In ~~J. Vázquez et al.~~

Metodoloxía da Investigación científica.

Santiago de Compostela, Spain: Universidade de S.d.C.

(pp. 105 - 126.

PIETER M. KROONENBERG
Leiden University

THREE-MODE ANALYSIS BY EXAMPLE

Introduction

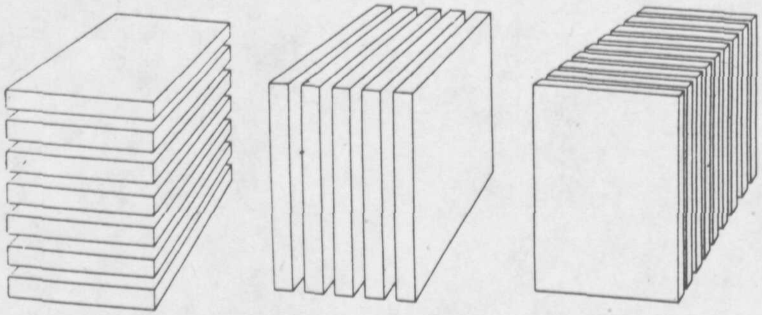
In this paper three-mode analysis, in particular, three-mode principal component analysis and, to a lesser extent, parallel factor analysis will be presented. The level of explanation will be exclusively on a conceptual level, and fomulas will be entirely avoided. The example is based on the data from a psychophysiological experiment. Twin pairs were given an acute dose of alcohol and several measures were taken before and three times after the drinking. Many other domains of enquiry also yield data which have been fruitfully handled by three-way techniques. For instance, the plant breeders' problem of evaluating genotypes of soyabeans in different locations on various attributes for further selection has been examined with three-mode techniques, as well as, intelligence scores from normal and retarded children. In the latter case, only the correlation matrices were available, but not the original scores. Thus both cross-sectional data bases and repeated measures data can be analysed fruitfully with three-way methods.

THEORY

Data

Two-way data. Generally, most researchers only come across two-way data, or at least they think they do. In this respect, there are two major kinds of two-way data we will consider: profile data and similarity data. Examples of two-way *profile data* are (1) test scores of subjects (first way) on arithmetic items (second way), and (2) measurements of water quality (second way) at several stations along the river Tambre (first way). An example of two-way *similarity data* is an array with similarities between pairs of Riojas Crianzas jud-

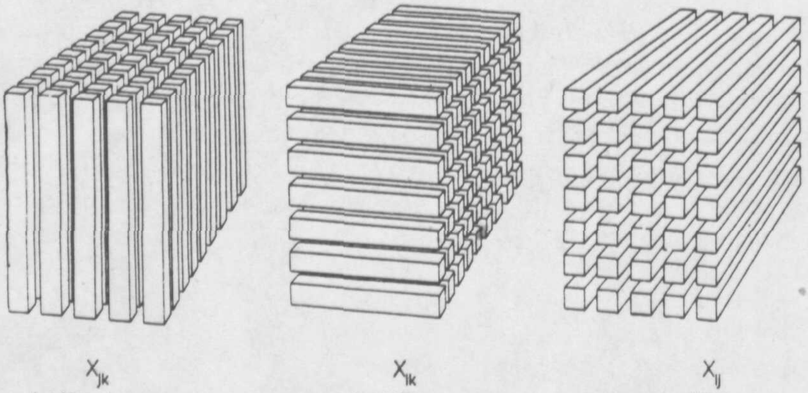
DISPLAY 1A



X_i I Horizontal Slices — X_j J Lateral Slices — X_k K Frontal Slices

Slices, the two-way submatrices of the three-way matrix X

DISPLAY 1B



X_{jk} J x K columns — X_{ik} I x K Rows — X_{ij} I x J Tubes

Fibers, the one-way submatrices of the three-way matrix X.

ged by one Dutchman with reespect to their attractiveness for her next diner party (stimuli -Riojas- constitute both first and second ways). Many similar examples exist in all fields of scientific endeavour.

Three-way data. However, more complex designs arise easily, for instance, when the subjects are measured under different conditions or points in time. For instance, when the water measurements are taken in different months, or when the similarity judgements between Riojas are made by several Dutchmen.

There are several ways in which such a data block or data box can be sliced or subdivided into submatrices. All possibilities are presented in Display 1. There are no particular rules in assigning types of variables to different ways of the data box, but in conformity with two-way *profile data* the data are generally arranged so that the first way pertains to subjects, the second way to variables and the third to conditions (points in time). For *similarity data* the first two ways are chosen to be the stimuli and the third the judges , and similarly when dealing *cross-sectional data* in the form of correlation or covariance matrices the first two ways are generally chosen to be variables and the third the different samples for which the variables were collected.

Examples. To make the situation a bit more concrete we will shortly look at research aimed at the improvement of inspection methods for surface roughness of metals; modelled after a paper by Inukai, Saito, and Mishima (1980) of the Industrial Products Research Institute, Tsukuba, Japan. Their basic material consisted of 15 metal 'loaves' with different combinations of types of roughness: (1) distance between irregularities, (2) height of irregularities, (3) waviness of the patterns, and (4) shape of the deviations. First, two hundred inspectors (subdivided into four different classes of experience) had to evaluate the comparative roughness of two loaves by feeling and looking at the loaves and score their difference on a five-point scale. Thus the scores are *dissimilarity* measures for each pair of loaves. Secondly, the same inspectors judged the fifteen loaves on the four measures indicated above, producing a set of *profile data.*

In a other curious example from Japan reported by Prof. Dai of the Tokyo Metropolitan University, several persons (first way) had to judge a number of chairs (second way) on twenty rating scales (third way) dealing with various aspects of their industrial design.

Models

Treating three-way data with two-way models. Traditionally, before the advent of three-way models and computer programmes to solve them, people generally either analysed each two-way matrix separately and compared the solutions, or the three-way matrix was strung-out in one or two of the three possible directions, and was analysed with standard two-way techniques, sometimes using the third way for interpretation.

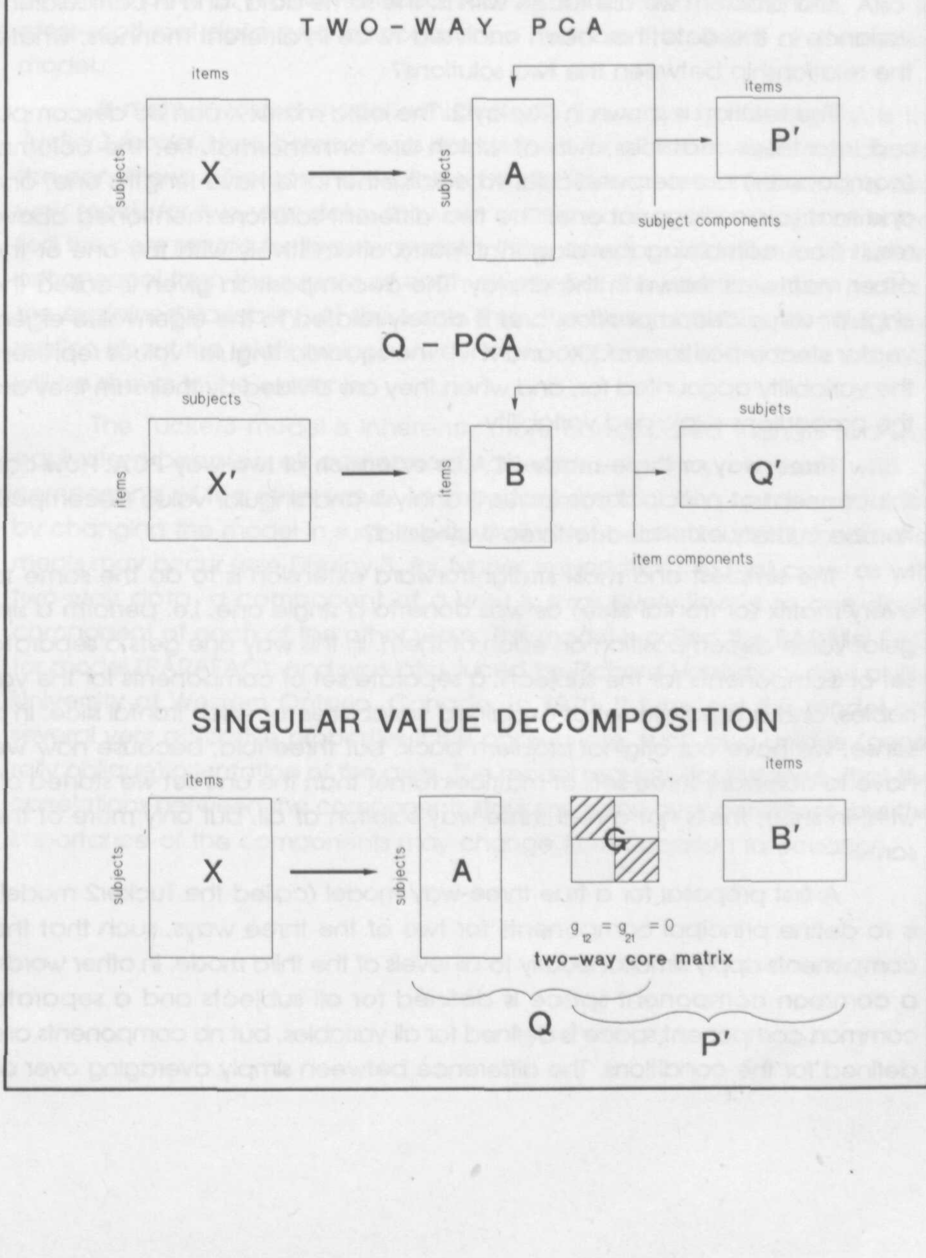
Stochastic three-way models. Many different models have been proposed for three-way data, but it is impossible to treat even a substantial part of them in this paper. One major distinction that can be made is that between *stochastic models* in which subjects are treated as replications and no parameters are estimated for the subjects. The means and covariances are deemed sufficient for analysis (in technical terms, they are sufficient statistics for the parameters). Such models fall within the true realm of inferential statistics, and many of the models can be tackled within the framework of linear structural relations (LISREL). In another chapter, Prof. Mellenbergh discusses several applications of this approach, be it not for three-way data.

Data-analytic three-way models. The other approach is that of *nonstochastic models* or *data-analytic models* which are primarily used for description, and in which parameter estimates are derived for the subjects as well. Three of such models will be the focus of attention of this paper. In particular, three-way principal component analysis with extended core matrix (Tucker2 model), Parallel factor analysis (PARAFAC), and full three-way principal component analysis (Tucker3 model).

Three-way component models

Two-way PCA as a matrix descomposition. To give an insight into three-way principal component analysis (which is generally called three-*mode* principal component analysis), it is easiest to start with ordinary two-way PCA. We will start with a subjects-by-variables matrix X which we want to reduce to a smaller number of components (or 'latent variables') on which the *subjects* have (component) scores, and a set of *weights for the variables*, in this context generally referred to as *loadings*.

DISPLAY 2


TWO-WAY PCA
items
subjects
X
subjects
A
items
P′
subject components
Q − PCA
subjects
items
X′
items
B
subjets
Q′
item components
SINGULAR VALUE DECOMPOSITION
items
subjects
X
subjects
A
G
items
B′
g12 = g21 = 0
two-way core matrix
Q
P′

Alternatively, we may start with a variables-by-subjects matrix X', which we want to reduce to a smaller number of components (or 'subject types') on which *variables* have *scores*, and a set of *weights for the subjects*. Such an analysis is sometimes known as 'Q-PCA', after Cattell (1966).

The problem we are faced with is: The same data, and in particular the variance in the data, has been analysed twice in different manners, what is the relationship between the two solutions?

The solution is shown in Display 2. The data matrix X can be descomposed into three matrices, two of which are orthonormal, i.e. the columns (components) are perpendicular to each other and have lengths one, and one matrix is a diagonal one. The two different solutions mentioned above result from combining the diagonal matrix alternatively with the one or the other matrix, as shown in the display. The decomposition given is called the *singular value decomposition*, and is closely related to the eigenvalue-eigenvector decompositions of XX' and X' X. The squared singular values represent the variability accounted for, and when they are divided by their sum they are the proportions explained variability.

Three-way or three-mode PCA as extension of two-way PCA. How can this concept of principal component analysis and singular value decomposition be fruitfully extended to three-way data?

The simplest and most straightforward extension is to do the same to every matrix (or frontal slice) as was done to a single one, i.e. perform a singular value decomposition on each of them. In this way one gets a separate set of components for the subjects, a separate set of components for the variables, and a separate set of explained variabilities for each frontal slice. In a sense, we have our original problem back, but three-fold, because now we have to compare three sets of matrices rather than the one set we started off with. In short, this is not a real three-way solution at all, but only more of the same.

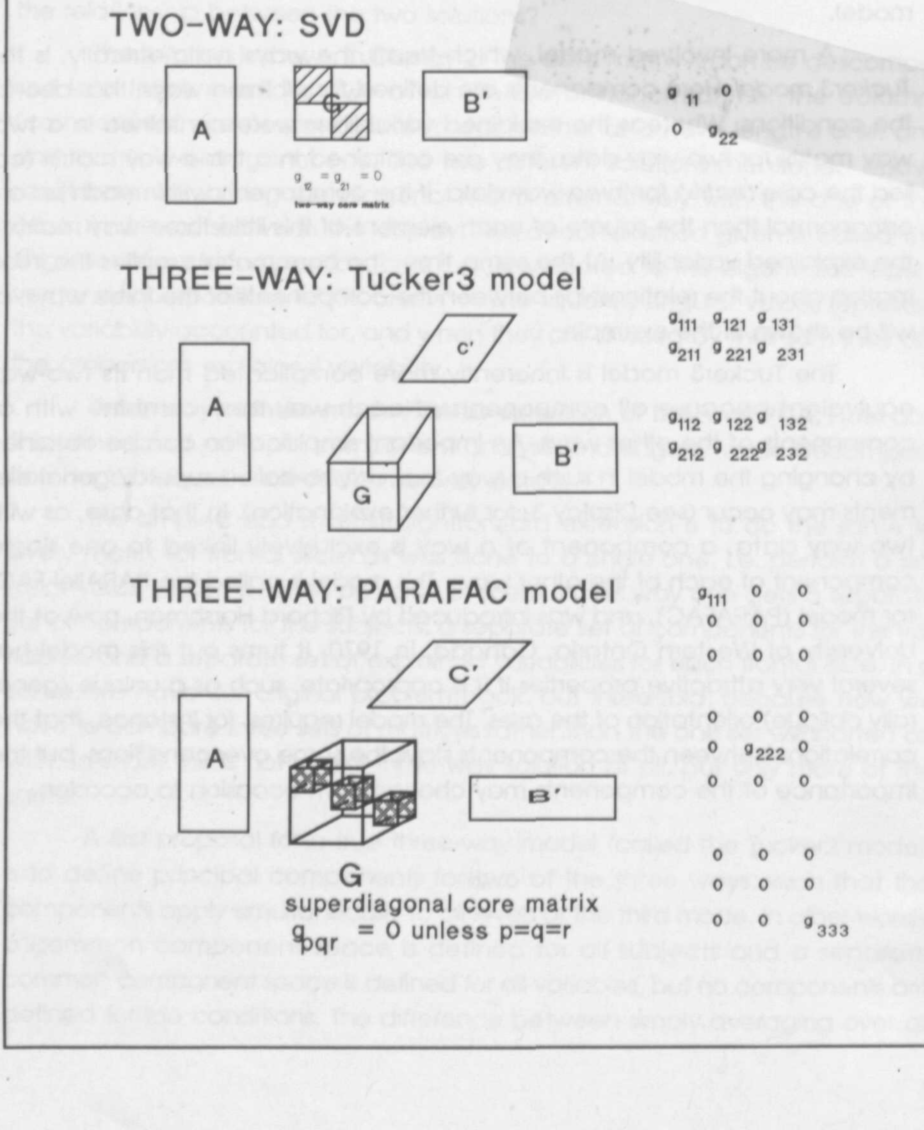
A first proposal for a true three-way model (called the Tucker2 model) is to define principal components for two of the three ways, such that the components apply simultaneously to all levels of the third mode. In other words, a common component space is defined for all subjects and a separate common component space is defined for all variables, but no components are defined for the conditions. The difference between simply averaging over all conditions, and then performing a singular value decomposition over the average conditions, is that in the model each condition has its own parameters to express how the common components are related to one another for that particular condition. For similarity matrices the Tucker2 model is a very general model for multidimensional scaling of individual differences. Also for cross-sectional data such as covariance matrices this seems an interesting model.

A more involved model, which treats the ways symmetrically, is the *Tucker3 model.* Here components are defined for all three ways, thus also for the conditions. Whereas the explained variabilities were contained in a two-way matrix for two-way data, they are contained in a three-way matrix (called the *core matrix*) for three-way data. If the components within each set are orthonormal than the square of each element of this little three-way matrix is the explained variability. At the same time, the core matrix supplies the information about the relationships between the components of the three ways, as will be shown in the example.

The Tucker3 model is inherently more complicated than its two-way equivalent because all components of each way may combine with all components of the other ways. An important simplification can be obtained by changing the model in such a way that only so-called superdiagonal elements may occur (see Display 3, for further explanation). In that case, as with two-way data, a component of a way is exclusively linked to one single component of each of the other ways. This model is called the PARAllel FACtor model (PARAFAC), and was introduced by Richard Harshman, now at the University of Western Ontario, Canada, in 1970. It turns out this model has several very attractive properties if it is appropriate, such as a unique (generally oblique) orientation of the axes. The model requires, for instance, that the correlations between the components stays the same over conditions, but the importance of the components may change from occasion to occasion.

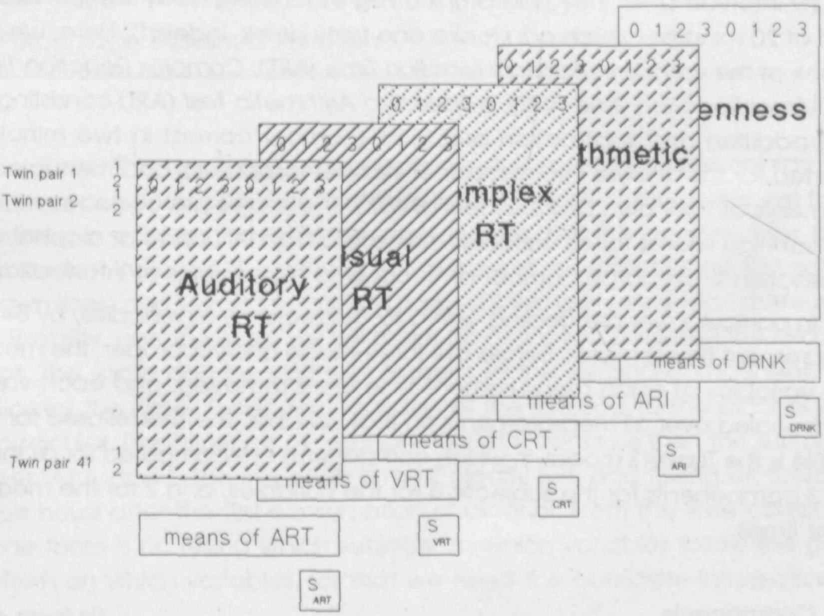
DISPLAY 3

CORE MATRICES



DISPLAY 4

DATA ARRAY FOR TWIN STUDY



82 x 8 x 5 Data Box

(with removed means and standard deviations)

APPLICATION: DRUNKEN TWINS

Data

Nick Martin and his colleagues of the Queensland Institute for Medical Research, Brisbane, Australia, collected a large amount of information from Australian twin pairs who received an acute or challenge dose of alcohol (see Martin, Oakeshott, Gibson, Starmer, Perl, & Wilks, 1985, for the full experimental details) In this paper, we will concentrate on 41 twin pairs who were measured at two separate occasions. At each occasion they were measured four time. The first time the subjects were sober. The other measurements were taken at hourly intervals after they had drunk 0.75g ethanol/kg body weight over a period of 20 minutes (which can make one fairly drunk, indeed). Here, we will only look at the variables: *Auditory reaction Time* (**ART**), *Complex Reaction Time* (**CRT**), *Visual Reaction Time* (**VRT**), a *speeded Arithmetic Test* (**ARI**) consisting of simple addition and subtraction problems (number correct in two minutes; converted for this analysis into number of incorret responses), and the subjects' judgements of their own *Drunkenness (DRNK)*. The scores are coded in such a way, that high scores for all variables indicate a high influence of alcohol, i.e. long reaction times, large number of errors, and high ratings of intoxication.

In particular, we are dealing with a 82 (subjects) by 5 (variables) by 8=2*4 (measurement times) matrix. Before the three-mode analysis proper, the means of the variables at each measurement time were removed, and each variable was scaled over all measurements on that variable. The model used for this example is the Tucker3 model, in which components are computed for all three ways: 3 components for the subjects, 3 for the variables, and 2 for the measurement times.

Components

Variables. First, we will present the components of the three modes. The structure in the three principal components of the variables has been enhanced by rotating them orthogonally according to a varimax criterion. The three axes V1, V2, and V3 an easily be labelled *"Reaction Time"* (RT), *"Arithmetic"* (ARI), and *"Self-rated Drunkenness"* (DRNK).

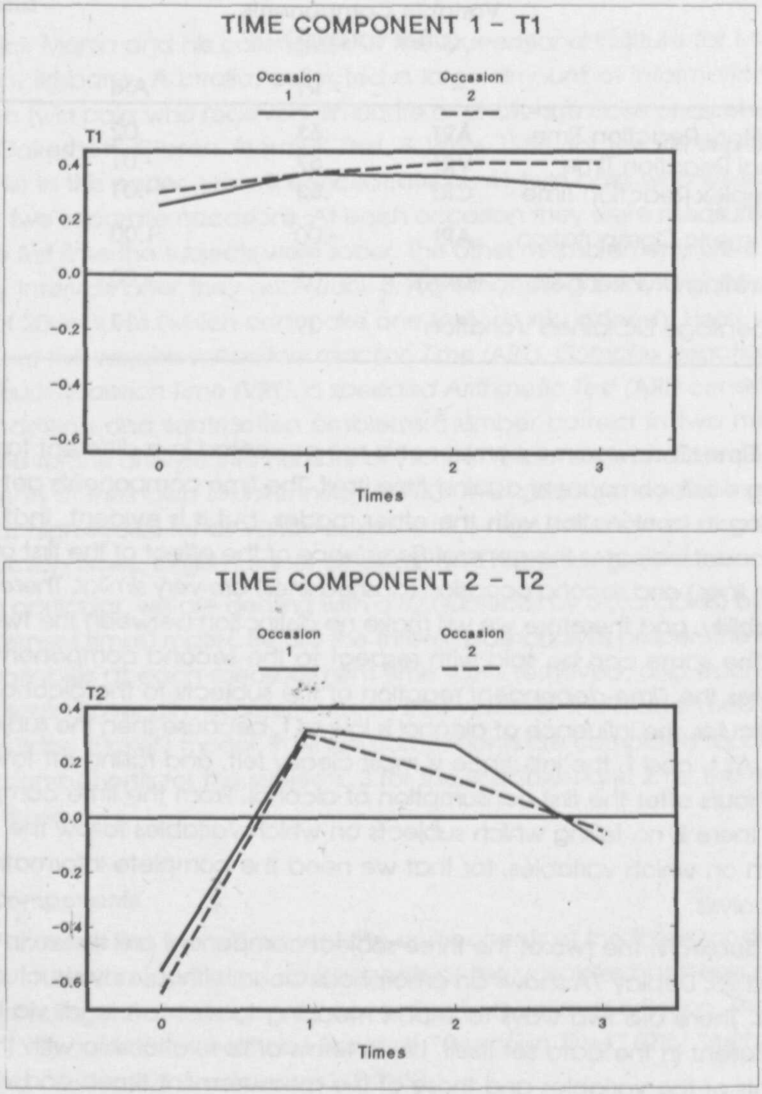
DISPLAY 5
Variable Components
(after varimax)

		RT	ARI	DRNK
Auditory Reaction Time	**ART**	**.63**	.02	-.08
Visual Reaction Time	**VRT**	**.57**	-.01	.05
Complex Reaction Time	**CRT**	**.53**	-.01	.02
Artihmetic Computation	**ARI**	-.00	**1.00**	.00
Self-rating Drunkenness	**DRNK**	.02	.00	**.99**
Porcentage Explained Variation		39	17	14

Time. The two time components are presented in a different fashion by plotting each component against time itself. The time components get their full meaning in conjunction with the other modes, but it is evident, that the first component indicates the general *Persistence* of the effect of the first occasion (drawn lines) and second occasion (dashed lines) are very similar. There is good replicability, and therefore we will make no distinction between the two occasions. The same can be said with respect to the second component, which indicates the *Time-dependent* reaction of the subjects to the alcohol intake. In particular, the influence of alcohol is low at t_0 because then the subjects are sober. At t_1 and t_2 the influence is most clearly felt, and falling off towards t_3, three hours after the first consumption of alcohol. From the time components alone there is no telling which subjects on which variables follow the general pattern on which variables, for that we need the complete information from the analysis.

Subjects. The two of the three subject component are shown in Displays 7A and 7B. Display 7A shows an amorphous cloud without any structure whatsoever. There are two ways to impart meaning to such a cloud: via information present in the data set itself, i.e. in terms of its relationship with the components of the variables and those of the measurement times, and via external variables with additional information on the twins.

Display 6



Time components: T1 = Persistent effect of alcohol
T2 = Time dependent effect of alcohol

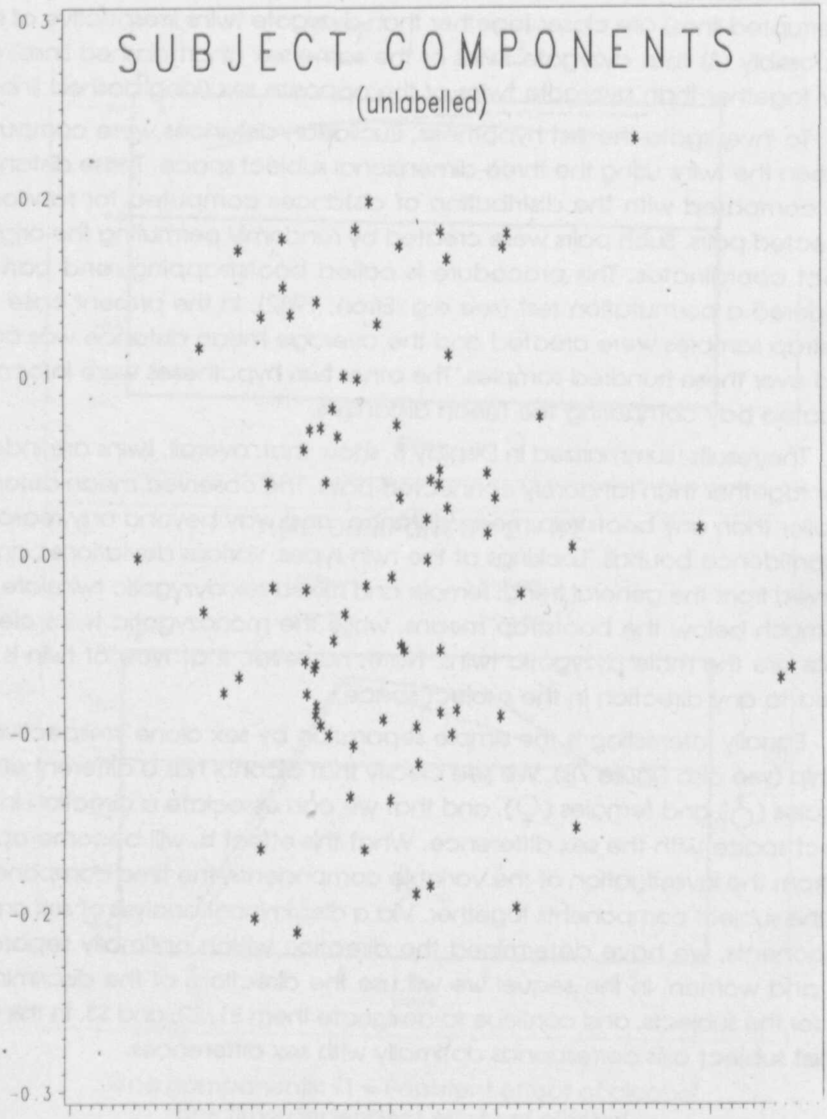
In particular, we can connect all twin pairs and label them according to type and sex (display 7B). One would expect (1) that twins are closer together than randomly paired subjects, (2) that monozygote twins (connected with uninterrupted lines) are closer together than dyzygote twins irrespective of sex, and possibly (3) that dyzygote twins of the same sex (short dashed lines) are closer together than dyzygote twins of the opposite sex (long dashed lines).

To investigate the first hypothesis, Euclidean distances were computed between the twins using the three-dimensional subject space. These distances were compared with the distribution of distances computed for randomly connected pairs. Such pairs were created by randomly permuting the original subject coordinates. This procedure is called bootstrapping, and can be considered a permutation test (see e.g. Efron, 1982). In the present case 100 bootstrap samples were created and the average mean distance was computed over these hundred samples. The other two hypotheses were informally evaluated bay comparing the mean distances.

The results, summarized in Display 8, show that, overall, twins are indeed closer together than randomly connected pairs. The observed mean distance is smaller than any bootstrap mean distance, and way beyond any reasonable confidence bounds. Lookings at the twin types, various deviations can be observed from the general trend: female and mixed-sex dyzygotic twins are not very much below the bootstrap means, while the monozygotic twins clearly are, as are the male dyzygotic twins. Note, however, that type of twin is not related to any direction in the subject space.

Equally interesting is the simple separation by sex alone irrespective of twinship (see also Figure 7B). We see clearly that alcohol has a different effect on males (♂) and females (♀), and that we can associate a direction in the subject space with the sex difference. What this effect is, will become apparent from the investigation of the variable components, the time components, and the subject components together. Via a discriminant analysis of sex on the components, we have determined the direction which optimally separates men and women. In the sequel we will use the directions of the discriminant axes for the subjects, and continue to designate them S1, S2, and S3. In this way the first subject axis corresponds optimally with sex differences.

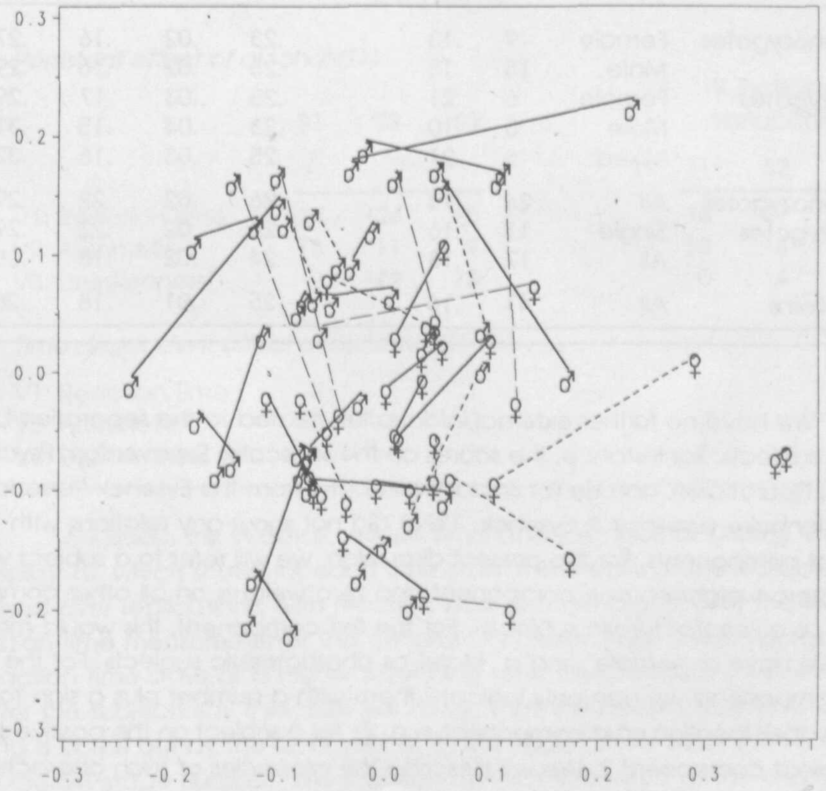
Display 7A


SUBJECT COMPONENTS
(unlabelled)
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
-0.3
-0.2
-0.1
0.0
0.1
0.2
0.3

Display 7B

SUBJECT COMPONENTS

(labelled by sex and zygosity)



monozygotes = ———————

dyzygotes/same sex = - - - -

dyzygotes/mixed sex = __ __ __

DISPLAY 8

Comparison of mean distances between twin pairs

Twin Type	Sex	N	Mean Twin Dist.	Bootstrap Distances Mean	S.D.	Min	Max
Monozygotes	Female	9	.13	.23	.02	.16	.27
	Male	15	.15	.25	.02	.16	.29
Dyzygotes	Female	6	.21	.25	.03	.17	.29
	Male	5	.10	.23	.04	.15	.31
	Mixed	6	.21	.25	.03	.16	.32
Monozygotes	All	24	.14	.26	.02	.22	.29
Dyzygotes	Single	11	.16	.24	.02	.22	.29
	All	17	.18	.24	.02	.18	.28
All twins	All	41	.16	.25	.01	.18	.28

We have no further external information related to the separation between subjects. For instance, the scores on the subscales Extraversion, Psychoticism, Neuroticism, and Lie (or Social Desirability) from the Eysenck Personality Questionnaire (Eysenck & Eysenck, 1975) did not show any relations with the subject components. For the present discussion, we will refer to a subject with a nonzero weight on one component and zero weights on all other components as a *'characteristic subjects'*. For the first component, this would mean that we have a *'Female'* and a 'Male' as characteristic subjects. For the other components we can only indicate them with a number plus a sign to indicate their location on a component, e.g. 2+ for a subject on the positive side of subject component 2. We will describe the properties of such characteristic subjects in terms of changes over time in their scores on the variable components, as expressed through the time components.

Core matrix

The relationships between the components of the various modes is contained in the core matrix as we pointed out before. For the present solution this core matrix is shown in Display 9. Note first of all that the T1 panel, referring to the persistent effect of alcohol on the subjects, has a rather complicated structure, suggesting that different 'characteristic subjects' have quite different reactions towards alcohol.

DISPLAY 9
Core Matrix

	S1	S2	S3	% Explained variability S1	S2	S3
Persistent effect of alcohol (T1)						
V1: Reaction Time	*24*	*-26*	-5	*18*	*21*	0
V2: Arithmetic	18	11	9	10	4	2
V3: Drunkenness	-2	*-12*	*15*	0	4	*7*
Time-dependent effect of alcohol (T2)						
V1: Reaction Time	-2	1	1	0	0	0
V2: Arithmetic	-3	-3	-0	0	0	0
V3: Drunkenness	-1	-6	7	0	1	1

To explain this in detail, we will simultaneously look at Display 9 and at Display 10, which shows for each time point the means of the variables averaged over replications, with reaction time also averaged over the three reaction-time measurements. The general conclusion from these figures is that reaction time stays at a higher level long after the alcohol intake, and long after the subjects say they feel less drunk. That the influence alcohol is declining is borne out by the arithmetic test. The figures show what the scores are of the *'Average Subject'*. This Average Subject is located at the origin of the subject space, and will be the reference point for all further explanations.

With this background we can now turn to the investigation of the fundamental relationships in the data as contained in the core matrix. The first thing to note is that a strong temporal effect (over and above that already displayed in the average curves), is only evident in the self-ratings of drunkenness. The other variables or variable components only have the trend as protrayed in the first time component, viz. the persistent effect of alcohol. This means that the curves of the characteristic subjects especially for the after-alcohol periods stay parallel to, either above or below, the mean curves. The easiest way to look at this core matrix is to describe the characteristic subjects one by one.

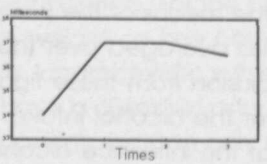
Characteristic Subject 1 (Female versus Male). Characteristic subject 1+ (S1+; Female) has persistently (T1) longer reaction times (V1) than average (core element (V1,S1,T1) = 24), also has persistently (T1) more arithmetic errors (V2) than average (core element (V2,S1,T1) = 18). On the other hand, the characteristic subject 1-(S1-; Male) has persistently shorter reaction times than average, and persistently less arithmetic errors than average. Thus the general trend, as is embodied in the means, is elevated for females with respect to males for the performance variable components, while there is no appreciable sex-related deviation from the average in perceived drunkenness (core element (V3,S1,T1) = -2).

DISPLAY 10

MEAN VALUES

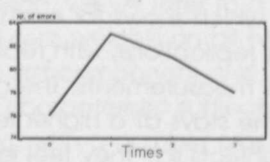
(=scores of Average Subject)

Reaction Time Pattern (RT)



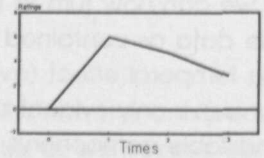
(Reaction time pattern is based on VRT, ART, and CRT scores)

Arithmetic (ARI)



(Number of errors are indirectly derived from numbers correct)

Self-rated Drunkenness (DRNK)



(Ratings based on both occasions)

Characteristic Subject 2 (No specific relationship with external variables known). The 2+ subject shows persistently shorter reaction times than average (core element (V1,S2,T1) = 26), has persistently more arthmetic errors than average [**(V2,S2,T1) = 11**]. He gives persistently lower drunkenness ratings [**(V3,S2,T1) = -12**], which show, in addition, an inverse pattern to that of the average curve [**(V3,S2,T2) = -6**], thereby attenuating the peak of the average curve. The 2- subject shows the reverse pattern: persistently longer reaction times and less arithmetic errors. This is accompanied by higher drunkenness ratings, which tend to emphasize the peak already present in the means, especially one hour after alcohol. Thus alcohol affects these subjects differently with respect to the performance measures, either reaction time is long and arithmetic low in errors, or vice versa. The self-ratings of drunkenness concur with the reaction times, but not with arithmetic. In addition, the subjects profess to be either fairly sensitive to the alcohol (2-), or are largely indifferent to it (2+), as their time-dependent curve counteracts the average one.

Characteristic subject 3 (no relationship with external variables known). Subject 3+ is about average on reaction time, but has persistently more errors and higher drunkenness ratings than average, and these ratings are time-dependent in that they elevate the peak of the Average Subject. Subject 3- is, of course, also average on reaction time, and makes persistently less errors and has lower ratings for drunkenness, with an attenuated peakedness directly after alcohol.

Thus high drunkenness ratings can occur both with large number of errors, and with long reaction times. For some subjects their feeling of drunkenness is reflected in elevated scores for arithmetic, and not for reaction times, while for other it is the reverse, that is a high feeling of drunkenness is reflected in elevated scores for reaction time, but not for arithmetic. And when both performance measures are high or low the drunkenness ratings tend to be average. Furthermore, note that when there are differences between subjects on the drunkenness ratings that higher than average scores tend to go together with higher peakedness, and reversely that low ratings go together with lower peakedness directly after alcohol. Thus emphasizing the sensitivity or insensitivity for alcohol.

CONCLUSION

By treating an example in considerable detail we have tried to convey some of the power of an integrated analysis of three-way data. In particular, we hope to have succeeded in showing that complex questions generally have complex answers. It demands a careful and thoughtful analysis, preferably with considerable theoretical insight into the subject matter.

In the present example, the theoretical background was not very profound, but this is primarily due to the very common sense notions and variables in the research. The fact that our samples consisted of twin pairs does not seem to be very relevant for explaining differences in tolerance to alcohol. In that respect, sex does a far better job. However, it became evident that twins in general have more similar reactions than arbitrarily paired persons, be it that for dyzygotes the situation is not unequivocal.

In addition to sex, one would like to find other external correlates to explain differences between subjects. Without such variables it is unrealistic to expect an understanding of differences between subjects on various measures. This becomes especially clear from the subject component for which we have external information. The differences between the sexes on the first component and the nature of these differences, i.e. the relative stronger deterioration of performance by women, suggests further research into its causes. Of course, such research is already being carried out, but it is interesting to see this difference emerge here, and moreover to see that it is not clearly related with differences in subjective perception of drunkenness by females and males.

A paper such as this is too short to show the full range of possibilities for analysing three-way data, but we sincerely hope that the models and techniques are sufficiently intriguing to provide the reader with a new vantage point for his or her own data. And the expectation is that you will find that many more data come in boxes than you had previously realised.

REFERENCES AND FURTHER READING

Cattell, R. B. (1966). The data box: Its ordering of total resources in terms of posible relational systems (pp. 67-128). In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology*. Chicago: Rand McNally.

Coppi, R., & Bolasco, S. (Eds.) (1989). *Multiway data analysis.* Amsterdam: North Holland/Elsevier.

Dai, K. (1982). Application of three-mode factor analysis to industrial design of chair styles. *Japanese Psychological Review, 25,* 91-103 (in Japanese).

Efron, B. (1982). The jackknife, the bootstrap, and other resampling plans. *Society for Industrial and Applied Mathematics Monographs, 38* (whole issue).

Eysenck, H. J., & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire.* London: Hodder & Stoughton.

Flury, B. (1989). *Common principal component analysis and related multivariate models.* New York: Wiley.

Harshman, R. A. (1970). Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis. *UCLA Working Papers in Phonetics, 16,* 1-84 (University Microfilms No. 10, 085).

Harshman, R. A., & Lundy, M. E. (1984 a). The PARAFAC model for three-way factor anlysis and multidimensional scaling. In H. G. Law, C. W. Snyder Jr., J. A. Hattie, & R. P. McDonald (Eds.) (1984). *Research methods for multimode data analysis* (pp. 122-215). New York: Praeger.

Harshman, R. A., & Lundy, M. E. (1984b). Data preprocessing and the extended PARAFAC model. In H. G. Law, C. W. Snyder Jr., J. A. Hattie, & R. P. McDonald (Eds.) (1984). *Research methods for multimode data analysis* (pp. 216-284). New York: Praeger.

Inukai, Y., Saito, S. , & Mishima, I. (1980). A vector model analysis of individual differences in sensory measurement of surface roughness. *Human Factors, 22* (1), 25-36.

Kroonenberg, P. M. (1983 a). *Three-mode principal component analysis. Theory and Applications.* Leiden: DSWO Press.

Kroonenberg, P. M. (1983 b). Annotated bibliography of three-mode factor analysis. *British Journal of Mathematical and Statistical Psychology, 36,* 81-113.

Kroonenberg, P. M. (1988). Three-mode analysis. In S. Kotz, N. L. Johnson, & C. B. Read (Eds.), *Encyclopedia of Statistical Sciences* (Vol. 9), pp. 231-236. New York: Wiley.

Lavit, C. (1988). *Analyse conjointe de plusieurs matrices de données.* Paris: Masson.

Law, H. G., Snyder Jr., C. W., Hattie, J. A., & McDonald, R. P. (Eds.) (1984). *Research methods for multimode data analysis.* New York: Praeger.

Martin, N. G., Oakeshott, J. G., Gibson, J. B., Starmer, G. A., Perl, J., & Wilks, A. V. (1985). A twin study of psychomoter and physiological responses to an acute dose of alcohol. *Behavior Genetics, 15,* 305-347.

Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika, 31,* 279-311.

Acknowledgements

My thanks go to Nick Martin of the Queensland Institute for Medical Research for making the data available. Piet Brouwer kindly supplied the drawings for Figure 1, and Jos Henselmans assisted with the óther drawings.

(2) Content: It was hypothesized that each of the two judgments of the content was influenced by this factor;

(3) Organization: It was hypothesized that each of the two judgments of the organization was influenced by this factor;

(4) Style: It was hypothesized that each of the two judgments of the style was influenced by this factor.

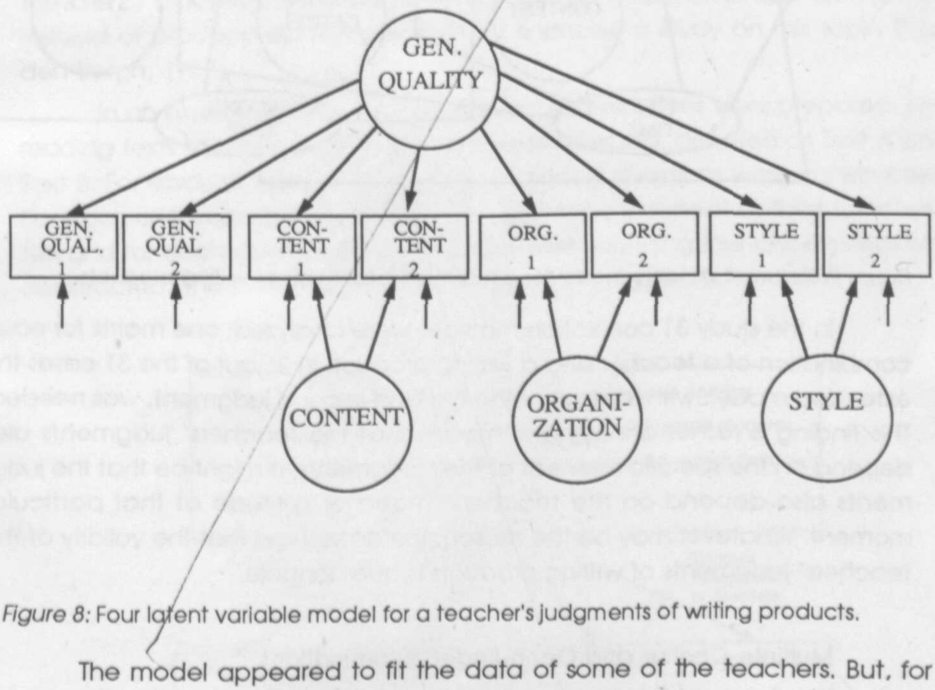


Figure 8: Four latent variable model for a teacher's judgments of writing products.

The model appeared to fit the data of some of the teachers. But, for most of the teachers the model did not fit the data, which means that another model must be specified for them. The model was extended with two other latent variables:

(5) a fifth factor for the first judgment, and

(6) a sixth factor for the second judgment.

The extended model is shown in Figure 9.

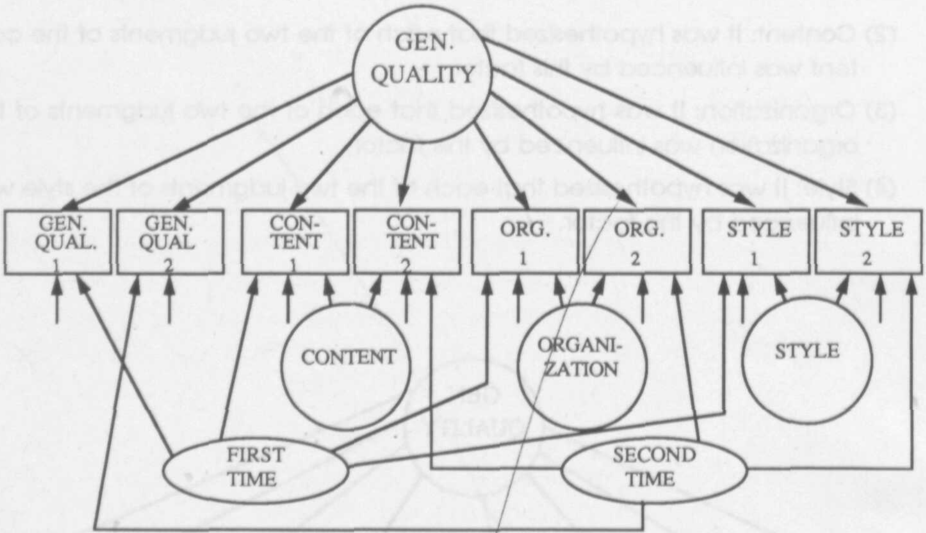


Figure 9: Six latent variable model for a teacher's judgments of writing products.

In the study 31 correlation matrices were analyzed: one matrix for each combination of a teacher and a writing product. In 26 out of the 31 cases the extended model, with factors for the first and second judgment, was needed. This finding is rather annoying: it means that the teachers' judgments also depend on the specific moment of their judgments. It might be that the judgments also depend on the teachers' mood or attitude at that particular moment. Whatever may be the reason, the data show that the validity of the teachers' judgments of writing products is questionable.

Multiple-Choice and Open-Ended Examinations

The final example is on the item format of the Dutch examinations at the end of the Lower Vocational and Lower Administrative Secondary Education. A part of the examination is an investigation of the examinees' reading comprehension. The examinee must read a Dutch text and he or she must answer a number of questions on the text. The questions investigate whether the examinee did understand the content of the text. The questions are of two different formats: First, a number of open-ended questions are posed and the examinee must write down the answer to the questions. Beside these open-ended questions the examinee must also answer a number of multiple-choice questions. A question is posed and the examinee must choose one option out of four options.

The Dutch Government was interested to Know whether multiple-choice and open-ended questions tap the same intellectual abilities of the examinees. For example, in open-ended questions the examinee must generate and produce their own answers, whereas in multiple-choice items the examinees can recognize the correct option. The government was concerned that the use of multiple-choice items would put too much emphasis on memory instead of productivity; the government financed a study on this topic (Van den Bergh, 1989).

In an experimental study four different examinations were prepared. Two reading texts of previous examinations were selected, denoted as Text A and Text B. For each of the two texts 25 open-ended questions were constructed. For each of these versions one correct and three incorrect options were written and for each open-ended question a corresponding four-choice item was constructed. The design of the study is shown in Figure 10.

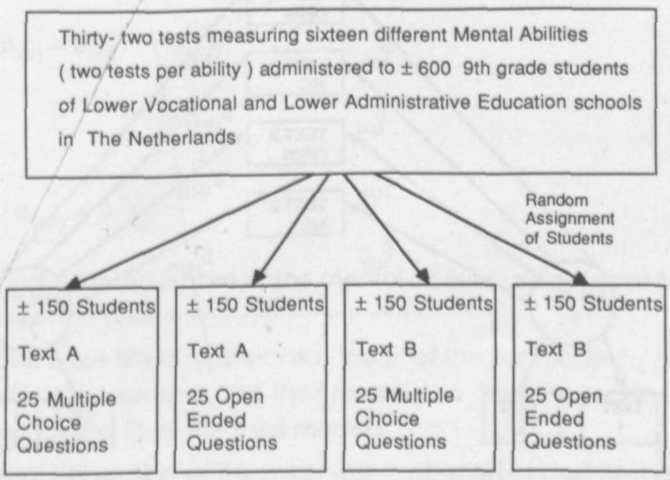


Figure 10: Design of Van der Bergh's (1989) study on mental abilities and item format.

The study has four different experimental examinations: (1) Text A, open-ended questions, (2) Text A, multiple-choice questions, (3) Text B, open-ended questions, (4) Text B, multiple-choice questions. A group of about 600 9th grade students of Lower Vocational and Lower Administrative Education schools was used. The four examinations were assigned at random to the students, which means that each of the four examinations was administered to about 150 students; see Figure 10. Moreover, 32 psychological tests were administered to all of the students. The tests measure sixteen different mental abilities of a semantic nature. The mental abilities were selected in such a way that differences between open-ended and multiple-choice items could be expected. For example, tests for the memory of semantic material were used and tests for the production of semantic material. It was hypothesized that the open-ended questions would appeal more to the production abilities and that the multiple-choice items would appeal more to the momory abilities.

The four correlation matrices of the 32 tests and the examination were computed. The data were analyzed using the multiple-group option of the program LISREL (Jöreskog & Sörbom, 1986).

The general form of the model is shown in Figure 11.

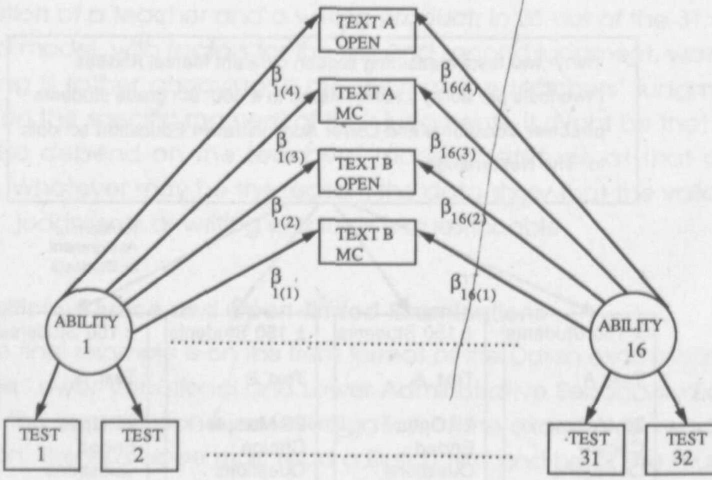


Figure 11: Model for the influence of sixteen mental abilities on open-ended and multiple-choice examinations.

The following three models were specified:

1. Each of the sixteen mental abilities is measured by two psychological tests. The arrows from the abilities to the examinations indicate the influence of the ability on the examinations. In the first model the influence of each of the mental abilities is the same for each of the four examinations, i.e.

$$
\begin{array}{c}
B_{1(1)} = B_{1(2)} = B_{1(3)} = B_{1(4)} \\
\cdot \quad \cdot \quad \cdot \quad \cdot \\
\cdot \quad \cdot \quad \cdot \quad \cdot \\
\cdot \quad \cdot \quad \cdot \quad \cdot \\
B_{16(1)} = B_{16(2)} = B_{16(3)} = B_{16(4)}.
\end{array}
\tag{6}
$$

If this model fits the data it means that the influence of the mental abilities is identical for multiple-choice and open-ended questions, for both texts.

2. In the second model the influence of the mental abilities on multiple-choice questions is the same as the influence on the open-ended questions, but the influence is different for the two texts, i.e.

$$
\begin{array}{c}
B_{1(1)} = B_{1(2)} \; ; B_{1(3)} = B_{1(4)} \\
\cdot \quad \cdot \quad \cdot \quad \cdot \\
\cdot \quad \cdot \quad \cdot \quad \cdot \\
\cdot \quad \cdot \quad \cdot \quad \cdot \\
B_{16(1)} = B_{16(2)} \; ; B_{16(3)} = B_{16(4)}.
\end{array}
\tag{7}
$$

3. In the third model the influences of the mental abilities are different for each of the four examinations.

The models were fitted to the data. The fit of the first model to the data is less than the fit of the second and third model. The fit of the second model is nearly as good as the fit of the third model.

The general conclusion of this study was that open-ended and multiple-choice questions on reading comprehension tap the same mental abilities. But, it was also concluded that different reading texts can tap different mental abilities. In other words: The reading text of the examination, and not the format of the questions, mainly determines which of the mental abilities are evoked.

CONCLUSION

In this paper some principles of the structural analysis of covariance and correlation matrices have been discussed, and some of the applications were shown. The general conclusion is that structural analysis is very fruitful for empirical research. The examples showed that structural analysis is useful for both theoretical and applied research. The examples on the judgment of writing products and examinations were from applied studies, financed by the Dutch government and the results are used for governmental policy making.

On the other hand, structural analysis is not an easy job. The field is beset with hard problems of methodological and statistical nature; also issues of the philosophy of science, such as the formulation and testing of hypotheses, are of importance. Structural analysis offers many opportunities for theoretical and applied research, but it has also its limitations.

REFERENCES

De Glopper, C.M. (1989). *Schrijven beschreven.* Dissertation University of Amsterdam, The Netherlands. 's-Gravenhage: SVO.

Jöreskog, K.G. Structural analysis of covariance and correlation matrices. *Psychometrika*, 1978, 443-477.

Jöreskog K.G., & Sörbom, D.G. (1986). *LISREL: Analysis of linear structural relationships by maximum likelihood, instrumental variables and least squares methods.* Uppsala, Sweden: University of Uppsala.

Lawley, D.N., & Maxwell, A.E. (1971). *Factor analysis as a statical method* (2nd ed.). London: Butterworth.

Marsh, H.W., Balla, J.R., & Mc.Donald, R.P. (1988). Goodnes-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, 103, 391-410.

Sarisw, W.E., & Stronkhorst, L.H. (1984). *Causal modelling in nonexperimental research.* Amsterdam, The Netherlands: Sociometric Research Foundation.

Satorra, A., & Saris, W.E. (1985). Power of the likelihood ratio test in covariance structure models. *Psychometrika*, 50, 83-90.

Van den Berg, H. (1989). *Examens geëxamineerd*. Dissertation University of Amsterdam, The Netherlands. s'-Gravenhage: SVO.

Vorst, H.C.M. (1989). *Three components of educational attitudes in four countries and two educational levels.* Amsterdam, The Netherlands: Section or Methodology, Departament or Psychology, University of Amsterdam.