceptually new version of this traditional intelligence scale. The new scale has a solid basis in theory, but there is little evidence that the Binet matches its intended theory. This study was designed to determine whether the Binet corresponds to the theory that guided its construction. First- and higher-order confirmatory factor analysis was conducted using the entire standardization sample (N=3,354) and three age groups from the standardization sample--12-23 years (n=910), 7-11 years (n=960), and 2-6 years (n=1,484)--to determine the extent to which the Binet measures verbal reasoning, quantitative reasoning, abstract/visual reasoning, short-term memory, crystallized versus fluid intelligence, and general intelligence. The results generally support the four first-order factors as reflecting the underlying structure of the new Binet. The biggest deviation from the theoretical structure occurred for ages 2-6 years; for this age group, it was difficult to separate memory from reasoning. The hierarchical analyses support the test authors' contention that there is a strong "g" component underlying the Binet, but did not support the second level of the Binet theory (which combines the first-order factors into crystallized versus fluid intelligence). Five tables present sample data, and six figures diagram theoretical relationships among factors. (Author/SLD)

Hierarchical Confirmatory Analysis of the Stanford-Binet Fourth Edition: Testing the Theory--Test Match

Timothy Z. Keith

Virginia Polytechnic Institute and State University

Valerie A. Cool

Christine G. Novak

Lyle J. White

The University of Iowa

April 9, 1988

Paper presented at the annual meeting of the American Educational Research Association, New Orleans. The first-order confirmatory analyses reported here were originally reported in 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. The hierarchical analyses are new. Correspondence should be sent to Timothy Z. Keith, 206 UCOB, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061; Bitnet ID: TZKEITH at VTVM1.

Running Head: BINET HIERARCHICAL ANALYSIS

Abstract

The Stanford-Binet Intelligence Scale: Fourth Edition is a conceptually new version of this traditional intelligence scale. The new scale has a solid basis in theory, but there is little evidence that the Binet matches its intended theory. This study was designed to determine whether the Binet corresponds to the theory which guided its construction. First- and higher-order confirmatory factor analysis was conducted using the entire standardization sample and three age groups (12-23, 7-11, 2-6) from the standardization sample to determine the extent to which the Binet measures verbal reasoning, quantitative reasoning, abstract/visual reasoning, short-term memory, crystallized versus fluid intelligence, and general intelligence. The results generally support the four first order factors as reflecting the underlying structure of the new Binet. The biggest deviation from theoretical structure occurred for ages 2-6; for this age group, it was difficult to separate memory from reasoning. The hierarchical analyses support the test authors' contention that there is a strong *g* component underlying the Binet, but did not support the second level of the Binet theory (which combines the first order factors into crystallized versus fluid intelligence).

Hierarchical Confirmatory Analysis of the Stanford Binet Fourth Edition: Testing the Theory--Test Match

The newly revised Stanford-Binet Fourth Edition (Binet) is a considerable departure from earlier editions of this traditional intelligence scale. As explained in the Technical Manual (TM; Thorndike, Hagen, & Sattler, 1986), the focus of the revision was to provide a clinically useful profile of individual abilities while maintaining an overall general ability score which deemphasized verbal skills in comparison to previous editions.

Like earlier versions, the new Binet continues to assess intellectual ability in individuals from ages two through adulthood, but the structure of the scale and the theory underlying the scale are quite different from earlier versions. The new Binet continues to measure general intelligence, or _g_. This _g_ may next be divided into three broad cognitive abilities or factors: crystallized abilities, fluid-analytic abilities, and short-term memory, a modification of Cattell and Horn's theory of intelligence (Cattell, 1982; Horn & Cattell, 1966). Crystallized abilities are further subdivided into verbal reasoning and quantitative reasoning, whereas fluid-analytic abilities are represented by abstract/visual reasoning. At a practical level, the 15 subtests comprising the Binet are combined to produce area standard scores for Verbal Reasoning, Abstract/Visual Reasoning, Quantitative Reasoning, and Short-Term Memory. These four area scores are then added together to produce the Composite Score, a measure of overall intellectual ability (the crystallized vs. fluid ability aspect of the Binet is not reflected in the test's scoring). Not all of the 15 tests are used at every age level; rather, up to 13 tests are used depending on subject age, subject ability, and examiner choice.

Few independent investigations of the construct validity of this new edition of the Stanford-Binet have been conducted, so potential users must rely primarily on the evidence provided in the TM. The new scale has a solid basis in theory, and the TM does provide some general support for the Binet as a measure of intelligence, but there is little evidence that the Binet matches its intended theory. For example, one can conclude little from the factor analytic results in the TM. The factor loadings are quite small, but then the vague reference to the procedure used makes it difficult to know exactly what should be expected. Without more detail as to how this "variant of confirmatory factor analysis" (p. 52) was conducted, the findings reported cannot be regarded as unequivocal support for the test's factor structure.

Thus, there have been major changes in the new edition of the Binet. There is a strong theory underlying the new scale, but only scant evidence that the Binet matches its intended theory. The present study was undertaken to determine the extent to which the new Binet corresponds to the theory which guided its construction. To fulfill this purpose, we performed first and higher-order confirmatory factor analysis on the Binet standardization data to determine whether the Binet tests measure the constructs they are designed to measure.

Method

Instrument

Like earlier versions, the new Binet continues to assess intellectual ability in individuals from ages two through adulthood, but the structure of the scale and the theory underlying the scale are quite different from earlier versions. At the most basic level, the new Binet measures general intellectual ability, or *g*, using various combinations of 15 tests depending on subject age, subject ability, and examiner choice. The 15 tests are described briefly below, grouped by the scales of which they are a part.

Verbal Reasoning (Crystallized Abilities)

Vocabulary. Pictures or visually- and orally-presented words are defined.

Comprehension. Initial comprehension items consist of pointing out body parts on a picture, whereas latter items require verbal responses to questions (e.g., "Why do people have stores?").

Absurdities. Absurd aspects of pictures are described and explained.

Verbal Relations. Four words are presented; the examinee must tell how the first three are similar to each other but different from the fourth word.

Quantitative Reasoning (Crystallized Abilities)

Quantitative. Manipulatives, pictures, and verbal descriptions provide problems that require use of numerical reasoning, concepts, and computation.

Number Series. Number Series requires examinees to discern a pattern in a series of numbers and to use that pattern to predict the next two numbers in the series.

Equation Building. Pictorially presented groups of numbers and operation signs (+, x, =, etc.) are arranged into balanced equations.

Abstract/Visual Reasoning (Fluid-Analytic Abilities)

Pattern Analysis. Form board or block-design-type tasks are solved, depending on the examinee's age and cognitive skills.

Copying. Geometric designs are reproduced with blocks or paper and pencil depending on examinees' age and ability levels.

Matrices. Visually presented analogies, similar to Raven's Progressive Matrices, are solved.

Paper Folding and Cutting. Pictures are selected as representations of how folded and cut pieces of paper would look unfolded.

Short-Term Memory

Bead Memory. Pictorially presented bead patterns are reproduced by placing beads on a stick.

Memory for Sentences. Orally presented sentences are repeated from memory.

Memory for Digits. Memory for Digits is a simple digits forward and digits reversed task.

Memory for Objects. Pictured objects are recalled in the sequence of presentation.

Insert Figure 1 about here

The hierarchical model shown in Figure 1 graphically displays the theoretical structure of the new Binet. The purpose of the Binet is first to measure general intelligence, or _g_. This _g_ may next be divided into three broad cognitive abilities or factors: crystallized abilities, fluid-analytic abilities, and short-term memory, a modification of Cattell and Horn's theory of intelligence (Cattell, 1982; Horn & Cattell, 1966). Crystallized abilities are further subdivided into verbal reasoning and quantitative reasoning, whereas fluid-analytic abilities are represented by abstract/visual reasoning.

Sample

The Stanford-Binet Technical Manual reports intercorrelation matrices for the Stanford-Binet tests, areas, and composite for subjects in each age level of the standardization sample and the median intercorrelations for the entire standardization sample (TM, pp. 53, 110-126). Four age groups were used for the present analyses: the total standardization sample, and preschool, elementary-aged, and adolescent to adult subgroups of the total standardization sample.

We first used the median intercorrelation matrix for all Binet tests for all age levels--the entire standardization sample (TM, p. 53).[1] Following this overall analysis, we used three age groups from the standardization data. We calculated median intercorrelations for ages 2-6 (preschool), 7-11 (elementary), and 12-23 (adolescent and adult) from the intercorrelation matrices in the back of the TM (pp. 110-126). This particular grouping of ages was chosen because it is used elsewhere in the TM and because there is more similarity among the tests given within these age levels than there is between these age levels. Thus the present sample consists of the nationally representative standardization sample for ages 2 through 23, as well as nationally representative subsamples of ages 2-6, 7-11, and 12-23.[2]

Procedure

A series of confirmatory factor analyses were performed on the input correlation matrices using the LISREL VI computer program (Jöreskog & Sörbom, 1984). The first set of analyses included first-order confirmatory analyses of the Binet subtests for the entire Binet standardization sample (and for three more homogeneous subgroups from that sample) to determine the extent to which those subtests measure verbal reasoning, quantitative reasoning, abstract/visual reasoning, and short-term memory. For example, for the initial analysis the Vocabulary, Comprehension, Absurdities, and Verbal

Relations tests were specified as loading only on a verbal reasoning factor; the Pattern Analysis, Copying, Matrices, and Paper Folding and Cutting tests were specified to load only on an abstract/visual reasoning factor; the Quantitative, Number Series, and Equation Building tests were allowed to load only on a quantitative reasoning factor; and the Bead Memory, Memory for Sentences, Memory for Digits, and Memory for Objects tests were allowed to load only on a short-term memory factor. All other factor loadings were constrained to zero. The first-order factors were allowed to correlate with each other, because each of these factors is presumed to also reflect _g_. This first-order analysis was seen as the first step in testing the Binet theory. Relaxations in the factor models were allowed following the testing of the strict model, but all relaxations were theoretically consistent with the Binet theory.

The second series of analyses included hierarchical (second- and third-order) analyses, using a technique outlined by Marsh and Hocevar (1985), in order to test the validity of the second and third levels of the Binet model: the grouping of the verbal and quantitative factors into a crystallized factor and the combination of the crystallized factor with the abstract/visual (or fluid) and memory factors into a _g_ factor.

Results and Discussion

First-Order Analyses

Total Standardization Sample

Figure 2 shows the results of the first analysis of the entire Binet standardization sample. The arrows in the figure represent the unconstrained factor loadings, or the paths from the latent factors to the observed tests; the curved lines represent the correlations among the factors. The "goodness-of-fit" statistics listed beneath the model in Figure 2 suggest that even this strict model provides a generally good fit to the standardization data.[3] The adjusted goodness-of-fit index for the first model was .879; the root mean square (rms) residual correlation was .044, indicating that the correlation matrix predicted by this model differed from the original correlation matrix, on the average, by only .044. Again, these results suggest that the model shown in Figure 2 provides a good fit to the total standardization data.

Insert Figure 2 about here

Given an adequate fit for the model, the next step in interpreting the model is to examine the factor loadings and correlations. All factor loadings and intercorrelations were significant ($t > 2.0$), again offering support for the validity of the Stanford-Binet theory.

For the verbal reasoning factor, Vocabulary had the highest loading (.89), whereas Number Series had the highest loading (.85) on quantitative reasoning. Matrices had the highest loading on the abstract/visual reasoning factor (.77), with Memory for Sentences having the highest loading on short-term memory (.75). Still, all tests loaded quite highly on their corresponding factors.

The factor intercorrelations were also very high; the smallest correlation was .79 between the verbal and quantitative factors, and the correlation between the abstract/visual and quantitative factors was .92! As might be expected given the history and the theory underlying the Binet, these high correlations suggest a strong *g* factor underlying all scales and tests. By way of comparison, factor intercorrelations among three factors on another new intelligence test, the Kaufman Assessment Battery for Children, range from .56 to 1.0, with most in the .70's (Strommen, in press). The factor correlation matrix also showed some inconsistency with the Binet theory. Becaus the verbal and quantitative scales both require crystallized ability according to the Binet theory, one would expect these two factors to correlate more highly with each other than with other scales. The correlation between verbal and quantitative was the lowest factor intercorrelation (.79), however, and the quantitative factor correlated most highly (.92) with the abstract/visual factor, which requires fluid abilities according to the Binet theory.

Although the fit statistics suggest that the Binet model provides a good fit to the overall standardization data, it is also worthwhile to examine the fit of this model in more detail. One way of doing this is to compare the original correlation matrix with the correlations predicted by the model; large differences between the predicted and observed correlations indicate a lack of model fit. The model overpredicted, or overfit, the correlation between the Copying and Equation Building tests--the predicted correlation was .39 versus the accual correlation of .24--probably as a result of the large correlation between the quantitative and the abstract factors. On the other hand, the model did not adequately account for the correlations among Memory for Sentences and several of the verbal tests (for example, the predicted correlation between Memory for Sentences and Vocabulary was .56 versus an actual correlation of .64), nor did the model adequately account for correlations between the Abstract/Visual tests and Bead Memory and Absurdities. This "underfitting" suggested several ways in which the factor model could be relaxed to provide a better fit to the data, but in ways consistent with the overall Binet theory. Therefore, in a series of analyses, we allowed the Memory for Sentences test to load on the verbal factor in addition to the memory factor, the Bead Memory test to load on the abstract/visual factor, and the Absurdities test to load on the abstract/visual factor. We believe that such changes were theoretically justified because the Memory for Sentences test obviously requires a verbal component, whereas the Bead Memory and Absurdities tests have strong visual components. The final analysis in this series incorporated all three changes, and is shown in Table 1, which also includes the information from Figure 2 for comparison.

Insert Table 1 about here

Although the chi-square (χ^2) is an inappropriate test of the adequacy of a particular model with a large sample size, the χ^2 may be used to compare two similar, competing models. If one model is a subset of another model, then the difference in the two chi-squares (along with the difference in the degrees of freedom) may be used to test whether one model provides a better fit than another (Bentler & Bonett, 1980). Briefly, if additions (e.g., allowing additional factor loadings as was done here) to the model (which reduce the degrees of freedom) result in a significant decrease in χ^2, then those additions are justified; the χ^2 decrease is probably not due to chance. In contrast, if restrictions are placed on the model (e.g., constraining factor correlations, resulting in an increase in df), and the χ^2 does not significantly increase, then those restrictions are also justified. Of course, all such modifications should make theoretical sense (MacCallum, 1986). As can be seen in Table 1, our original strict model produced a χ^2 of 2,189.22 with 84 degrees of freedom; the relaxed, or modified, version of the model produced a χ^2 of 1,636.95 with 81 degrees of freedom. This decrease in χ^2 (change in χ^2 = 552.27 at $df = 3$) is highly significant ($p < .001$), and suggests that the relaxed model, with its three additional factor loadings, provides a significantly better fit to the standardization data than does the original strict model. The other fit statistics also suggest the good fit of this model; in particular, the correlation matrix predicted by the revised model differed from the actual correlation matrix, on average, by only .037. Again, all factor loadings and factor correlations were significant with this revised model, and most were similar to their values in the original model. A few changes were evident, however. Interestingly, in the revised model, Bead Memory loaded highest on the abstract/visual reasoning factor (.49) rather than on short-term memory (.27). Memory for Sentences loaded almost as highly on the verbal reasoning factor (.40) as on the short-term memory factor (.43). Absurdities continued to load more highly on the verbal factor (.47), with a lower loading on abstract/visual (.27). The relaxation of the model slightly lowered the correlations among several of the factors, but the correlation between the abstract/visual and quantitative factors was still .90.

Thus, first-order confirmatory factor analysis of the entire Binet standardization sample generally supported the theoretical model of the Binet as conceptualized by the test authors. There were, however, some inconsistencies, most notably in the degree of relation among the first-order latent factors. Furthermore, a relaxation of the strict theoretical structure of the Binet produced a significantly better fit to the standardization data.

First-Order Analyses of Three Age Groups

Ages Twelve Through Adult

Table 2 presents the results of the confirmatory factor analysis using the Binet model for the 12 through 23-year-old age group from the standardization sample. The results of this initial analysis are quite similar to those for the entire standardization sample. The fit statistics suggest that this strict model also provides a fairly good fit for the adolescent and adult age group in the standardization sample; the adjusted goodness-of-fit index was .822, and the observed correlations and predicted correlations differed, on average, by only .051. All factor loadings and correlations were significant, and all tests had high loadings on the appropriate factors. As in the overall analysis, Vocabulary, Number Series, Matrices, and Memory for Sentences had the highest loadings on the verbal, quantitative, abstract/visual, and memory factors, respectively.

Insert Table 2 about here

Like in the overall analysis there were very high correlations among the first-order latent factors. Again, the highest correlation was between the quantitative and the abstract/visual factors (.93) and the correlation between the verbal and quantitative factors was somewhat lower (.85).

The detailed fit statistics (residuals and modification indexes) again suggested that the Memory for Sentences test should be allowed to load on the verbal factor and the Bead Memory and Absurdities tests should be allowed to load on the abstract/visual reasoning factor. The results of the analysis of this relaxed model are shown in the second column of Table 2; those results were similar to those found with the composite. The relaxed model provided a significantly better fit to the 12-23 standardization data than did the strict Binet model (change in χ^2 = 172.44, *df*=3, *p*<.001). All factor loadings and correlations were also significant in this relaxed model, although changes in factor loadings are evident. When allowed to load on both the abstract/visual and memory factors, Bead Memory loaded more highly (.52) on the abstract/visual than on the memory (.25) factor. Absurdities loaded almost equally on the verbal (.37) and the abstract/visual (.40) factors, and Memory for Sentences loaded only slightly higher on the verbal (.44) than on the memory (.38) factor. As in the strict Binet model, the correlation between the quantitative and the abstract/visual factors was the highest (.92), with the quantitative--verbal correlation somewhat lower (.83).

<u>Ages Seven Through Eleven</u>

The confirmatory factor analytic results for the 7- to 11-year-old age group from the standardization sample are shown in Table 3. Again, the overall fit statistics support the underlying theory of the test. For these elementary children, the adjusted goodness-of-fit was

.873, and the correlations predicted by the strict Binet factor model differed from the actual correlations, on average, by only .050. Although the overall fit of the strict Binet model to the 7-11 data was good, further inspection of the results suggests some problems with this model. In particular, the correlation between the quantitative and the abstract/visual factors was 1.01, an impossible value. Still, inspection of the standard errors suggests that this value is not significantly different from 1.0, and when the model was reanalyzed with this correlation constrained to 1.0, the fit of the model was not significantly worse than the one shown here, and most factor loadings and factor correlations were unchanged. Thus, it appears that the true correlation between the quantitative and the abstract/visual factors for elementary age children is 1.0, a statistically possible value, but certainly one which would not be predicted based on the Binet theory. In essence, this correlation of 1.0 suggests that the quantitative and abstract/visual factors are identical, a finding inconsistent with the Binet theory which posits that the quantitative and verbal factors measure crystallized intelligence, in contrast with the abstract/visual factor, which presumably measures fluid intelligence.

Insert Table 3 about here

Most of the other factor loadings and correlations were quite similar to those for the oldest age group and for the entire standardization sample. Furthermore, examination of the residuals and modification indexes suggested that Memory for Sentences might reasonably be allowed to load on the verbal factor, and that Bead Memory and Absurdities should be allowed to load on the abstract/visual factor. The results of this more relaxed model are also in Table 3, and as in the previous analyses, this relaxed model provided a significantly better fit to the elementary age standardization data than did the strict Binet model (χ^2 change=196.58, df=3, p<.001). This relaxation in the factor model also reduced the correlation between the quantitative and abstract/visual factors to less than 1.0, although the new value (.97) was still somewhat higher than the value for the verbal--quantitative correlation (.82).

Ages Two Through Six

It was somewhat more difficult to estimate the Binet model for 2- through 6-year-olds from the standardization sample. At this age level, only one test, Quantitative, is used to form the Quantitative Reasoning scale. Thus, in the factor analysis, the quantitative factor would be estimated by only one test. We used the most common method of dealing with such problems: we constrained the quantitative reasoning--Quantitative test factor loading to the square root of the reliability of the quantitative test (median reliability for ages 2-6 = .84 from Table 5.1, p. 40 in the TM), and constrained the unique variance of the Quantitative test to the compliment of the reliability (cf. Newman, 1984). Yet even with this necessary modification, there were difficulties estimating the Binet model. The confirmatory factor analysis of the strict Binet model for 2 through 6-year-olds is shown in Table 4. Although the fit statistics suggest that this model provides a relatively good fit to the 2-6 standardization data, there are also obvious problems with the model. In particular, the correlation between the abstract/visual and the memory factors was 1.02, and the correlation between the memory and the verbal factors was close to 1.0 (the correlation was .97, but the standard error of the correlation was .03). The model was therefore reestimated with these two correlations set to 1.0. The factor correlations and factor loadings for this revised model (also shown in Table 4) were almost identical to those shown in the first column in Table 4, and the χ^2 was not significantly greater (χ^2 change = 3.12, _df_=2,

$p > .05$), thus suggesting that the correlations between the memory factor and the verbal and abstract/visual factors are probably 1.0. The finding of a perfect correlation between the memory factor and the verbal and abstract/visual reasoning factors suggests that short-term memory may be indistinguishable from reasoning for preschool children, at least using the Binet tests.

Insert Table 4 about here

These difficulties with model identification for the 2- to 6-year-old age group for the Binet suggest that the theory guiding the development of the Binet does not adequately explain the structure of the test for preschool children. Rather, it appears that substantial modifications are needed in the theory underlying the Binet for the preschool age group. The results of these analyses--suggesting the identity of the short-term memory with the verbal and abstract/visual factors--implies that the short-term memory tasks cannot be separated adequately from the Binet reasoning tasks for this age group. A factor model (the No-Memory model) incorporating this hypothesis is shown in Figure 3. For this model, we assumed that three factors (verbal reasoning, quantitative reasoning, and abstract/visual reasoning) underly the

Binet for preschoolers rather than four factors. Memory for Sentences, which according to the Binet authors should measure short-term memory, was allowed to load only on the verbal reasoning factor, and Bead Memory was allowed to load on the abstract/visual factor instead of on a memory factor. The results of this analysis are shown in Figure 3.

Insert Figure 3 about here

The model shown in Figure 3 is not a direct subset of the strict Binet model (Table 4), and therefore formal statistical comparisons are inappropriate. Nevertheless, all of the goodness-of-fit statistics shown for the No-Memory model are considerably better than those shown in the strict Binet model in Table 4 (the results for the No-Memory model are also shown in the last column of Table 4 for comparison). In fact, the adjusted goodness-of-fit index (.969) and the root mean square residual (.026) suggest that this No-Memory model provides an excellent fit to the 2- through 6-year-old standardization data. All of the factor loadings were high and significant, although the memory tests loaded at a slightly lower level than the more purely reasoning tasks. The correlations among the three factors (verbal, quantitative, and abstract/visual) were also quite high. Still, like in the previous analyses, the correlation between the verbal and quantitative factors (.63) was somewhat lower than the abstract--quantitative correlation (.80) (in addition to the verbal--abstract correlation [.75]).

Higher-Order Analyses

Entire Standardization Sample

The first-order analyses reported above provide important evidence of the validity of the first level of the theory underlying the new Binet: the grouping of the 15 tests into various scales. The second series of analyses used hierarchical confirmatory analysis to test the validity of the second and third levels of the Binet model: the grouping of the verbal and quanititative factors into a crystallized factor and the combination of the crystallized factor with the abstract/visual (or fluid) and memory factors into a *g* factor. For the first step in that series of analyses, a model which matched that shown in Figure 1 (with one minor exception) was specified, and was estimated using the entire Binet standardization sample.

The only change in model specification from that shown in Figure 1 is that a fluid ability factor was not included between the abstract/visual factor and *g*. Fluid analytic abilities are defined entirely by abstract/visual abilities in the Binet model, and thus the two are identical for all practical purposes. This identity could be recognized in LISREL by specifying a path of 1.00 from fluid to abstract/visual, but it was easier simply to drop one of the two levels; the same results would be obtained either way.

As might be expected from the factor intercorrelations from the first-order analyses, there were problems in estimating this model. Although the fit statistics suggested an adequate fit to the data (e.g., rms=.044), the loading of the crystallized factor on g was 1.05, and the crystallized factor had a negative variance (-.10). Therefore, the model was reanalyzed with the g--crystallized factor loading set to 1.00 and the crystallized variance set to 0. This model had an adjusted goodness of fit of .875 and an rms of .045, thus suggesting that it provides an adequate fit to the standardization data. The higher-order factor loadings therefore suggest the identity of the overall g factor and the crystallized factor.

But if the g factor and the crystallized factor are identical, as suggested by these results, then the model could be simplified by deleting the crystallized factor from the model. This revised model is shown in Figure 4; instead of three levels of latent factors beyond the tests, this model has two levels; a _g_ factor affects the first-order factors verbal reasoning, quantitative reasoning, abstract/visual reasoning, and short-term memory. The fit statistics again suggested an adequate fit to the data. The first-order factor loadings are very similar to those shown in Table 1 because the model shown here essentially interprets the factor correlations. Of more interest are the loadings of the first-order factors on the second- ~der _g_ factor; all were quite large (only one, for verbal reasoning, was below .9!), thus offering support for the presence of a stong general factor underlying the new Binet.

Insert Figure 4 about here.

To provide an additional check on the presence of a third-order crystallized factor, we estimated the model shown in Figure 4 allowing a correlation between the errors of measurement (the unique variance) of the verbal and quantitative reasoning factors. This change essentially specifies that there is an unknown factor, not included in the model (e.g., crystallized abilities) which affects both verbal and quantitative reasoning. But, like the earlier analysis, this attempt at estimating a strict Binet model suggested that the model shown in Figure 1 does not adequately explain the standardization data. Although this analysis did result in a significant decrease in χ^2 in comparison the model in Figure 3 (change=52.61, *df*=1, *p*<.001), the correlation between the errors was small and *negative*, hardly suggesting that the verbal and quantitative factors measure something in common other than *g*. In contrast, a model with correlated errors between the abstract/visual and the quantitative factors produced a significant χ^2 decrease (115.64 at 1 *df*, *p*<.001) and also showed a significant, positive correlation between the errors of measurement of the quantitative and abstract/visual factors (*r*=.13).

For the final step in this set of analyses with the overall standardization sample, we relaxed several of the first-order factor loadings to be consistent with the first-order factor analyses. That is, Memory for Sentences was allowed to load on the verbal in addition to the memory factor, Bead Memory and Absurdities were allowed to load on the abstract/reasoning in addition to the verbal and memory factors, respectively, and we freed the correlation between the errors of measurement of the quantitative and abstract/visual factors.

The results of this final analysis are shown in Figure 5. The fit statistics shown beneath the model suggest that it provides a good fit to the standardization data; the adjusted goodness of fit index was .904 and the rms was .037. This relaxed model also provided a significantly better fit to the standardization data than did the previous model (χ^2 change = 535.26, df=3, p<.001). All factor loadings (and the correlation among errors) were significant, and the first-order factor loadings were almost identical to those shown for the relaxed first-order factor model in Table 1. The quantitative reasoning factor had the highest loading on the second-order g factor (.92), but all of the first-order factors had strong g loadings; the lowest was .80 by the short-term memory factor.

Insert Figure 5 about here

We also calculated the loading of each of the tests on the second-order factor (or the total effects of g on each of the tests). These results, shown in Table 5, suggest that (given the adequacy of the model in Figure 4) the Number Series, Vocabulary, Quantitative, Comprehension, and Matrices tests provide the five best measures of g for the total sample.

Insert Table 5 about here

Ages Two Through Six

Because the first-order factor structures for ages 12--adult and 7--11 were so similar to that shown for the total standardization sample, we conducted hierarchical analyses only for the youngest age group. Based on the first-order analyses, we excluded the memory factor, and based on the results of the hierarchical analyses for the total sample, we only included two levels of factors (no second-order crystallized factor).

The factor model and the results of the intial analysis with the 2-6 year-olds is shown in Figure 5. Similar to the first-order analyses, this initial no-memory model provided an excellent fit to the standardization data (adjusted goodness of fit index=.967, rms=.026); the correlations predicted by the model differed from the actual correlations by only .026, on average!

Insert Figure 6 about here

The first-order factor loadings are identical to those shown in Table 4 and Figure 3. Interestingly, there was more variability among the first-order factors' g loadings for these young children than there was for the total sample. The abstract/visual (or fluid intelligence) factor was almost identical to g, with a factor loading of .98. Quantitative and verbal reasoning had somewhat lower loadings (.81 and .77).

We also estimated models which allowed correlations among the unique variance of the first-order factors (first freeing the verbal--quantitative correlation and then the quantitative--abstract correlation). Although the values obtained for these correlations were very similar to those found for the overall sample (-.12 for the verbal-quantitative correlation and .13 for the quantitative--abstract/visual), neither resulted in a significant decrease in χ^2 over the model shown in Figure 6.

The second column in Table x shows the loadings of each of the Binet tests on the g factor for the no-memory model from Figure 6. Although the Quantitative test had the highest g loading (.75), this value is probably inflated because the Quantitative test--quantitative factor loading was constrained to the square root of the reliability. The next four highest g loadings were by Pattern Analysis, Copying, Bead Memory, and Comprehension.

Conclusions

The Fourth Edition of the Stanford-Binet is the latest descendant in this long line of traditional intelligence tests. It is also probably the greatest departure from the previous versions of the test, so much so, in fact, that it should probably best be considered as a completely new instrument rather than a simple revision of the earlier Binets. The new Binet has a strong basis in intelligence theory. The theory underlying the test is well explained in the various Administration and Technical manuals, and there is theoretical support for the Binet theory, both through research on the nature of *g* (e.g., TM, chap. 1), and through its obvious relation to Cattell and Horn's theory of fluid and crystallized intelligence (Cattell, 1982; Horn & Cattell, 1966). But while the TM has attempted to support the construct validity of the new Binet, there is little interpretable evidence to support the correspondence between the structure of the new Binet and the theory underlying its development. In particular, the factor analytic evidence reported in the TM is vaguely and inadequately explained; the results as presented are essential uninterpretable. The purpose of the present study was to test more explicitly the match between the factor structure of the Binet and the theory underlying the scale. To do so, we performed confirmatory factor analysis on the entire Binet standardization sample and on three, more homogeneous age groups from that standardization sample.

The results of these analyses offer mixed support for the construct validity of the new Binet. For the total standardization sample, and for 2 of the 3 subsamples, the first level of the Binet theory provided a generally good fit to the standardization data. And with relatively few changes, changes which were consistent with the Binet theory, the factor model derived from the Binet theory provided an even better fit to the standardization data. Thus, for most age levels, it appears the new Binet indeed does measure verbal reasoning, quantitative reasoning, abstract/visual reasoning, and short-term memory. This conclusion is supported for the adolescent and adult (ages 12-23) and the elementary school (ages 7-11) age groups from the standardization sample, as well as the entire standardization sample (ages 2-23).

The very strong correlations among first-order factors (almost all above .75 and several above .90) also suggest that the new Binet is a strong measure of _g_, or general intelligence, a central feature of both this Binet and its predecessors. And this finding is further confirmed by the results of the hierarchical analyses; all first-order factors had large loadings on the second-order _g_ factor, and, as a result, it appears that each of the individual tests has a substantial _g_ component.

Our analyses also highlight a number of inconsistencies between the Binet and its theory. One significant departure from the Binet theory was shown for the preschoolers from the standardization sample (ages 2 through 6). Although we were able to fit the first level of the Binet model to the 2-6 standardization data, there were obvious problems with the estimated solution. In particular, it appeared difficult, if not impossible, to separate memory from verbal and abstract/visual reasoning for this age group. In contrast, a model in which the memory factor was deleted and the memory tests were allowed to load on the verbal reasoning and the abstract/visual reasoning factors provided an excellent fit to the 2-6 standardization data. The reason for the lack of differentiation in memory and reasoning for these preschool children may be a function of the Binet memory measures. Only two tests--Bead Memory and Memory for Sentences--were used to assess memory for the 2 to 6-year-olds, and both also had significant reasoning loadings for the older age groups in the standardization sample. Thus, one reason for the lack of separation in memory and reasoning may be that the Binet does not assess memory well for this age group. On the other hand, other research has suggested that cognitive abilities are less differentiated for younger than for older children (e.g., Garrett, 1946/1965). Furthermore, exploratory factor analyses of other test batteries have often failed to find a discrete memory factor. Although the WPPSI and WISC-R use similar tests, including several memory measures, factor analyses of the WPPSI have generally produced Verbal and Performance factors (Carlson & Reynolds, 1981) whereas those of the WISC-R have generally produced three factors (e.g., Kaufman, 1975), with the third factor often interpreted as a memory factor (e.g., Jensen & Reynolds, 1982). Similarly, analyses of the K-ABC have often suggested a one-factor solution for its intelligence subtests for younger age groups (Kaufman & Kamphaus, 1984, p. 631), whereas a factor interpretable as verbal memory (cf. Keith, 1985) emerges for older age groups. Still, a similar memory or sequential factor can be forced even for younger age groups (cf. Kaufman & Kamphaus, 1984) Nevertheless, such results do suggest that the reason for the lack of differentiation between the Binet's memory and reasoning factors for ages 2-6 may be that memory skills are undifferentiated from reasoning skills for such young children.

A comparison of the g loadings for the total group versus the youngest age group also suggests some age-related differences. For the total sample the first-order factors all had high and similar loadings on g, but for the preschool children the loading of the first-order abstract/visual reasoning factor was considerably higher on g than were the other two factors. Similarly, verbal tests were less important measures of general intelligence for younger children than they were for the standardization sample as a whole. Obviously, further research will be needed to determine whether these differences are unique to the Binet tests or whether they represent a more general characteristic of young children.

Neither the factor correlations in the first-order analysis nor the difficulties resulting from the inclusion of a crystallized factor were particularly supportive of the Binet theory. The theory underlying the new Binet is presumably based on Cattell and Horn's theory of crystallized and fluid intelligence. The Binet authors suggest that the Verbal and Quantitative Reasoning Scales measure crystallized ability or intelligence, whereas the Abstract/Visual Scale measures fluid intelligence. If so, we would expect the correlations between the first-order verbal and quantitative factors to be consistently higher than the correlations between these two factors and the abstract/visual or short-term memory factors. In contrast, the verbal--quantitative factor correlations were generally among the smaller factor correlations for all age levels examined, and the quantitative and abstract/visual factors generally correlated at a considerably higher level. Furthermore, a hierarchical factor model which mirrored the four levels of the Binet theory--and, in particular, the combination of the verbal and quantitative factors into a crystallized factor--was untenable. Subsequent analyses suggested that the verbal and quantitative factors are not affected by any common factor other than g, but that quantitative and abstract/visual reasoning may share skills other than general intelligence. Our results do not reveal whether this discrepant finding is due to inadequacies in the theory which quided the construction of the Binet or whether it is due to problems in implementing that theory through the new scale. Nevertheless, these findings do suggest an inconsistency between the new scale and the theory which quided its construction.

These results, then, offer mixed support for the construct validity of the new Binet. The first-order factors, as represented by the Verbal Reasoning, Quantitative Reasoning, Abstract/Visual Reasoning, and Short-Term Memory scales are generally well supported by these confirmatory analyses, with somewhat less support shown for the independence of the Short-Term Memory scale, especially for young children. These results also offer strong support for the general intellectual, or g, factor at the apex of the Binet theory. The present results do not support the middle level of the Binet theory: the division of the scales into measures of crystallized versus fluid intelligence. And although this level of the theory is not reflected in the actual structure of the scale (the area scores are not converted into crystallized or fluid scores), these departures are nevertheless pertinent because they suggest an inconsistency between the scale and the theory which guided its construction.

The results of these analyses also have implications for users of the new Binet. Based on these results, it appears those users can be confident that the Composite score from the Binet provides a good estimate of g. These results suggest that users of this scale may also confidently interpret the Verbal Reasoning, Quantitative Reasoning, and Abstract/Visual scales as estimates of similarly named abilities. For children ages 7 and above, the Memory tests also appear to provide a valid measure of short-term memory, although several of the Memory tests also seem to measure verbal reasoning (Memory for Sentences) and visual reasoning (Bead Memory). For young children, we found little support for interpreting the Memory tests as a separate scale; rather, the two

Memory tests appear to require verbal and abstract/visual reasoning rather than short-term memory at ages 6 and below.

These results also suggest that caution is needed in interpreting the second level of the Binet theory. Our analyses do not support the interpretation of the Verbal and Quantitative scales as measures of crystallized ability which is then contrasted with fluid ability.

References

Bentler, P. M. (1980). Multivariate analysis with latent variables: Causal modeling. Annual Review of Psychology, 31, 419-456.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 88, 588-606.

Carlson, L., & Reynolds, C. R. (1981). Factor structure and specific variance of the WPPSI subtests at six age levels. Psychology in the Schools, 18, 48-54.

Cattell, R. B. (1982). The inheritance of personality and ability: Research methods and findings. New York: Academic.

Garrett, H. E. (1965). A developmental theory of intelligence. In A. Anastasi (Ed.), Individual differences (pp. 73-83). New York: Wiley. (Reprinted from American Psychologist, 1946, 1, 372-378)

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. Journal of Educational Psychology, 57, 253-270.

Jensen, A. R., & Reyonolds, C. R. (1982). Race, social class and ability patterns on the WISC-R. Personality and Individual Differences, 3, 423-438.

Jöreskog, K. G., & Sörbom, D. (1979). Advances in factor analysis and structural equation models. Cambridge, MA: Abt.

Jöreskog, K. G., & Sörbom, D. (1984). LISREL VI: Analysis of linear structural relationships by the method of maximum likelihood: User's guide. Mooresville, IN: Scientific Software.

Kaufman, A. S. (1975). Factor analysis of the WISC-R at 11 age levels between 6 1/2 and 16 1/2 years. Journal of Consulting and Clinical Psychology, 43, 135-147.

Kaufman, A. S., & Kamphaus, R. W. (1984). Factor analysis of the Kaufman Assessment Battery for Children (K-ABC) for ages 2 1/2 through 12 1/2 years. Journal of Educational Psychology, 76, 623-637.

Keith, T. Z. (1985). Questioning the K-ABC: What does it measure? School Psychology Review, 14, 9-20.

Keith, T. Z., & Dunbar, S. B. (1984). Hierarchical factor analysis of the K-ABC: Testing alternate models. The Journal of Special Education, 18, 367-375.

Kerlinger, F. N. (1986). Foundations of behavioral research (3rd ed.). New York: Holt, Rinehart and Winston.

Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First- and higher-order factor models and their invariance across groups. Psychological Bulletin, 97, 562-582.

MacCallum, R. (1986). Specification searches in covariance stucture modeling. Psychological Bulletin, 100, 107-120.

National Opinion Research Center. (1980). High School and Beyond Information for Users: Base year (1980) data (Report to National Center for Education Statistics, Contract No. 300-78-0208). Chicago: Author.

Newman, R. S. (1984). Children's achievement and self-evaluations in mathematics: A longitudinal study. Journal of Educational Psychology, 76, 857-873.

Sandoval, J., & Irvin, M. G. (1986, November). A review: Stanford-Binet Intelligence Scale: Fourth Edition technical manual. CASP Today, 36(2), 3-5.

Slate, J. R. (1986, September). A reaction to the revised Stanford-Binet Intelligence Scale: New does not necessarily mean better. NASP Communique, 15(1), 3.

Strommen, E. F. (in press). Confirmatory factor analysis of the Kaufman Assessment Battery for Children (K-ABC): A reevaluation. Journal of School Psychology.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986).

Technical Manual: Stanford-Binet Intelligence Scale: Fourth Edition. Chicago: Riverside.

Footnotes

[1] The intercorrelation matrix on page 53 of the Technical Manual excludes several correlations because the pairs of tests involved were only given to a few of the age levels in the standardization sample. We calculated these missing median correlations from the correlation matrices for each age presented in pages 110-126 of the TM. Throughout this manuscript we have capitalized the names of tests and scales. Factor names and the abilities presumably measured by the various tests are not capitalized.

[2] Several reviewers have been critical of the weighting procedure used to control for the overrepresentation of high SES children in the standardization sample (Sandoval & Irvin, 1986; Slate, 1986; the weighting procedure is described in chapters 3 and 4 of the TM). In contrast, we agree with the publishers that this is probably the best method of dealing with over- and under-representation in the sample. It would certainly be preferable to have all groups represented to the same extent as in the U.S. population, but given this lack of consistency, the weighting procedure is better than the other two possible alternatives: discarding usable data from the over-represented groups or calculating scores without regard to the SES differences. Similar, albeit more

complex and better described, weighting procedures are common among national data sets (e.g., National Opinion Research Center, 1980, chap. 4). On the other hand, we can only assume that the correlations matrices are based on the weighted norms, and it is unclear whether the sample sizes as presented in the same tables are weighted or unweighted.

[3] The "goodness-of-fit" statistics provide a measure of how well the data fit the proposed model. If the chi-square is large in comparison to the degrees of freedom, the null hypothesis that the model fits the data is rejected. Thus, a small χ^2 is desired. An associated probability can also be computed for the χ^2, but both the χ^2 and the probability are very dependent on sample size, with even good models being rejected with large samples such as those used here. The χ^2 is therefore not recommended as the significance test when large samples are used, but can be used for making comparisons between two nested models (Bentler & Bonett, 1980). A goodness-of-fit index is also calculated which increases as the model better fits the data; the adjusted goodness-of-fit index reported here ranges from zero to one with a value of one indicating a perfect fit. Perhaps the best measure of the fit of the model is shown by the root mean square residual correlation (rms). This

statistic compares the original correlation matrix with the correlations predicted by the model; the average of these differences is the root mean square residual. An rms below about .10 is generally considered to suggest a good fit (Kerlinger, 1986, chap. 36). For more information, see Bentler (1980), Bentler and Bonett (1980), Jöreskog and Sörbom (1979, 1984), or Kerlinger (1986). For examples of the technique, see Keith and Dunbar (1984) or Marsh and Hocevar (1985).

The sample size for these analyses varies depending on the pairs of tests correlated. For example, for the composite, the sample size varies from 80 for the Copying--Equation Building correlation to 4,979 for the Vocabulary--Comprehension correlation. We calculated the minimum sample size for each age level and added the appropriate age levels together to derive the sample size used in each analysis. Thus, we used intermediate but conservative sample size estimates.

[4]The method used for conducting hierarchical analysis was a little different than that outlined in the LISREL manual and was based on a procedure outlined by Marsh and Hocevar (1985). Briefly, all factors--first, second, and third order--were specified as Etas, but the tests were allowed to load only on the first order factors. The loadings of the first order factors on higher order

factors were then estimated in the Beta matrix (for more information, see Marsh & Hocevar)

Table 1

Confirmatory Factor Analysis of the Stanford Binet for the Entire Standardization Sample

Factor, Test, or Fit	Strict Model	Relaxed Model
	Factor Loadings	
Verbal Reasoning to:		
Vocabulary	.89	.91
Comprehension	.82	.82
Absurdities	.70	.47
Verbal Relations	.77	.77
Memory for Sentences	--	.40
Quantitative Reasoning to:		
Quantitative	.82	.82
Number Series	.85	.85
Equation Building	.72	.72

(table continues)

Factor, Test, or Fit	Strict Model	Relaxed Model
Factor Loadings		
Abstract/Visual Reasoning to:		
Pattern Analysis	.73	.74
Copying	.59	.60
Matrices	.77	.77
Paper Folding & Cutting	.75	.75
Bead Memory	--	.49
Absurdities	--	.27
Short-Term Memory to:		
Bead Memory	.70	.27
Memory for Sentences	.75	.43
Memory for Digits	.67	.76
Memory for Objects	.58	.62
Factor Correlations		
Verbal Reasoning with:		
Quantitative Reasoning	.79	.77
Abstract/Visual Reasoning	.80	.77
Short-Term Memory	.83	.68

(*table continues*)

Factor, Test, or Fit	Strict Model	Relaxed Model
	Factor Correlations	
Quantitative Reasoning with:		
Abstract/Visual Reasoning	.92	.90
Short-Term Memory	.84	.75
Abstract/Visual Reasoning with:		
Short-Term Memory	.82	.68
	Goodness-of-Fit Statistics	
Chi-Square	2189.22	1636.95
_df_	84	81
Adjusted Goodness-of-Fit	.879	.904
Root Mean Square residual	.044	.037

Note. _N_ = 3,354.

Table 2

Confirmatory Factor Analysis of the Stanford Binet for Twelve to Twenty-three Year Olds

Factor, Test, or Fit	Strict Model	Relaxed Model
	Factor Loadings	
Verbal Reasoning to:		
Vocabulary	.90	.92
Comprehension	.83	.82
Absurdities	.72	.37
Verbal Relations	.75	.76
Memory for Sentences	--	.44
Quantitative Reasoning to:		
Quantitative	.85	.85
Number Series	.85	.85
Equation Building	.69	.69

(table continues)

Factor, Test, or Fit	Strict Model	Relaxed Model
Factor Loadings		
Abstract/Visual Reasoning to:		
Pattern Analysis	.74	.75
Copying	.67	.68
Matrices	.81	.81
Paper Folding & Cutting	.72	.71
Bead Memory	--	.52
Absurdities	--	.40
Short-Term Memory to:		
Bead Memory	.73	.25
Memory for Sentences	.75	.38
Memory for Digits	.69	.77
Memory for Objects	.60	.65
Factor Correlations		
Verbal Reasoning with:		
Quantitative Reasoning	.85	.83
'bstract/Visual Reasoning	.84	.79
Short-Term Memory	.86	.70

(table continues)

Factor, Test, or Fit	Strict Model	Relaxed Model
Factor Correlations		
Quantitative Reasoning with:		
Abstract/Visual Reasoning	.93	.92
Short-Term Memory	.88	.79
Abstract/Visual Reasoning with:		
Short-Term Memory	.86	.73
Goodness-of-Fit Statistics		
Chi-Square	924.46	752.02
df	84	81
Adjusted Goodness-of-Fit	.822	.850
Root Mean Square residual	.051	.045

Note. *N* = 910.

Table 3

Confirmatory Factor Analysis of the Stanford Binet for Seven to Eleven Year Olds

Factor, Test, or Fit	Strict Model	Relaxed Model
	Factor Loadings	
Verbal Reasoning to:		
Vocabulary	.86	.88
Comprehension	.83	.84
Absurdities	.69	.36
Memory for Sentences	--	.44
Quantitative Reasoning to:		
Quantitative	.65	.66
Number Series	.70	.70
Abstract/Visual Reasoning to:		
Pattern Analysis	.74	.76
Copying	.58	.60
Matrices	.71	.70
Bead Memory	--	.50
Absurdities	--	.41

(table continues)

Factor, Test, or Fit	Strict Model	Relaxed Model
	Factor Loadings	
Short-Term Memory to:		
Bead Memory	.66	.22
Memory for Sentences	.77	.43
Memory for Digits	.64	.74
Memory for Objects	.55	.60
	Factor Correlations	
Verbal Reasoning with:		
Quantitative Reasoning	.86	.82
Abstract/Visual Reasoning	.75	.7[illegible]
Short-Term Memory	.83	.62
Quantitative Reasoning with:		
Abstract/Visual Reasoning	1.01	.97
Short-Term Memory	.89	.77
Abstract/Visual Reasoning with:		
Short-Term Memory	.83	.67

(table continues)

Factor, Test, or Fit	Strict Model	Relaxed Model
Goodness-of-Fit Statistics		
Chi-Square	435.58	239.00
df	48	45
Adjusted Goodness-of-Fit	.873	.926
Root Mean Square residual	.050	.036

Note. *N* = 960.

Table 4

Confirmatory Factor Analysis of the Stanford Binet for Two to Six Year Olds

Factor, Test, or Fit	Strict Model	Constrained Model	No-Memory Model
	Factor Loadings		
Verbal Reasoning to:			
Vocabulary	.81	.81	.81
Comprehension	.8	.82	.82
Absurdities	.76	.76	.75
Memory for Sentences	--	--	.69
Quantitative Reasoning to:			
Quantitative	.92[a]	.92[a]	.92[a]
Abstract/Visual Reasoning to:			
Pattern Analysis	.72	.72	.72
Copying	.70	.70	.71
Bead Memory	--	--	.69
Short-Term Memory to:			
Bead Memory	.59	.57	--
Memory for Sentences	.61	.61	--

(table continues)

Factor, Test, or Fit	Strict Model	Constrained Model	No-Memory Model
	Factor Correlations		
Verbal Reasoning with:			
Quantitative Reasoning	.62	.62	.63
Abstract/Visual Reasoning	.79	.79	.75
Short-Term Memory	.97	1.00[b]	--
Quantitative Reasoning with:			
Abstract/Visual Reasoning	.79	.79	.80
Short-Term Memory	.86	.86	--
Abstract/Visual Reasoning with:			
Short-Term Memory	1.02	1.00[b]	--
	Goodness-of-Fit Statistics		
Chi-Square	262.46	265.58	105.52
df	16	18	19
Adjusted Goodness-of-Fit	.906	.917	.969
Root Mean Square residual	.045	.045	.026

Note. *N* = 1,484.

[a]The quantitative reasoning--Quantitative test factor loading was fixed to the square root of the reliability of the Quantitative test; $\sqrt{.84} = .916$.

[b]These two factor correlations were fixed to 1.0 for the constrained model.

Table 5

Total Effects of g on the 15 Binet Tests for the Total Sample and for the Preschool Sample

Tests	Total Sample	Two-Six Year Olds
Vocabulary	.77	.62
Comprehension	.70	.63
Absurdities	.64	.57
Verbal Relation	.65	-
Pattern Analysis	.66	.71
Copying	.53	.69
Matrices	.69	-
Paper Cutting	.67	-
Quantitative	.75	.75
Number Series	.78	-
Equation Building	.65	-
Bead Memory	.65	.67
Memory for Sentences	.68	.53
Memory for Digits	.60	-
Memory for Objects	.50	-

Figure Captions

Figure 1. The hierarchical theoretical structure underlying the Stanford-Binet Fourth Edition. Tests are shown on the right side; the increasingly global abilities (latent factors) presumably measured by the test are shown to the left of the observed tests.

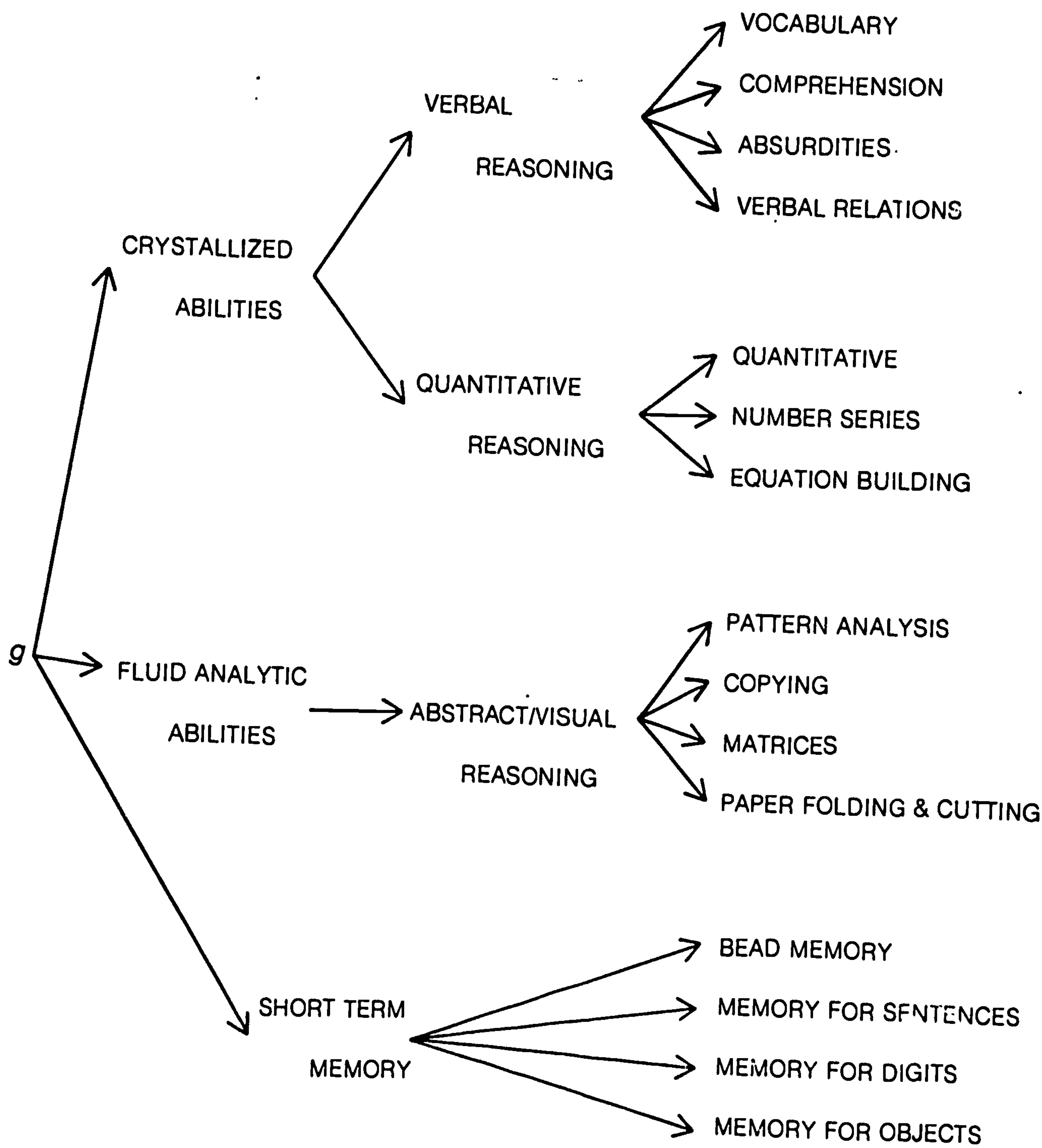
Figure 2. Confirmatory factor solution for the entire Binet standardization sample. The results shown are for the "strict" Binet model, one which corresponds exactly to the first two levels of the Binet's theoretical structure shown in Figure 1.

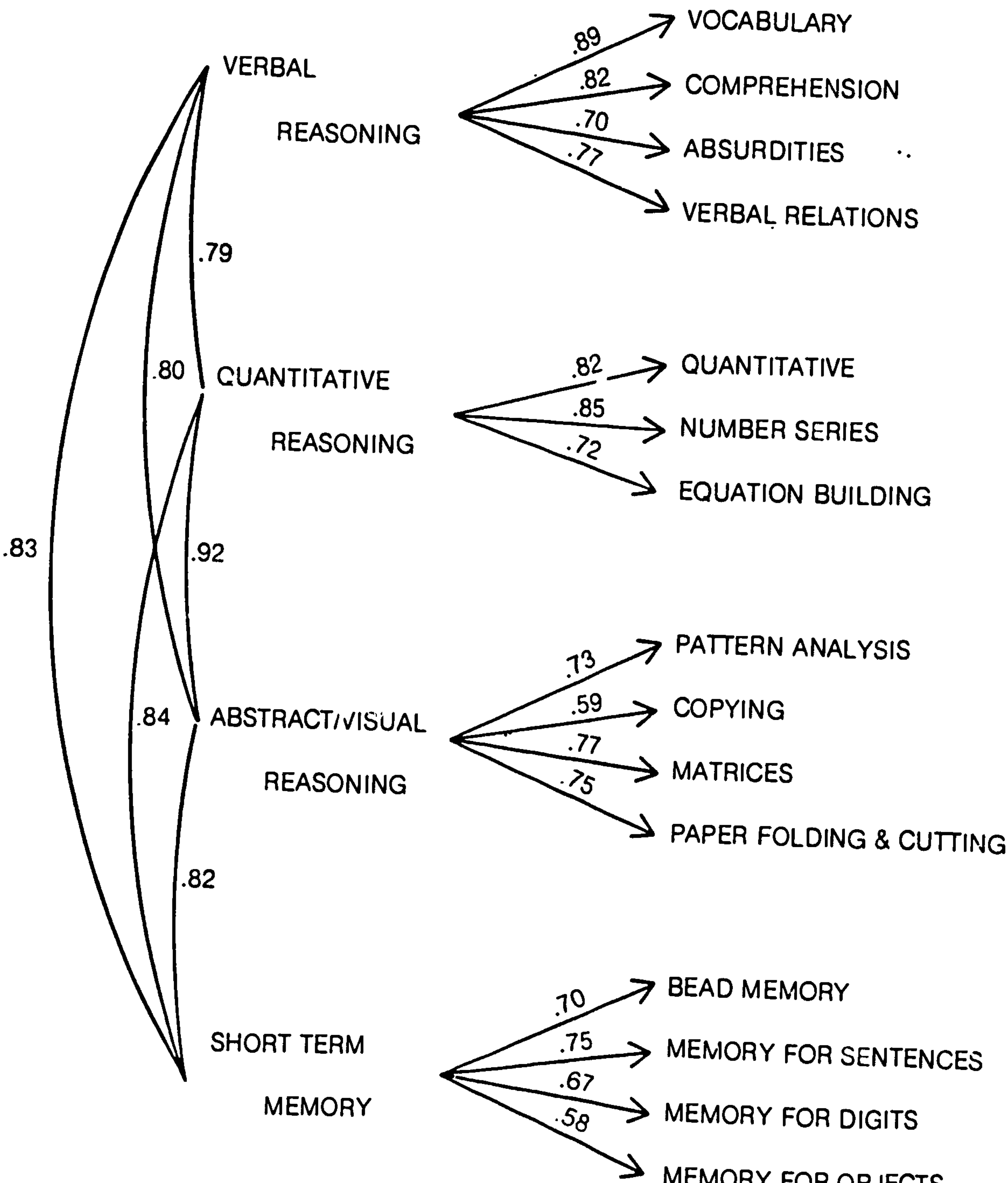
Figure 3. Confirmatory analysis for 2- to 6-year-olds. The model departs from the Binet's theoretical structure by excluding a short-term memory factor.

Figure 4. Hierarchical factor structure of the Binet for the entire standardization sample. This model does not include a third order crystallized factor as predicted by the Binet theory.

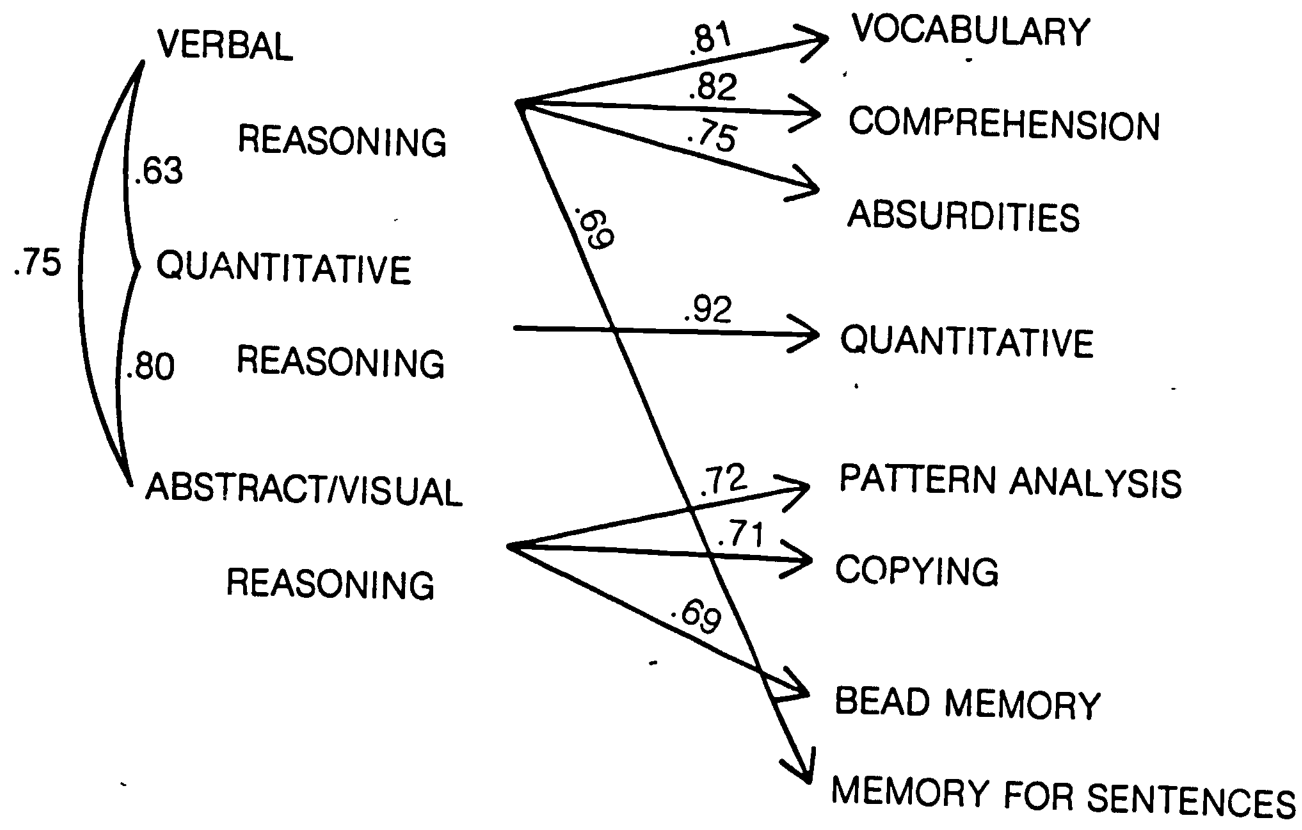
Figure 5. A relaxed hierarchical factor model of the Binet.

Figure 6. The hierarchical structure of the Binet for children ages 2 through 6. No short-term memory factor is included.

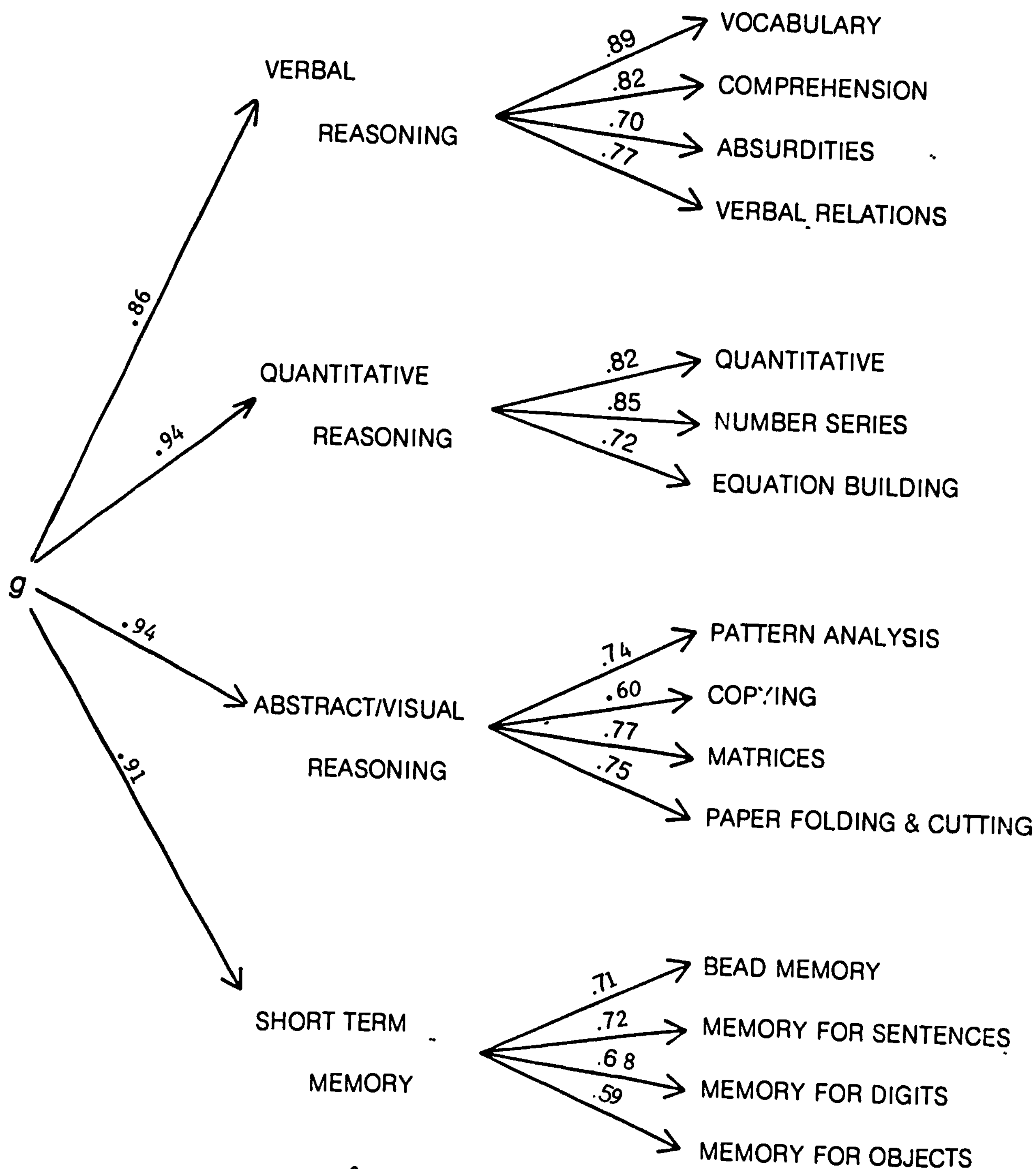

VOCABULARY
COMPREHENSION
ABSURDITIES
VERBAL RELATIONS
VERBAL REASONING
CRYSTALLIZED ABILITIES
QUANTITATIVE REASONING
QUANTITATIVE
NUMBER SERIES
EQUATION BUILDING
g
FLUID ANALYTIC ABILITIES
ABSTRACT/VISUAL REASONING
PATTERN ANALYSIS
COPYING
MATRICES
PAPER FOLDING & CUTTING
SHORT TERM MEMORY
BEAD MEMORY
MEMORY FOR SENTENCES
MEMORY FOR DIGITS
MEMORY FOR OBJECTS



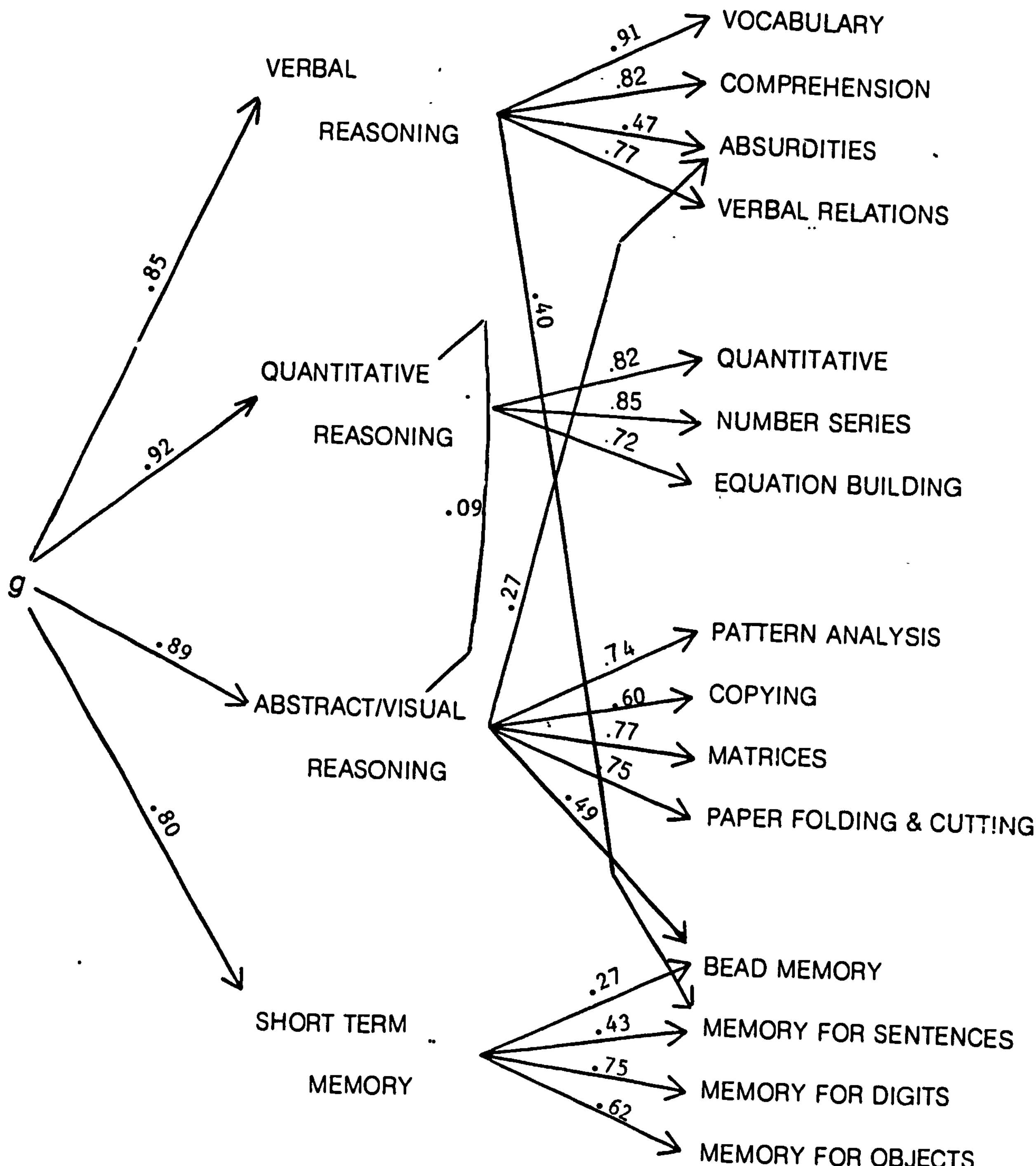
χ^2	2189.22
df	84
Fit Index	.879
rms	.044



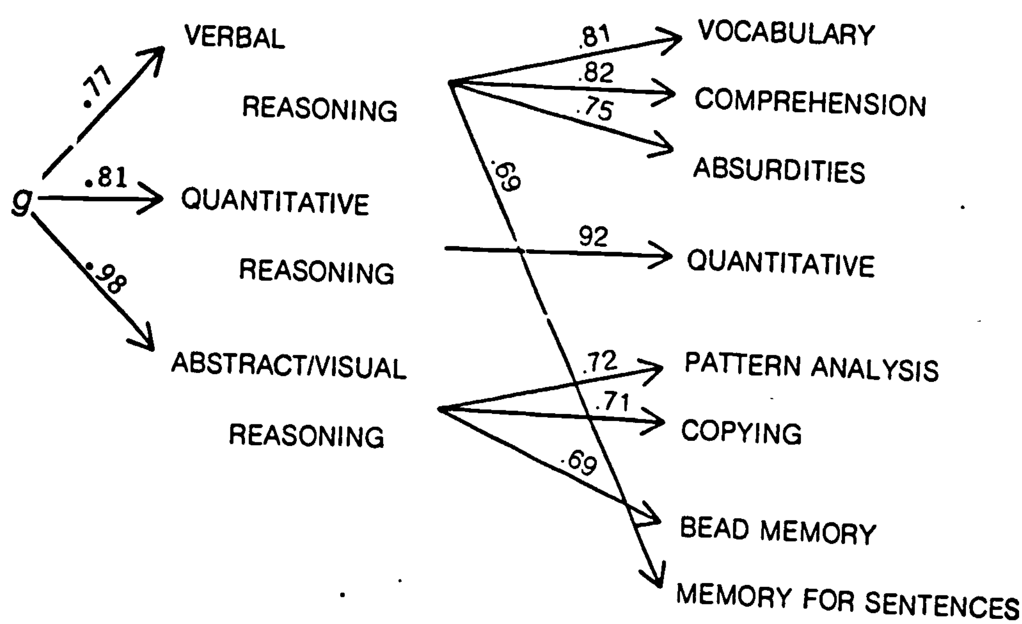
χ^2	105.52
df	19
Fit Index	.969
rms	.026



χ^2	2310.81
df	86
Fit Index	.875
rms	.045



χ^2	1659.91
df	82
Fit Index	.904
rms	.037



χ^2	105.52
df	18
Fit Index	.967
rms	.026