DOCUMENT RESUME

ED 388 699 TM 024 090

AUTHOR	Witta, E. Lea; Keith, Timothy Z.
TITLE	Multi-Sample Hierarchical Confirmatory Factor Analysis of the WISC-R: An Old Problem Revisited.
PUB DATE	Nov 94
NOTE	29p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (Nashville, TN, November 9-11, 1994).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Age Differences; Analysis of Covariance; Chi Square; *Construct Validity; Factor Structure; *Goodness of Fit; *Intelligence Tests; *Sampling
IDENTIFIERS	*Confirmatory Factor Analysis; *Hierarchical Analysis; LISREL Computer Program; Standardization; Wechsler Intelligence Scale for Children (Revised)

ABSTRACT

Although the Wechsler Intelligence Scale for Children-Revised (WISC-R) is being rapidly replaced by the third edition of the WISC, questions concerning the construct validity of the WISC-R have not yet been resolved, including the number of factors it measures and whether the same constructs fit across all age levels. This study sought to determine whether the WISC-R measures the same constructs across age levels, what constructs it does measure, and how many constructs provide the best fitting model. Multi-sample, hierarchical confirmatory factor analyses using the LISREL computer program (version 7.2) were performed on the WISC-R standardization data. This sample consisted of 2,200 subjects, 200 in each of 11 age groups from 6.5 to 16.5 years. The covariance matrices for the 11 age levels were statistically indistinguishable (p>.05). The test did measure the same constructs across ages. The three-factor model provided a statistically better fit than the two-factor model, and a more parsimonious fit than the four-factor model. In addition, the three-factor model produced a consistently good fit as tested by chi-square holding both measurement and error matrices invariant across all 11 age groups. (Contains 3 tables, 4 figures, and 11 references.) (Author/SLD)

Multi-Sample Hierarchical Confirmatory Factor Analysis of the WISC-R: An Old Problem Revisited

E. Lea Witta
East Tennessee State University
Department of Educational Leadership

Timothy Z. Keith
Alfred University
Department of School Psychology

Paper presented to the annual conference of the Mid-South Educational Research Association in Nashville, TN, November, 1994.
Correspondence should be sent to Lea Witta, Statistical Analysis & Evaluation, 19556 Stone Mountain Road, Abingdon, VA 24210 - (703) 628-9760.

Running Head: WISC-R

difference in fit, the three factor model yielded a more parsimonious fit than the four factor model. In addition, the three factor model produced a consistently good fit as tested by χ^2 ($p>.05$) holding both measurement and error matrices invariant across all eleven age groups.

Abstract

Although the Wechsler Intelligence Scale for Children - Revised (WISC-R) is being rapidly replaced by the third edition of the WISC, questions concerning the construct validity of the WISC-R have not yet been resolved. Does it measure two factors? Does it measure three factors? Do the same constructs fit across all age levels?

This study sought to determine (a) whether the WISC-R measures the same constructs across its age span, (b) what constructs are measured, and (c) how many constructs provide the best fitting model. Multi-sample, hierarchical confirmatory factor analyses (LISREL 7.2) were performed on the WISC-R standardization data. This sample consisted of 2200 subjects, 200 subjects in each of the eleven age groups (ages 6 1/2 to 16 1/2).

The covariance matrices for the 11 age levels were statistically indistinguishable (p>.05). The test does measure the same constructs across ages. The three factor model provided a statistically (χ^2_{diff}=92.46, df=11, p<.01) better fit than the two factor model. Although there was no statistically significant

The latest version of the Wechsler Intelligence Scale for Children, the WISC-III, is rapidly replacing the prior revised version, the WISC-R. Questions, however, concerning the WISC-R have still not been resolved. Does the WISC-R measure the same constructs across its eleven age groups? Does the WISC-R measure two or three constructs? What are the construsts that are being measured?

Despite wide use and extensive research, it is not clear whether the WISC-R measures the same abilities across its 11 year age span. According to Kaufman (1979), Piaget's theory of the development of intelligence dictates that different tests be used to measure intelligence at different ages. This position is strengthened by evidence that intelligence changes with age (Garrett, 1965). If the same tests are used across age groups, these findings suggest _different_ constructs would be measured for the age groups. The WISC-R, however, uses the same subtests for all 11 age groups producing a final general intelligence measure, _g_.

It is also unclear whether the WISC-R measures two factors or includes an additional third factor. Two constructs, Verbal and Performance, are suggested by the WISC-R manual (Wechsler, 1974). These constructs, in turn, yield a second-order general intelligence factor, *g* (see Figure 1).

Insert Figure 1 about here

Three constructs (Verbal, Performance, and Freedom from Distractibility) have been suggested by confirmatory factor analyses (Kaufman, 1975). These constructs also yield a second-order general intelligence factor, *g* (see Figure 2).

Insert Figure 2 about here

In addition, there is disagreement concerning what the additional third factor - if there is a third factor - actually measures (Jensen & Reynolds, 1982; Kaufman, 1979; Steward & Moely, 1983; Wielkiewics, 1990).

Finally, although confirmatory factor analysis has been previously performed using data produced by the WISC-R, this procedure tests only the first-order factors or constructs. Since *g* is increasingly recognized as a second or higher-order factor (Carroll, 1993), if the WISC-R measures *g*, then that structure should be tested in a hierarchical model. If confirmatory factor analysis alone is used, the loading of the first-order factors on the second-order factor, *g*, is not determined.

This research serves multiple purposes: 1) to determine if the subtests of the WISC-R measure the same components across all 11 age groups; 2) to determine how many constructs are measured by these components; 3) to determine what constructs are measured by the subtests; 4) to determine if the same hierarchical model will fit all 11 age groups; and 5) to demonstrate a relatively new method of testing the construct validity of tests.

Method

Subjects and Instrument

The WISC-R is an individually administered measure of the intellectual ability of children ages 6 1/2 to 16 1/2. The WISC-R was standardized on a nationally representative sample of 2200 children - 200 in each of the 11 age groups. All subtests were administered to each child.

Analyses

Correlation matrices and standard deviations for each group were used as input for the analyses. All analyses were conducted using the LISREL 7.2 computer program (Jöreskog & Sörbom, 1989).

To answer the first research question - whether the subtests of the WISC-R measure the same components across its age span - the covariance matrices for each group were compared. Since all subtests were administered to every age group, covariance matrices were compared across all age groups using LISREL-7.2 multi-sample analysis (Jöreskog & Sörbom, 1989, Chapter 9). The hypothesis tested was that the variance-covariance matrix of the subtests was identical across all 11 groups. This procedure determined if the distribution of scores around the mean of each subtest and the relations among subtests for each age group were identical to the corresponding distribution and relations of all other age groups. Since all matrices and constructs are included in this initial matrix, if these matrices are statistically indiscernible, then the WISC-R must measure identically across all ages (Keith, 1990). No assumption was made about the correct factor structure for these matrices. Nor was any assumption made about what was being measured. This procedure simply determined if the same things were being measured for each age group.

To answer the second question - how many constructs are measured by the WISC-R - the generalized covariance matrix for all groups was placed in a hierarchical factor model consisting of two (Wechsler, 1974), three (Kaufman, 1979), and four first-order factors leading to one second-order general ability, or *g* factor. Since factor analytic procedures traditionally explain more variance and provide a better fit when more factors are included, a four factor model (see Figure 3) was also included in this study. The four factor model was determined by traditional exploratory factor analysis using SPSS-PC+.

Insert Figure 3 about here

To answer the third question - what constructs are measured by the subtests - the first- and second-order factor loadings provided by the best-fitting factor model (determined in answering question two) were used. In addition, the total effects for each subtest were considered.

To answer the fourth question - how adequately does the same factor model fit across all age groups - elements of the best-fitting hierarchical factor model were constrained to equality and compared. This procedure consisted of three steps[1]. In the first step, the age groups were constrainted to a similar factor structure (ie, the first-order factors loaded on the same subtests across age groups). In step two, the factor loadings were constrainted to equivalence across age groups (ie, the second-order factor loadings on the first-order factor and the first-order factor loadings

[1]In LISREL terminology, for the first step, all matrices had the same pattern and starting values, but were permitted to vary otherwise. Second, Psi and Theta Epsilon were specified as having the same pattern and starting values across groups, while Gamma and Lambda Y were invariant. Third, the Gamma, Lambda Y, Psi, and Theta Epsilon matrices were all invariant across groups.

on the subtests were identical across age groups). The third step constrained the factor structure to equivalence across age groups (ie, in addition to identical factor loadings, the unique and error variances for each subtest and factor were identical across age groups).

The fit statistics provided by the LISREL program were used to judge whether the hypothesis of identical matrices should be rejected. This program produces a single chi-square (χ^2) statistic which is used to test the "fit of all LISREL models in all groups, including all constraints" (Jöreskog & Sörbom, 1989, p228). A large χ^2 indicates the model is not invariant across groups. Since the value of χ^2 is seriously inflated by sample size, however, meaningless differences between groups can result in statistical significance, leading to rejection of a good model (Hayduk, 1987). Since this sample contained 2200, the Differential Fit Value (DFV) suggested by Muthén (1989) was used as the primary criterion for decision making. The DFV[2] is a

[2]To calculate the actual χ^2, multiply the χ^2 (DVF) by 2.2 ((2200-1)/(1000-1)).

χ^2 value for a sample size of 1000. All χ^2 values reported in this study are Diffential Fit Values.

Decisions concerning the best-fitting model were made using the difference between the two χ^2 values with degrees of freedom equal to the difference in degrees of freedom of the two models (Hayduk, 1987). This difference χ^2 is distributed as χ^2 with degrees of freedom determined by the difference degrees of freedom.

Results and Discussion

Does the WISC-R Measure the Same Things Across Age Groups?

The specification that the 11 age matrices of the WISC-R were identical resulted in a χ^2 (DFV) of 453.16 (df=780, p>.99). The variance/covariance matrices of the 11 age groups were statistically indistinguishable. Whatever the WISC-R measures, it measures the identically across all age groups.

How many constructs does the WISC-R Measure?

All factor models provided a surprisingly good fit ($p \geq .05$) as measured by χ^2 (DFV). The three- and the four-factor models provided a significantly better fit

(χ^2_{dif}=92.46, df=11, $p \leq .01$; χ^2_{dif}=85.72.46, df=22, $p \leq .01$), however, than the two-factor model (see Tabel 1).

Although the χ^2 values produced by the three and four-factor models were statistically indiscernible, the χ^2 (DFV) produced by the four-factor model was <u>larger</u> than the one produced by the three-factor model. Since the three factor model provides the more parsimonious fit and is supported by theory, it is the preferred model.

Insert Table 1 about here

<u>What constructs are measured by the WISC-R?</u>

The Verbal and Performance (or perceptual) factors appear appropriately named (see Table 2 and Figure 4). The factor, however, termed Freedom from Distractability is questionable. Although the Digit Span subtest may suggest this term, factors are usually named for those items that load most heavily on them. In this instance, the Arithmetic subtest loads on the first-order factor at 0.77 and on *g* at 0.65; the Digit Span subtest loads on the first-order factor at 0.57

and on *g* at 0.48 (see Table 2 and Figure 4). This would suggest this factor is most appropriately named Quantitative Reasoning.

Insert Table 2 and Figure 4 about here

<u>How adequately does the same factor model fit across all age groups?</u>

Comparison of the three-factor model by constraining portions of it to equality resulted in a χ^2 change that was not statistically significant (see Table 2). Proceeding from a model in which the eleven age groups are constrainted to conform to a similar factor structure (ie, the same subtests load on the first order factors) through the constraint that not only are the factor loadings on each first-order factor equivalent across age groups, but the unique and error variances of the subtests and factors are also equivalent, produced no significant change in χ^2. Thus not only does the proposed three-factor structure of the WISC-R provide an excellent fit across all 11 age levels in the standardization sample, but the hierarchial factor structure - including variances - appears <u>invariant</u> across those ages.

Insert Table 3 about here

Conclusions

The findings from this study indicate that the WISC-R does measure the same constructs identically across all age levels. Three constructs, as suggested by Kaufman (1979), provide the best fitting model. The Verbal and Performance constructs appear to be appropriately named. It is suggested, however, that the construct previously termed Freedom from Distractability would be best interpreted as a measure of Quantitative Reasoning by those still using the WISC-R.

A final purpose of this study was to demonstrate a relatively new method of testing the construct validity of a test. A multi-sample test of the equivalence of the subtest covariance matrices provided a test of identical constructs across groups. Hierarchical, multi-sample, confirmatory factor analysis was then used to understand what the constructs are - or are not - for this test. These findings point to the superiority of hierarchical, rather than simple first-order, analysis for understanding constructs measured by a test. By using a hierarchical structure, the methodology used in this study provides much stronger evidence of what a test measures than does exploratory or first-order confirmatory analysis. In addition, the comparison of groups within one statistical analysis contributes a more powerful test than does a factor analysis for each group separately. This methodology is also appropriate for comparison of the factor structure of a test across ethnic or gender groups. Used in this manner, this technique would provide an extremely powerful test of construct bias.

References

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

Garrett, H.E. (1965). A developmental theory of intelligence. In A. Anastasi (Ed.), Individual differences (pp. 78-83). New York: Wiley.

Jensen, A.R. & Reynolds, 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. (1989). LISREL 7 A Guide to the Program and Applications (2nd ed) [Computer program manual]. Chicago, IL: SPSS Inc.

Kaufman, A.S. (1979). Intelligent testing with the WISC-R. New York: Wiley Interscience.

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

Keith, T.Z. (1990). Confirmatory and hierarchical confirmatory analysis of the Differential Ability Scales. Journal of Psychoeducational Assessment, 8, 391-405.

Mŭthén, B.O. (1989). Dichotomous factor analysis of symptom data. Sociological Methods & Research, 18, 19-65.

Stewart, K.J. & Moely, B.E. (1983). The WISC-R third factor: What does it mean? Journal of Consulting and Clinical Psychology, 51, 940-941.

Wechsler, D. (1974). Wechsler Intelligence Scale for Children - Revised [manual], New York: Psychological Corporation.

Wielkiewics, R.M. (1990). Interpreting low scores on the WISC-R third factor: It's more than distractibility. Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2, 91-97.

Table 1

Comparison of Two, Three, and Four Factor Models

Factors	χ^2	(df)	$^a\chi^2_{dif}$	(df)	$^b\chi^2_{dif}$	(df)
2	519.21	(572)				
3	426.75	(561)	92.46**	(11)		
4	433.79	(550)	85.72**	(22)	7.04	(11)

Note: [a] χ^2 difference from the two-factor model.

[b] χ^2 difference from the three-factor model.

* $p \leq .05$. ** $p \leq .01$.

Table 2

Total Effects of Subtest on g

Subtest	g
Information	.741
Similarities	.736
Vocabulary	.801
Comprehension	.691
Arithmetic	.653
Digit Span	.480
Coding	.386
Block Design	.647
Mazes	.414
Object Assembly	.561
Picture Completion	.527
Picture Arrangement	.480

Note. [a]Total Effects of subtest on g.

Table 3

Structure of the Three Factor WISC-R

Hypothesis Tested	χ^2 (df)	χ^2_{dif} (df)	p
Similar Factor Structure[a]	426.75 (561)		
Identical Factor Loadings[b]	488.41 (681)	61.66 (120)	>.99
Identical Factor Structure[c]	618.30 (831)	129.89 (150)	>.88
Total Change		191.55 (270)	>.99

Note. [a]Factor loadings, error, & unique variances - unconstrained.

[b]Variances unconstrained, factor loadings constrained to equality.

[c]Factor loadings & variances - constrained to equality.

Figure 1: Two Factor Model of the WISC-R

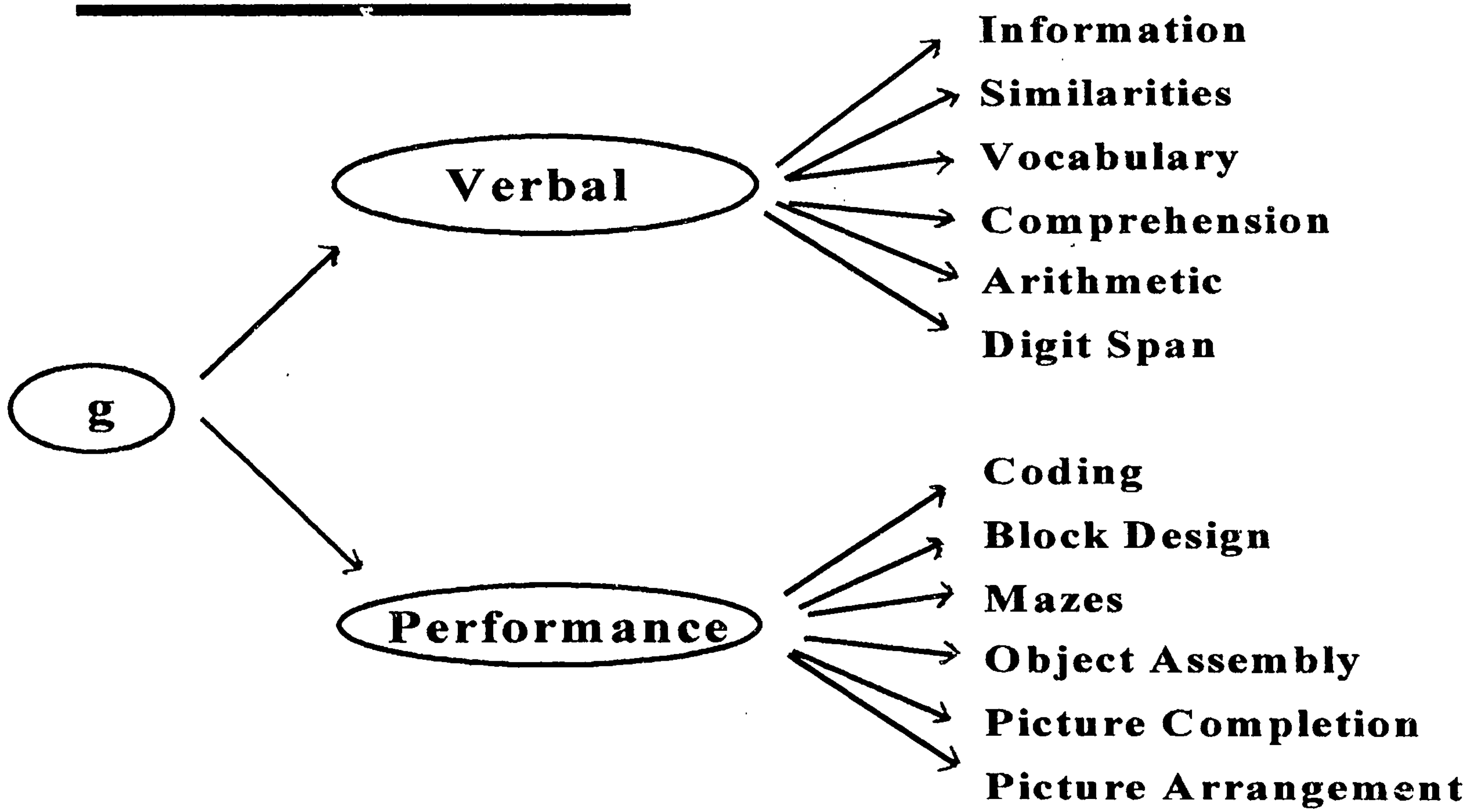

Figure 2: Three Factor Model of the WISC-R

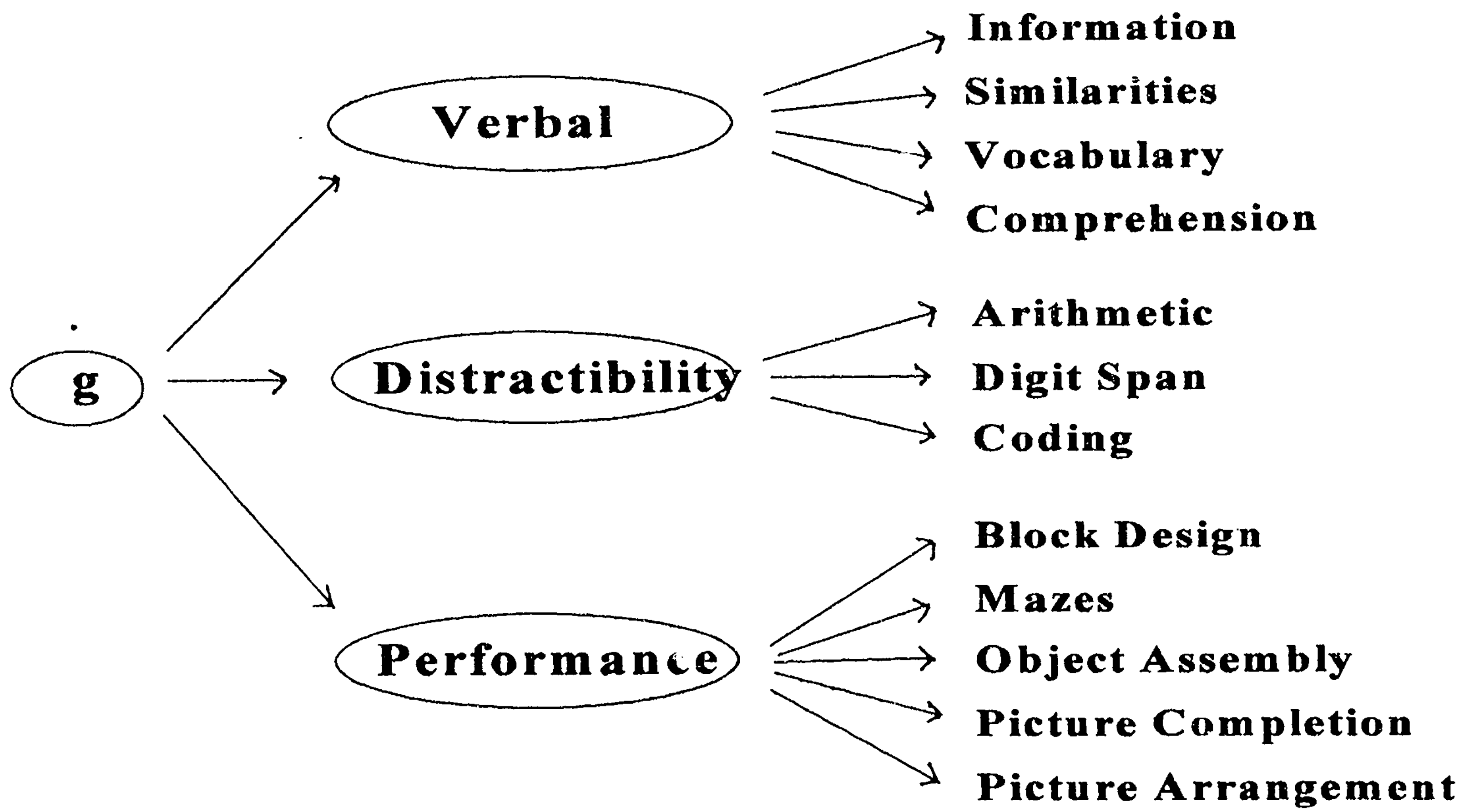

Figure 3: Four Factor Model of the WISC-R

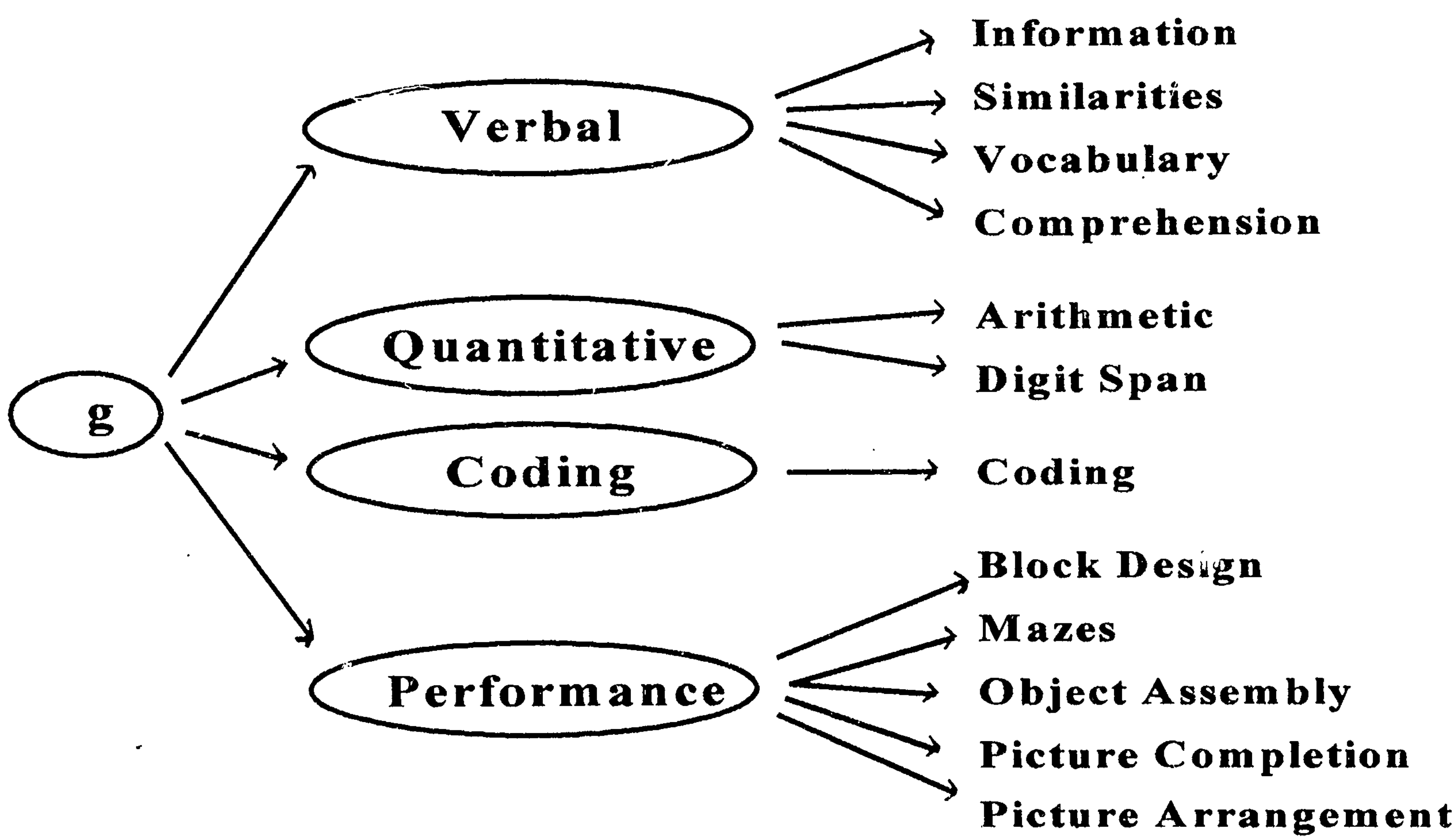

Figure 4: Proposed Three Factor Model of the WISC-R