Multivariate Experimental Clinical Research, Volume 11, No. 3, 1997, 121-130.

CONFIRMATION OF THE FACTOR PATTERN OF A BATTERY OF INTELLIGENCE AND ACHIEVEMENT INDICATORS FOR A CHILD CLINICAL POPULATION

Ralph Mason Dreger
Louisiana State University

ABSTRACT

An exploratory factor analysis of the data for 50 clinical children, from a pool of 210, on a battery of tests required of clients in a private practice clinic, including four intelligence indicators, three formal indicators of achievement, and Factor II from the Children's Behavioral Classification Project Inventory, revealed a factor pattern loading intelligence indicators on one factor and achievement indicators on another. This result was confirmed in an independent sample of 46 children by confirmatory factor analysis. As assessed by chi-square and specific indices of the EQS program, the empirically-derived covariance matrix and hypothesized covariance matrix were found to be essentially congruent. In addition, traditional matrix comparison coefficients supported the relative equality of the two factor matrices.

INTRODUCTION

A previous article, "A research-oriented private psychological clinic" (Dreger, 1990), reported among other findings one relating Factor II of the Children's Behavioral Classification Project (CBCP) Inventory (described below) to measured intelligence and scholastic achievement. Factor II is labeled among the CBCP factors, "Intellectual and Scholastic Retardation vs. Alert, Socialized Scholastic Achievement." According to current research, the factor is mislabeled.

The *generally* positive relation among intelligence tests is well known (Sattler, 1988).[1] However, that a similar relation holds between parental responses to a *behavioral* Inventory from which Factor II is derived and children's responses to standardized intelligence and achievement measures has yet to be demonstrated. It is true (Dreger, 1990), to be sure, that scores on Factor II derived from the CBCP Inventory predict almost as well to the Reading subtest of the Wide Range Achievement Test as does the Full Scale IQ of the child Wechsler tests.

This latter finding suggested that Factor II is related more closely to the achievement tests than to intelligence measures. In the exploratory factoring of the test battery (also described below), two factorial dimensions were distinctly identified as "Intelligence" and "Achievement." Factor II was much better correlated with the second factor than with the first one. Thus, the factor pattern which emerged from the analysis consisted of two factors, each primarily identified by four tests in the battery, four intelligence indicators and four achievement indicators.

The purpose of the present investigation was then to attempt to confirm the factor pattern in independent samples of the child clinical population. To ascertain, first, in one sample the factor pattern underlying this battery, i.e., its "causes" in latent variable analysis terms, and second, to confirm its existence in an independent sample are thus important and necessary tasks.

METHOD

PARTICIPANTS

Parents who applied for service in PRS signed informed consent forms for their children and they themselves responded to the research instruments required by the clinic, the author's private practice, including the CBCP Inventory describing their children's behaviors.

An original sample of 210 clinical children in a private practice clinic was reduced to 96 by eliminating all those whose protocols were incomplete, i.e., lacking even one of the tests required for research and service purposes. Even under the strict requirements of the clinic for administration of all tests to all children, for various reasons some children did not have all tests. Use of "missing values" routines was eschewed in favor of completeness, lowering, to be sure, the degrees of freedom available but guaranteeing no artifactual increase in the degrees of freedom. The children ranged in age from 4 to 13 and were almost exclusively Caucasian. The gender ratio was approximately 3 to 1, male to female, a common distribution for psychological clinics.

The 96-participant sample was divided into subsamples of 50 and 46 respectively and analyses carried out on each sample, exploratory factor analysis with the first and confirmatory analysis with the second.[2] These subsamples were deemed comparable inasmuch as the only selective factor was temporal, the first set taken from earlier cases and the second from later cases. There is, however, a significant difference between their mean ages, CA 10.58 and 9.52 respectively, though the ranges are the same for both. What effect the age differential might have on the factor patterns of the two groups is unknown. The proportions of male to female subjects are almost identical in the subsamples: .7 to .3.

INSTRUMENTS

In the author's private practice, purposely named Psychological Research and Services, a minimal test battery was required, in addition to whatever instruments the clinician needed for specific individuals. This is the multimethod approach to psychological evaluation, endorsed appropriately by Sattler (1988) among others, which the author has termed the "principle of redundant validity" (Dreger, 1993). According to this principle, regardless of what incremental validity is contributed by any particular test *for groups*, it is necessary *for individuals* and especially clinical individuals to have more than one indicator of a characteristic to make important decisions about such individuals. Accordingly, several both direct and indirect assessments of intellectual functioning were built into the battery, as well as, for children, several tests of academic achievement. Included in the battery, therefore, were the following tests:

WISC, *WISC-R*, or *WPPSI* and *Goodenough-Harris Drawing Test (GH)*. These are so well known, no need exists to describe them here.

Early School Personality Questionnaire or *Children's Personality Questionnaire* (Cattell, 1973). Whatever validity may or may not be found in these tests as a whole (e.g., Gough, 1978; Sines, 1978; Thorndike, 1978), *Factor B (ESPQ B)* which purports to measure Spearman's g or general intelligence factor does correlate significantly with the three other intelligence tests in the battery, its highest r being .53 with the Wechsler tests.

Full-Range Picture Vocabulary Test (Ammons & Ammons, 1948). This test has a four-choice picture mode comparable in part to the better-known Peabody. According to Sattler (1988), it has satisfactory reliabilities and correlates well with more extensive intelligence scales.

Wide Range Achievement Test (WRATR, WRATS, WRATA) (1965 revision) (Jastak, Jastak, & Bijou, 1965). This test in the versions used for this study has received quite negative reviews (e.g., Merwin, 1972; Thorndike, 1972), although Thorndike states that the test may have some validity in clinical and research settings with individual administration. This latter consideration, as well as the fact that the WRAT does tap important areas of educational attainment, in subtests of *Reading*, *Spelling,* and *Arithmetic*, and the brevity of the scales prompted the team of researchers/clinicians to choose the instrument for the battery. Although the reported split-half reliabilities verge on the unreal (.98 being the lowest r), more realistic coefficients using the two levels of the test for reliability purposes (.90 for Reading and Spelling and .85 for Arithmetic) give some assurance of internal consistency. Merwin (1972) questions whether the test distinguishes between intelligence and achievement; the factor pattern of the present study tends moderately to support the implied criticism.[3] Despite its shortcomings, the WRAT was included in the required test battery, though often supplemented in clinical practice by full-length achievement tests.

Children's Behavioral Classification Project Inventory (CBCP. Full descriptions of the Children's Behavioral Classification Project and its Inventory, from which *Factor II* was derived, are found in Dreger (1977, 1981, 1996) and Dreger and Dreger (1962). Briefly, the Inventory is a 274-item behavioral questionnaire to which parents or parents-surrogate respond "True" or "False" as to whether the behaviors have been observed in the past six months. The behavioral items were developed by an interdisciplinary team of clinicians and the research design and instrument were thoroughly critiqued and refined by a much broader interdisciplinary Technical Assistance Project sponsored by NIMH. The guiding principle for item creators was "What is it that the child does which prompts someone to say, 'He has a problem'?" Numerous reliability and validity studies have been carried out with the CBCP Inventory with samples from special populations, like "minimal brain dysfunction," blind, psychotic, brain damaged, EMR, sickle cell anemic, diabetes mellitus; in Spanish translation it has been standardized in Venezuela. Factor analyses with samples ranging from 341 to 1379 have been accomplished, with comparable factor composition from study to study.

Factor II (FACTII) of the Inventory is one of 30 derived by both principal factor and principal component analyses. It has 14 scored items and 22 other significantly-weighted identifying items. Scored items are usually those with the highest structure weights in variable rows; no item is scored on more than one factor, in order to avoid index correlation. Items particularly apropos to *Factor II* may be represented by such as "Makes failing grades in arithmetic, makes many mistakes with numbers, or says he does not like arithmetic," "Spells poorly," and "Reads poorly." Although undoubtedly influenced by the child's reports of school achievement or lack thereof, *Factor II*, derived from parental reports, is an assessment of scholastic achievement independent of the test data derived from the child.

PROCEDURE AND ANALYSES

Principal factor analysis using the SAS FACTOR procedure was employed with the 50-participant raw score matrix consisting of scores from the eight tests. Varimax and Promax rotations were applied to the two-factor solution dictated by Scree and eigenvalues greater than one. (A third latent root of slightly over 1.00 was tentatively ignored.) The rotated factor structure was examined first. Weights for the several intelligence tests tended to be higher on the first factor; and for the several achievement tests, including *Factor II*, they were higher on the second factor, but substantial loadings were found on the alternate factor in each case. However, both the reference vector structure and factor pattern clearly distinguished between the two dimensions: significant weights on the first factor for intelligence tests and on the second factor for achievement tests, with insignificant weights on the alternate factor, except for *WRATA* (Arithmetic subtest) which had a statistically significant but practically negligible weight on the first factor. Consequently, the factor pattern was employed for prediction in the confirmatory factor analysis. This choice was dictated, however, not only by the clarity of pattern but also by the fact that structural equation modeling parallels the factor pattern consisting of standardized regression weights.

Based on the two-factor pattern,[4] then, equations were written for carrying out confirmatory analysis on the data of the 46-participant sample's data, by means of LISREL (Joreskog & Sorbom, 1989) and EQS (Bentler, 1989), both of which default to the use of a covariance matrix.

The measurement model was that described in LISREL,

$$x = \Lambda_x \xi + \delta$$

where x is a vector of test scores, usually standardized, Λ_x is a matrix of factor weights, ξ is a vector of latent factors, and δ is a vector of unique elements consisting of specific and error variance. As explained in the Footnotes, LISREL could not complete the analyses.[5] Accordingly, analyses proceeded with EQS. EQS does give appropriate warnings, for example, that a matrix may not be positive definite, but ordinarily continues the analyses after suggesting that the results may not be trustworthy. Since EQS provides start values for parameters to be estimated, the user-supplied estimates for the lambda matrix were not included in the equations. Values for the theta delta diagonal were set to .7 and for the phi

TABLE 1
TESTS OF TWO HYPOTHESIZED LATENT VARIABLES AS ASSESSED BY EIGHT MANIFEST VARIABLES

Statistic or Test	1st Sample (n = 50)	2nd Sample (n = 46)
Determinant of covariance matrix	.116924D+15	.92560D+14
Residuals (S - Sigma)		
Average, absolute standardized	.0357	.0585
Average, off-diagonal absolute standardized	.0459	.0753
Range of distribution	5 - 8	5 - 8
Goodness of fit		
Chi-square (17 df)	18.819	28.772
p	.34	.04
Normal theory reweighted LS chi-square	18.095	23.438
Satorra-Bentley scaled chi-square	16.837	25.256
p	.47	.09
Normed Fit Index	.915	.785
Non-normed Fit Index	.985	.816
Comparative Fit Index	.991	.889

5 = -.2 to -.1; 6 = -.1 to .0; 7 = .0 to .1; 8 = .1 to .2.

matrix at .6. The maximum likelihood method with the ROBUST backup for possible non-normal distributions was ordered.

For making direct comparisons of factor patterns from one set of data to the other, both of the data sets were subjected to EQS, even though the exploratory analysis had originally been done with SAS. It would have been desirable to utilize LISREL, since it has some advantages over EQS, especially in providing squared multiple correlations, a plot of standardized residuals, and modification indices for changing parameter estimates (as well as the lambda matrix in its familiar factor form!). Nevertheless, EQS, too, has its advantages, including especially more easily allowing equality constraints. Thus, for the test of equality of factor patterns in the two data sets, equality constraints were imposed, and tested by the Lagrange multiplier test, corresponding to the hypothesized model of two factors measured by four variables apiece, i.e., WISC, GH, ESPQ B, and FRPVT for the first factor and FACTII, WRATR, WRATS, and WRATA for the second factor.

In addition to the structural equation modeling, more traditional methods of matrix comparison were employed: the Ahmavaara transformation (Ahmavaara, 1954), the congruence coefficient (Burt, 1948), and the Salient Variable Similarity Index (Cattell, 1949).

TABLE 2
FACTOR PATTERNS OF INDEPENDENT DATA SETS FROM EQS ANALYSES

Variable	1st Sample			2nd Sample		
	Fact 1	Fact 2	Uniqueness	Fact 1	Fact 2	Uniqueness
Factor II	.000	-.541	.841	.000	-.481	.876
WISC, WISC-R, WPPSI	.957	.000	.290	.826	.000	.564
GH	.673	.000	.740	.727	.000	.687
ESPQ B	.559	.000	.829	.581	.000	.814
FRPVT	.802	.000	.598	.473	.000	.881
WRATR	.000	.911	.411	.000	.811	.584
WRATS	.000	.915	.402	.000	.781	.624
WRATA	.000	.645	.764	.000	.618	.787

RESULTS

Table 1 reveals the results of testing the hypothesis of a two-factor solution applied to the second data set. (The comparable EQS analysis of the first set is included in Table 1 for a direct comparison.) Both the average of absolute standardized residuals as a whole and of the same for off-diagonal elements are satisfactorily small, indicating a close fit between obtained and model covariance matrices. These figures fairly represent the entire distributions of residuals, in that the largest absolute standardized residual is .180 (compared in the first sample to .164). Likewise, the non-significant chi-squares suggest that there is a fairly close approximation between theory and empirical reality. On the other hand, the three fit indices (above and beyond the chi-squares which in this usage constitute goodness-of-fit indices) give reason to conclude that the model fits less well than it ought to. None of the coefficients, though, indicates a truly "bad" fit. Table 1 does not include figures on the univariate and multivariate skewnesses and kurtoses, none of which, however, obviate the results given in the table.

Table 2 presents the results of the modeling in terms of factor matrices in more familiar form; these represent the standardized measurement equations. As postulated, the zero entries imply negligible elements; in actuality, in exploratory factor analyses these elements may achieve statistical significance and can be used in regression equations to predict scores more accurately on the original variables. But from a structural standpoint (and in Cattell's, 1978, and Gorsuch's, 1983, all-or-nothing weighting), these elements constitute "noise." Even at this point, though, before formally testing comparability of patterns, it is evident that they are reasonably similar. It must be emphasized that these similarities did not arise from inserting initial start values in the EQS equations, for no such values were given; they came entirely from the data themselves.

TABLE 3
EQUALITY OF FIRST AND SECOND FACTOR PATTERNS DETERMINED BY JOINT MODELLING AND COMPARISON COEFFICIENTS

Statistic or Test			1st Sample (n = 50)	2nd Sample (n = 46)
Average, absolute standardized residual			.0553	.0962
Average, off-diagonal absolute standardized residual)			.0596	.1060
Constraints, to test equality hypothesis			**Chi-square**	**Probability**
1st Sample		2nd Sample		
FACTOR II, F2	-	FACTOR II, F2 = 0	.016	.900
WISC, F1	-	WISC, F1 = 0	.532	.466
GH, F1	-	GH, F1 = 0	1.049	.306
ESPQ B, F1	-	ESPQ B, F1 = 0	1.505	.220
FRPVT, F1	-	FRVPT, F1 = 0	1.226	.268
WRATR, F2	-	WRATR, F2 = 0	.332	.565
WRATS, F2	-	WRATS, F2 = 0	.731	.392
WRATA, F2	-	WRATA, F2 = 0	.242	.623
Comparison Method			**Coefficient**	
Congruence Coefficient			.9885	
Ahmavaara Transformation			1.0000	
Salient Variable Similarity Index			.6957	

Table 3 brings together the tests of equality of factor patterns derived from the separate and wholly independent data sets. Again, it is evident that the model fits the first set better than it does the second, as seen in the average absolute standardized residuals in the upper part of the table. In the second set, although values clustering about 0.1 are not large by the usual criteria, especially in consideration of the fact that the largest absolute standardized residual in either matrix is only .238, a closer fit in the second data would make for a closer parallel between the two sets. In the middle portion of the table, the Lagrange method tests whether the actual values obtained for the relation between the manifest and latent variables are the same for both data sets. From the small chi-squares and associated p values, it can be seen that the differences are statistically non-significant. At the bottom of Table 3, the more traditional comparison coefficients provide evidence that the two pattern matrices are essentially quite similar.

DISCUSSION

By several different tests the comparability of the two sets of factor patterns, each derived from an independent sample of clinical children, has been demonstrated. Thus, the purpose of this study, to confirm the factor pattern of the two-factor solution by revealing it in independent samples, appears to have been accomplished.

There are several additional comments called for. In Table 2, it might seem as if no factor weights would be found on the alternate factor when the variable is more strongly associated with one factor. No doubt there could be some "significant" weights found on alternate factors if the predictive equations included elements for the alternate. The fact, however, that the exploratory factor analysis so clearly delineated a pattern in which no significant weights were found on the alternate factor virtually required the equations to be set up with only one factor apiece (plus an error term in each instance).

It is both interesting and important to note that both the Lagrange test and the traditional comparison methods (Table 3) yield the same information; examination of Table 2 provides essentially that same information in different form.

Finally, attention should be paid to the Footnote in which the significance of this research is stressed: Revealing the underlying "causes" of the battery of well-standardized tests appears to be an important accomplishment.

REFERENCES

Ahmavaara, Y. (1954). Transformation analysis of factorial data and other new analytical methods of differential psychology with their application to Thurstone's basic studies. *Annales Academiae Scientiarum Fennicae, Sarja-Ser. B nide-tom, 88*, 2.

Ammons, R. B., & Ammons, H. S. (1948). *Full-Range Picture Vocabulary Test.* Missoula, Montana: Psychological Test Specialists.

Bentler, P. M. (1989). *EQS Structural equations program manual.* Los Angeles: BMDP Statistical Software.

Bentler, P. M. (1993). *EQS Structural equations program manual.* Los Angeles: BMDP Statistical Software.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Burt, C. (1948). The factorial study of temperamental traits. *British Journal of Psychology, Statistical Section, 1*, 178-203.

Cattell, R. B. (1949). A note on factor invariance and the identification of factors. *British Journal of Psychology, 2*, 134-138.

Cattell, R. B. (1973). *Personality and Mood by Questionnaire*. San Francisco: Jossey-Bass.

Cattell, R. B. (1978). *The scientific use of factor analysis in behavioral and life sciences*. New York: Plenum.

Dreger, R. M. (1953). Different IQ's for the same individual associated with different intelligence tests. *Science, 118*, 594-595.

Dreger, R. M. (1981). The classification of children and their emotional problems. *Clinical Psychology Review, 1*, 415-430.

Dreger, R. M. (1991). The latent structure of the Salvation Opinionnaire, a measure of religious attitudes in the American Christian tradition. *Educational and Psychological Measurement, 51*, 707-719.

Dreger, R. M. (1993). Methodology. In J. L. Matson (Ed.), *Handbook of hyperactivity in children*. Boston: Allyn & Bacon.

Dreger, R. M. (1996). Background for and development of a brief Children's Behavioral Classification Project Inventory. *Psychological Reports, 79*, 963-977.

Dreger, R. M., & Dreger, G. E. (1962). *Proceedings of the Technical Assistance Project*. Behavioral Classification Project Report No. 1. Jacksonville, Florida: Behavioral Classification Project, Jacksonville University.

Gough, H. F. (1978). Review of *Children's Personality Questionnaire, 1975 Edition*. In O. K. Buros (Ed.), *Eighth mental measurements yearbook*. Lincoln, Nebraska: University of Nebraska Press.

Guilford, J. P. (1964). Zero intercorrelations among tests of intellectual abilities. *Psychological Bulletin, 61*, 401-404.

Hayduk, L. A. (1987). *Structural equation modelling with LISREL: Essentials and Advances*. Baltimore and London: Johns Hopkins University Press.

Jastak, J., Jastak, S. R., & Bijou, J. W. (1965). *Wide-Range Achievement Test*. (Rev. ed.) New York: Psychological Corporation.

Joreskog, K. G., & Sorbom, D. (1989). *LISREL 7 user's reference guide*. Mooresville, Indiana: Scientific Software.

Merwin, J. C. (1972). Review of *Wide Range Achievement Test*. In O. K. Buros (Ed.), *Seventh mental measurements yearbook*. Lincoln, Nebraska: University of Nebraska Press.

Messick, S. (1988). Validity. In R. L. Linn (Ed.), *Educational Measurement*. 3rd ed. New York: Macmillan.

Sattler, J. M. (1988). *Assessment of children*. 3rd ed. San Diego: Jerome M. Sattler, Publisher.

Sines, J. O. (1978). Review of *Early School Personality Questionnaire*. In O. K. Buros (Ed.), *Eighth mental measurements yearbook*. Lincoln, Nebraska: University of Nebraska Press.

Thorndike, R. L. (1972). Review of *Wide Range Achievement Test*. In O. K. Buros (Ed.), *Eighth mental measurements yearbook*. Lincoln, Nebraska: University of Nebraska Press.

Thorndike, R. L. (1978). Review of *Early School Personality Questionnaire*. In O. K. Buros (Ed.), *Eighth mental measurements yearbook*. Lincoln, Nebraska: University of Nebraska Press.

Wothke, W. (1992). Nonpositive definite matrices in structural modeling. In K. A. Bollen and S. Long (Eds.), *Testing structural models*. Beverly Hills, California: Sage.

Footnotes

1. Guilford (1964) reported that he had found over 7,000 correlation coefficients in the literature between and among intelligence tests, almost one-fourth of which were not significantly different from zero.
2. Forty-six subjects is lower than Bollen (1989) advocates as a minimum, i.e., 50. He states that he knows no hard and fast rule, but suggests several cases per each free parameter. In the present instance, there are 20 free parameters to be estimated. Consequently, it must be recognized that the chi- squares generated may be too large. However, Hayduk (1987) reports that he has run factor models with as few as 22 cases without discernible problems. Since the solutions in the present instance converged well within the EQS' specified limit of 30 iterations and the goodness-of-fit indexes are fairly satisfactory, the presumption is that the case is similar to Hayduk's.
3. Only for the Arithmetic subtest, however, which distributes its variance between the two factors in the factor structure with a lesser but still significant weight on the intelligence factor.
4. A three-factor model was also tested, utlizing EQS. Although the Bentler fit indices were larger than those for the two factors, the other indices of fit were far inferior to those seen in Table 1. E.g., the average standardized absolute residuals, both including and excluding diagonals, are .3330 and .3033 respectively, as compared to the very small average residuals for two factors. Eigenvalues for the first data set ($\underline{n}$ = 50) are 3.70, 1.37, 1.02, 0.64, 0.56, 0.35, 0.25, and 0.12; for the second data set ($\underline{n}$ = 46), they are 3.44, 1.46, 0.93, 0.79, 0.58, 0.34, 0.26, and 0.21.
5. That LISREL could not produce results stems in part from its sensitivity to non-positive definite matrices and underidentification of parameters. Both Joreskog and Sorbom (1989) and Wothke (1992) suggest ways to overcome lack of positive definiteness in LISREL. Several of these ways were attempted but without success. On the other hand, though EQS does give warnings about lack of identification and non-positiveness, it ordinarily continues the analyses after suggesting that the results may not be trustworthy.