

THE UTILITY OF CANONICAL CORRELATION ANALYSIS, COUPLED WITH TARGET ROTATION, IN COPING WITH THE EFFECTS OF DIFFERENTIAL SKEWNESS OF VARIABLES

JOHANN M SCHEPERS

abo@rau.ac.za

Department of Human Resource Management
 University of Johannesburg

ABSTRACT

The principal objective of the study was to determine the utility of canonical correlation analysis, coupled with target rotation, in coping with the effects of differential skewness of variables representing two batteries of tests. Generally speaking joint factor analyses of two or more batteries of tests result in factors of skewness rather than factors of content. To examine the problem, the General Scholastic Aptitude Test (GSAT) and Senior Ability Tests (SAT) were jointly applied to a sample of 1598 first-year university students, and subjected to both a principal factor analysis (PFA) and a canonical correlation analysis (CCA), coupled with target rotation. Three factors were obtained in both instances. The PFA yielded factors of skewness and the CCA factors of content. The target rotation gave a good fit with the theoretically specified values. The implications of the findings are discussed.

Key words

Canonical correlation, Target rotation, Tarrot rotation

An issue which often arises, for example in studies of the Big Five, is whether two or more test batteries, given to the same sample of participants, have a common factor structure. Traditionally, researchers have simply conducted a joint factor analysis of such batteries of tests. However, due to the effects of differential skewness of the variables involved, the resulting factor structures have often been distorted. Factors of skewness rather than content have usually been obtained. Finch and West (1997, p.470) pointed out in this regard that joint factor analyses confound two sources of covariation, namely covariation within batteries and covariation between batteries.

To overcome the confounding of the two sources of covariation mentioned, Tucker (1958) proposed his interbattery factor analysis. His model will now be briefly described.

Assume that two batteries of tests, with a postulated common factor structure, have been applied to a representative sample of participants. The variables were intercorrelated and yielded the super matrix depicted in Figure 1.

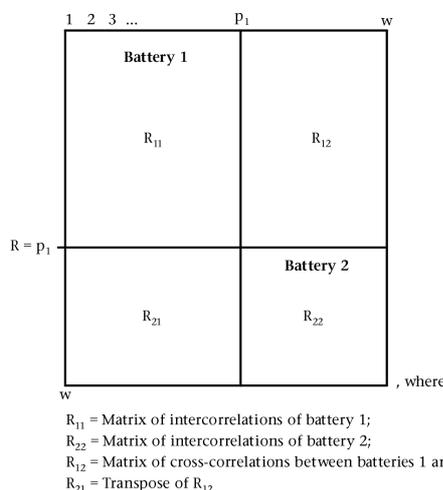


Figure 1: Super matrix (R)

According to the fundamental theorem of factor analysis the super matrix R can be resolved into its factors as follows:

$R = FF'$, where

$$F = \begin{bmatrix} A_1 & S_1 & O \\ A_2 & O & S_2 \end{bmatrix}$$

A_1 = Factors of battery 1 shared in common with battery 2;

A_2 = Factors of battery 2 shared in common with battery 1;

S_1 = Factors specific to battery 1;

S_2 = Factors specific to battery 2.

R can therefore be presented as follows:

$$R = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} = \begin{bmatrix} A_1 & S_1 & O \\ A_2 & O & S_2 \end{bmatrix} \cdot \begin{bmatrix} A_1' & A_2' \\ S_1' & O \\ O & S_2' \end{bmatrix}$$

$F \qquad F'$

It is therefore clear that $R_{12} = A_1A_2'$.

A_1A_2' contains the factors common to batteries 1 and 2.

Browne (1979) provided a maximum likelihood solution to Tucker's model of interbattery factor analysis. He obtained estimates of the interbattery factor loadings by scaling the correlations of the original variables with the canonical variates. He subsequently extended his technique to more than two batteries of tests (Browne, 1980).

More recently Schepers (2004b) showed that the Multiple Battery Factor Analysis (MBFA) technique of Browne (1980) can cope with the effects of differential skewness of variables from two different batteries of tests.

He applied the General Scholastic Aptitude Test (GSAT) and Senior Ability Tests (SAT) jointly to a sample of 1598 first-year university students, and subjected the inter-correlation matrix to a principal factor analysis. The obtained

factor matrix was rotated to simple structure by means of a Direct Oblimin rotation.

The principal factor analysis yielded three factors, viz. a non-verbal (spatial) factor, and two verbal factors. The verbal tests of the GSAT loaded on one factor and the verbal tests of the SAT on another.

Following this the intercorrelation matrix was subjected to a multiple battery factor analysis (MBFA) and rotated to simple structure by means of a Direct Quartimin rotation. Again a three-factor-structure was obtained. A Tucker-Lewis reliability coefficient of 0,967 was obtained, which is highly acceptable. The average absolute off-diagonal residual was 0,046 which indicates a very good fit.

Three clear-cut factors were obtained, which were identified as a non-verbal reasoning factor, a verbal factor, and a number factor.

The three factors were strongly positively correlated, suggesting an underlying factor of general intelligence.

From the coefficients of skewness of the various measures of the GSAT and SAT it would appear that the distributions of the GSAT are quite skew. The indices range from 1,818 to -2,111. By contrast the distributions of the SAT are moderately skew. The indices range from 0,450 to -1,248.

From the foregoing it should be clear that even moderate variations in degrees of skewness can distort the factor structure of two batteries of tests if a joint factor analysis is done. By contrast MBFA seems to cope quite well with moderate degrees of skewness.

According to Browne (1979, p.75) the interbattery factor analysis model is "a genuine factor analysis model in that a single set of unobservable factor variables accounts for all correlation coefficients between two batteries of tests". By contrast canonical correlation analysis "is strictly a method of component analysis since two sets of observable linear combinations of variables are employed to investigate relationships between the two batteries of tests" (p.75).

Despite the fact that the rationale of the two models are quite different, the numerical procedures of canonical correlation analysis are very similar to that involved in obtaining maximum-likelihood estimates of interbattery factor loadings (Browne, 1979, p.75). It would therefore be very interesting to examine the utility of canonical correlation analysis in coping with the effects of differential skewness of variables.

The objective of canonical correlation analysis is to form linear combinations of two sets of continuous variables so as to maximise the correlation between the two composites (Cliff, 1987, p.453). According to Cliff (1987, p.455) canonical correlation analysis can be used if "one set of variables is dependent and the other independent or when there is no distinction in the roles of the two sets". It can therefore also be applied to variables from two batteries of tests.

A statistical test is performed to determine how many significant components there are (Bartlett, 1950; 1951). Each component (dimension) is represented by two vectors of weights – one in respect of the first battery of tests, and the other in respect of the second battery of tests. The two vectors of weights representing a component are normally referred to as a variate, and the correlation between the two composites of a variate yields the canonical correlation in respect of that component. Thus there are as many canonical correlations as there are statistically significant components.

From an interpretive point of view it is normally very difficult to identify the components underlying the canonical structure

matrix as it resembles an unrotated factor matrix. Rotation to simple structure is therefore necessary. In this regard Cliff (1987, p.456) states that the "structure correlations" between the observed variables and the canonical variates "can be transformed by the rotational methods of factor analysis, although the same transformation must be applied to the structure correlations of both batteries". Target rotation would seem to be ideal for this purpose.

From a theory testing point of view target rotation is more appropriate than the usual rotations to simple structure such as Varimax, Promax, Direct Oblimin, Quartimax, Quartimin, and other procedures. With target rotation the common factor structure of two batteries of tests can be specified on theoretical grounds. This is particularly useful whenever theoretical models are being tested.

From the foregoing it should be clear that differential skewness of variables is very disruptive when doing joint factor analyses of two or more batteries of tests (Ferguson, 1941; Gorsuch, 1974; Schepers, 2004a and 2004b; Finch & West, 1997). There is thus a real need for techniques that can cope with the effects of differential skewness of variables of a continuous nature.

Objectives of the study

The principal objective of the study was to determine the utility of canonical correlation analysis, coupled with target rotation, in coping with the effects of differential skewness of variables from two batteries of tests.

RESEARCH DESIGN

Research approach

The primary goal of the study was to evaluate a particular statistical technique. A cross-sectional field survey was used in the collection of the data.

Participants

As the sample has been fully described in a previous study (cf. Schepers, 2004b, pp.78-79) only the essential details are given here:

A representative sample of first-year university students at the Rand Afrikaans University, during 1995, was used in the study. Complete records in respect of 1598 participants were available in respect of the General Scholastic Aptitude Test (GSAT) and Senior Aptitude Tests (SAT), amongst others.

Measuring instruments

As a complete description of the measuring instruments have been given in a previous study (cf. Schepers, 2004b, p.79) only the essential details are given here:

The General Scholastic Aptitude Test (GSAT)

The GSAT yields a measure of academic intelligence or scholastic aptitude. It consists of six subtests – three verbal and three non-verbal, and measures both verbal and non-verbal intelligence (Claassen, De Beer, Hugo & Meyer, 1998).

The Senior Aptitude Tests (SAT)

The SAT was designed for the measurement of a number of aptitudes of pupils in Grades 10, 11 and 12, and of adults. It consists of verbal, numerical, non-verbal reasoning, spatial and memory tests. Coordination and Writing Speed were excluded for the purposes of the present study (Fouché & Verwey, 1991).

Procedure

For the purposes of the present study only the records of students who had completed both the GSAT and the SAT were used. A total of 1598 complete records were obtained.

Statistical analysis

In order to attain the stated objective a Canonical Correlation Analysis (CCA) was done (Cliff, 1987; Tabachnick & Fidell, 1983). The obtained canonical structure matrix was rotated to simple structure by means of a target rotation (Browne, 1972a, 1972b, 1993).

RESULTS

Principal objective: To determine the utility of canonical correlation analysis, coupled with target rotation, in coping with the effects of differential skewness of variables from two batteries of tests

As a first step in the analysis, the canonical correlations of the subtests of the GSAT with the various measures of the SAT were computed. Bartlett's (1950, 1951) test of significance was used to determine the number of significant canonical correlations, and is given in Table 1.

**TABLE 1
STATISTICAL SIGNIFICANCE OF CANONICAL CORRELATIONS:
BARTLETT'S TEST IN RESPECT OF THE GSAT AND SAT**

Eigenvalues	Canonical correlations	Eigenavlue removed	Significance of eigenvalues remaining			
			χ^2	df	p	Lambda prime
0,548444	0,740570	0	1712,319	60	<0,000001	0,340294
0,154387	0,392921	1	449,373	45	<0,000001	0,753602
0,075667	0,275076	2	182,992	32	<0,000001	0,891190
0,024308	0,155911	3	58,004	21	0,00003	0,964144

Note: N = 1598

From Table 1 it is clear that there are at least three significant canonical correlations. Accordingly three canonical variates, together with their associated canonical correlations, were computed. The complete analysis is given in Table 2.

Table 2 shows that the first canonical variate yielded a canonical correlation of 0,741, the second a canonical correlation of 0,393 and the third a canonical correlation of 0,275. The first canonical variate suggests a general factor, with loadings ranging from 0,421 to 0,869. The second and third variates, however, are more difficult to interpret as no simple structure is visible. It was therefore decided to rotate the matrix of canonical variates to simple structure. For this purpose use was made of a target matrix in conjunction with a Tarrot rotation. The target matrix was specified on theoretical grounds after studying the subtests of the GSAT and SAT. The target matrix is given in Table 3.

The target matrix was specified with high loadings on Factor 1 in respect of the non-verbal reasoning tests. Factor 2 was specified with high loadings on all the verbal tests, together with the two memory tests, and Factor 3 was specified with high loadings on the numerical tests.

Accordingly an oblique Tarrot rotation was performed of the matrix of canonical variates. The rotated matrix is given in Table 4.

Table 4 shows that rotation of the canonical variates to simple structure resulted in a well defined structure, yielding a good fit with the theoretically specified target matrix. The square root of the average squared deviation was equal to 0,144480.

**TABLE 2
CANONICAL CORRELATIONS OF GSAT AND THE
RESPECTIVE MEASURES OF THE SAT**

Correlations of original measures with canonical variates				
	Variate 1	Variate 2	Variate 3	
Battery 1				
GSAT 1: WORD ANALOGIES	0,653	0,579	0,139	
GSAT 2: NUMBER SERIES	0,860	-0,049	0,417	
GSAT 3: VERBAL REASONING	0,869	0,216	0,075	
GSAT 4: PATTERN COMPLETION	0,784	-0,189	-0,405	
GSAT 5: WORD PAIRS	0,703	0,517	-0,165	
GSAT 6: FIGURE ANALOGIES	0,823	-0,241	-0,274	
Average % variance accounted for	61,79 %	12,42%	7,76%	Total: 81,97%
Average % redundancy	33,89%	1,92%	0,59%	Total: 36,40%
Battery 2				
SAT 1: VERBAL COMPREHENSION	0,772	0,402	0,103	
SAT 2: CALCULATIONS	0,629	0,180	0,719	
SAT 3: DISGUISED WORDS	0,499	-0,660	0,161	
SAT 4: COMPARISON	0,421	-0,125	0,259	
SAT 5: PATTERN COMPLETION	0,782	-0,155	-0,155	
SAT 6: FIGURE SERIES	0,712	-0,132	-0,016	
SAT 7: SPATIAL 2D	0,705	-0,203	-0,161	
SAT 8: SPATIAL 3D	0,762	-0,255	-0,301	
SAT 9: MEMORY (PARAGRAPH)	0,479	0,386	-0,063	
SAT 10: MEMORY (SYMBOLS)	0,511	0,236	-0,140	
Average % variance accounted for	40,29%	9,97%	7,85%	Total: 58,11%
Average % redundancy	22,09%	1,54%	0,59%	
Canonical correlations	0,741	0,393	0,275	Total: 24,22%

Note: N = 1598

**TABLE 3
TARGET MATRIX SPECIFIED FOR TARROT ROTATION**

Variable	Factor 1	Factor 2	Factor 3
GSAT 1 WORD ANALOGIES	0,000	9,000	0,000
GSAT 2 NUMBER SERIES	0,000	0,000	9,000
GSAT 3 VERBAL REASONING	0,000	9,000	0,000
GSAT 4 PATTERN COMPLETION	9,000	0,000	0,000
GSAT 5 WORD PAIRS	0,000	9,000	0,000
GSAT 6 FIGURE ANALOGIES	9,000	0,000	0,000
SAT 1 VERBAL COMPREHENSION	0,000	9,000	0,000
SAT 2 CALCULATIONS	0,000	0,000	9,000
SAT 3 DISGUISED WORDS	0,000	9,000	0,000
SAT 4 COMPARISON	0,000	0,000	9,000
SAT 5 PATTERN COMPLETION	9,000	0,000	0,000
SAT 6 FIGURE SERIES	9,000	0,000	0,000
SAT 7 SPATIAL 2D	9,000	0,000	0,000
SAT 8 SPATIAL 3D	9,000	0,000	0,000
SAT 9 MEMORY (PARAGRAPH)	0,000	9,000	0,000
SAT 10 MEMORY (SYMBOLS)	0,000	9,000	0,000

TABLE 4
TARROT ROTATION OF CANONICAL FACTOR LOADINGS
(GSAT & SAT)

Variable	Factor 1	Factor 2	Factor 3
BATTERY 1			
GSAT 1: WORD ANALOGIES	0,013	0,937	-0,159
GSAT 2: NUMBER SERIES	0,196	0,213	0,717
GSAT 3: VERBAL REASONING	0,260	0,558	0,260
GSAT 4: PATTERN COMPLETION	0,923	0,063	-0,142
GSAT 5: WORD PAIRS	0,113	0,880	-0,151
GSAT 6: FIGURE ANALOGIES	0,883	0,006	0,028
BATTERY 2			
SAT 1 VERBAL COMPREHENSION	0,003	0,732	0,183
SAT 2 CALCULATIONS	-0,070	-0,043	1,020
SAT 3 DISGUISED WORDS	-0,380	0,965	0,092
SAT 4 COMPARISON	0,134	-0,022	0,445
SAT 5 PATTEN COMPLETION	0,695	0,090	0,115
SAT 6 FIGURE SERIES	0,525	0,086	0,236
SAT 7 SPATIAL 2D	0,695	0,005	0,103
SAT 8 SPATIAL 3D	0,881	-0,030	-0,014
SAT 9 MEMORY (PARAGRAPH)	0,010	0,638	-0,061
SAT 10 MEMORY (SYMBOLS)	0,211	0,471	-0,083

Note: Square root of average deviation = 0,144480

FACTOR CORRELATION MATRIX

	Factor 1	Factor 2	Factor 3
FACTOR 1	1,000	0,575	0,464
FACTOR 2	0,575	1,000	0,453
FACTOR 3	0,464	0,453	1,000

DISCUSSION

The principal objective of the study turned out positive: Rotation of the canonical variates by means of a target rotation yielded a structure that is very similar to that obtained with the MBFA. A CCA followed by a target rotation might even be preferable to a MBFA when doing confirmatory studies as the target matrix can be specified on theoretical grounds prior to initiating the study. Target rotation can of course also be used with MBFA, but then the current program would have to be adapted.

ACKNOWLEDGEMENT

I hereby wish to thank Riëtte Eiselen and Wilhelm Koster of the Statistical Consultation Service of the University of Johannesburg for all the hours of computational work done for me. I value it very highly.

A special word of thanks to Annetjie Boshoff and her assistant Afton Walters for typing the manuscript at short notice.

REFERENCES

- Bartlett, M.S. (1950). Tests of significance in factor analysis. *British Journal of Psychology*, 3, 77-85.
- Bartlett, M.S. (1951). A further note on tests of significance in factor analysis. *British Journal of Psychology, Statistical Section*, 4, 1-2.
- Browne, M.W. (1972a). Oblique rotation to a specified target. *British Journal of Mathematical and Statistical Psychology*, 25, 207-212.
- Browne, M.W. (1972b). Orthogonal rotation to a partially specified target. *British Journal of Mathematical and Statistical Psychology*, 25, 115-120.
- Browne, M.W. (1979). The maximum likelihood solution in interbattery factor analysis. *British Journal of Mathematical and Statistical Psychology*, 32, 75-86.
- Browne, M.W. (1980). Factor analysis of multiple batteries by maximum likelihood. *British Journal of Mathematical and Statistical Psychology*, 33, 184-199.
- Browne, M.W. (1993). *Rotation to a partially specified target: Tarrot version 2*, Department of Psychology, University of Illinois, Columbus.
- Claassen, N.C.W., De Beer, M., Hugo, H.L.E. & Meyer, H.M. (1998). *Manual for the General Scholastic Aptitude Test*. Pretoria: Human Sciences Research Council.
- Cliff, N. (1987). *Analyzing multivariate data*. New York: Harcourt Brace Jovanovich, Publishers.
- Ferguson, G.A. (1941). The factorial interpretation of test difficulty. *Psychometrika*, 6, 323-329.
- Finch, J.F. & West, S.G. (1997). The investigation of personality structure: Statistical models. *Journal of Research in Personality*, 31, 439-485.
- Fouche, F.A. & Verwey, F.A. (1991). *Manual for the Senior Aptitude Tests*. Pretoria: Human Sciences Research Council.
- Gorsuch, R.L. (1974). *Factor analysis*. Philadelphia: Saunders.
- Kaiser, H.F. (1961). A note on Guttman's lower bound for the number of common factors. *British Journal of Statistical Psychology*, 14 (1), 1.
- Schepers, J.M. (2004a). Overcoming the effects of differential skewness of test items in scale construction. *SA Journal of Industrial Psychology*, 30 (4), 27-43.
- Schepers, J.M. (2004b). The power of multiple battery factor analysis in coping with the effects of differential skewness of variables. *SA Journal of Industrial Psychology*, 30 (4), 78-81.
- Tabachnick, B.G. & Fidell, L.S. (1983). *Using multivariate statistics*. New York: Harper & Row.
- Tucker, L.R. (1958). An inter-battery method of factor analysis. *Psychometrika*, 23, 111-136.