1 Old and New Approaches to Latent Variable Modelling

David J. Bartholomew

London School of Economics and Political Science

1.1 The Old Approach

To find the precursor of contemporary latent variable modeling one must go back to the beginning of the 20th century and Charles Spearman's invention of factor analysis. This was followed, half a century later, by latent class and latent trait analysis and, from the 1960's onwards, by covariance structure analysis. The most recent additions to the family have been in the area of latent time series analysis. This chapter briefly reviews each of these fields in turn as a foundation for the evaluations and comparisons which are made later.

1.1.1 Factor analysis

Spearman's (1904) original paper on factor analysis is remarkable, not so much for what it achieved, which was primitive by today's standards, but for the path-breaking character of its central idea. He was writing when statistical theory was in its infancy. Apart from regression analysis, all of today's multivariate methods lay far in the future. Therefore Spearman had not only to formulate the central idea, but to devise the algebraic and computational methods for delivering the results. At the heart of the analysis was the discovery that one could infer the presence of a latent dimension of variation from the pattern of the pairwise correlations coefficients. However, Spearman was somewhat blinkered in his view by his belief in a single underlying latent variable corresponding to general ability, or intelligence. The data did not support this hypothesis and it was left to others, notably Thurstone in the 1930's, to extend the theory to what became commonly known as multiple factor analysis.

Factor analysis was created by, and almost entirely developed by, psychologists. Hotelling's introduction of principal components analysis in 1933 approached essentially the same problem from a different perspective, but his work seems to have made little impact on practitioners at the time.

It was not until the 1960's and the publication of Lawley and Maxwell's (1971) book *Factor Analysis as a Statistical Method* that any sustained attempt was made to treat the subject statistically. Even then there was little effect on statisticians who, typically, continued to regard factor analysis as an alien and highly subjective activity which could not compete with principal components analysis. Gradually the range of applications widened but without going far beyond the framework provided by the founders.

1.1.2 Latent class analysis

Latent class analysis, along with latent trait analysis (discussed later), have their roots in the work of the sociologist, Paul Lazarsfeld in the 1960's. Under the general umbrella of latent structure analysis these techniques were intended as tools of sociological analysis. Although Lazarsfeld recognized certain affinities with factor analysis he emphasized the differences. Thus in the old approach these families of methods were regarded as quite distinct.

Although statistical theory had made great strides since Spearman's day there was little input from statisticians until Leo Goodman began to develop efficient methods for handling the latent class model around 1970.

1.1.3 Latent trait analysis

Although a latent trait model differs from a latent class model only in the fact that the latent dimension is viewed as continuous rather than categorical, it is considered separately because it owes its development to one particular application. Educational testing is based on the idea that human abilities vary and that individuals can be located on a scale of the ability under investigation by the answers they give to a set of test items. The latent trait model provides the link between the responses and the underlying trait. A seminal contribution to the theory was provided by Birnbaum (1968) but today there is an enormous literature, both applied and theoretical, including books, journals such as *Applied Psychological Measurement* and a multitude of articles.

1.1.4 Covariance structure analysis

This term covers developments stemming from the work of Jöreskog in the 1960's. It is a generalization of factor analysis in which one explores causal relationships (usually linear) among the latent variables. The significance of the word *covariance* is that these models are fitted, as in factor analysis, by comparing the observed covariance matrix of the data with that predicted by the model. Since much of empirical social science is concerned with trying to establish causal relationships between unobservable variables, this form of analysis has found many applications. This work has been greatly facilitated by the availability of good software packages whose sophistication has kept pace with the speed and capacity of desk-top (or lap-top) computers. In some quarters, empirical social research has become almost synonymous with LISREL analysis. The acronym LISREL has virtually become a generic title for linear structure relations modeling.

1.1.5 Latent time series

The earliest use of latent variable ideas in time series appears to have been due to Wiggins (1973) but, as so often happens, it was not followed up. Much later there was rapid growth in work on latent (or "hidden" as they are often called) Markov chains. If individuals move between a set of categories over time it may be that their movement can be modeled by a Markov chain. Sometimes their category cannot be observed directly and the state of the individual must be inferred, indirectly, from other variables related to that state. The true Markov chain is thus latent, or hidden. An introduction to such processes is given in MacDonald and Zucchini (1997). Closely related work has been going on, independently, in the modeling of neural networks. Harvey and Chung (2000) proposed a latent structural time series model to model the local linear trend in unemployment. In this context two observed series are regarded as being imperfect indicators of "true" unemployment.

1.2 The New Approach

The new, or statistical, approach derives from the observation that all of the models behind the foregoing examples are, from a statistical point of view, mixtures. The basis for this remark can be explained by reference to a simple example which, at first sight, appears to have little to do with latent variables. If all members of a population have a very small and equal chance of having an accident on any day, then the distribution of the number of accidents per month, say, will have a Poisson distribution. In practice the observed distribution often has greater dispersion than predicted by the Poisson hypothesis. This can be explained by supposing that the daily risk of accident varies from one individual to another. In other words, there appears to be an unobservable source of variation which may be called "accident proneness". The latter is a latent variable. The actual distribution of number of accidents is thus a (continuous) mixture of Poisson distributions.

The position is essentially the same with the latent variable models previously discussed. The latent variable is a source of unobservable variation in some quantity, which characterizes members of the population. For the latent class model this latent variable is categorical, for the latent trait and factor analysis model it is continuous. The actual distribution of the manifest variables is then a mixture of the simple distributions they would have had in the absence of that heterogeneity. That simpler distribution is deducible from the assumed behaviour of individuals with the same ability - or whatever it is that distinguishes them. This will be made more precise below.

1.2.1 Origins of the new approach

The first attempt to express all latent variable models within a common mathematical framework appears to have been that of Anderson (1959). The title of the paper suggests that it is concerned only with the latent class model and this may have caused his seminal contribution to be overlooked. Fielding (1977) used Anderson's treatment in his exposition of latent structure models but this did not appear to have been taken up until the present author used it as a basis for handling the factor analysis of categorical data (Bartholomew 1980). This work was developed in Bartholomew (1984) by the introduction of exponential family models and the key concept of sufficiency. This approach, set out in Bartholomew and Knott (1999), lies behind the treatment of the present chapter. One of the most general treatments, which embraces a very wide family of models, is also contained in Arminger and Küsters (1988).

1.2.2 Where is the new approach located on the map of statistics?

Statistical inference starts with data and seeks to generalize from it. It does this by setting up a probability model which defines the process by which the data are supposed to have been generated. We have observations on a, possibly multivariate, random variable $\mathbf{x}$ and wish to make inferences about the process which is determined by a set of parameters ξ. The link between the two is expressed by the distribution of $\mathbf{x}$ given ξ. Frequentist inference treats ξ as fixed; Bayesian inference treats ξ as a random variable.

In latent variables analysis one may think of $\mathbf{x}$ as partitioned into two parts $\mathbf{x}$ and $\mathbf{y}$, where $\mathbf{x}$ is observed and $\mathbf{y}$, the latent variable, is not observed. Formally then, this is a standard inference problem in which some of the variables are missing. The model will have to begin with the distribution of $\mathbf{x}$ given ξ and $\mathbf{y}$. A purely frequentist approach would treat ξ and $\mathbf{y}$ as parameters, whereas the Bayesian would need a joint prior distribution for ξ and $\mathbf{y}$. However, there is now an intermediate position, which is more appropriate in many applications, and that is to treat $\mathbf{y}$ as a random variable with ξ fixed. Regarding $\mathbf{y}$ as fixed would enable one to make inferences about the values of $\mathbf{y}$ for these members of the sample. But often one is interested in predicting the values of $\mathbf{y}$ for other members of the population on the basis of their $\mathbf{x}$-values. This is the case, for example, when one is constructing a measuring instrument for some mental ability. That is one wants a means of locating an individual on a scale of ability on the basis of a set of scores, $\mathbf{x}$, obtained in a test. In that case it is more appropriate to treat $\mathbf{y}$ as a random variable also.

This problem can also be expressed in terms of probability distributions as follows. For an observable random variable $\mathbf{x}$, the distribution can be represented in the form

$$f(\mathbf{x} \mid \xi) = \int f(\mathbf{x} \mid \mathbf{y}, \xi) h(\mathbf{y})\, d\mathbf{y}. \tag{1.1}$$

Here, for simplicity, the notation is that of continuous random variables, but the point being made is quite general.

1.2.3 Basic relationships

Equation 1.1 reveals an important indeterminacy in the problem defined. One can make any one-to-one transformation of $\mathbf{y}$ without changing $f(\mathbf{x} \mid \xi)$. In other words the distribution of $\mathbf{y}$ is indeterminate. There can be no *empirical* justification for choosing one prior distribution for $\mathbf{y}$ rather than another. This might seem to be a crippling handicap but, as shall now be seen, important deductions can be made which hold for all prior distributions.

The problem is to say something about the missing values, $\mathbf{y}$, when one knows $\mathbf{x}$. All of the information about $\mathbf{y}$ given $\mathbf{x}$ is contained in

$$f(\mathbf{y} \mid \mathbf{x}) = h(\mathbf{y})f(\mathbf{x} \mid \mathbf{y})/f(\mathbf{x}) \tag{1.2}$$

where ξ has been suppressed in the notation because it is not relevant for the point being made.

At first sight one is at an impasse because of the presence of the unknown $h(\mathbf{y})$ on the right hand side. However, suppose that it were the case that $f(\mathbf{x} \mid \mathbf{y})$ could be expressed in the form

$$f(\mathbf{x} \mid \mathbf{y}) = g(\mathbf{X} \mid \mathbf{y})\phi(\mathbf{x}) \tag{1.3}$$

where $\mathbf{X}$ is a function of $\mathbf{x}$ of the same dimension as $\mathbf{y}$. Then one would find, by substitution into equation 1.2, that

$$f(\mathbf{y} \mid \mathbf{x}) = f(\mathbf{y} \mid \mathbf{X}) \propto h(\mathbf{y})g(\mathbf{X} \mid \mathbf{y}). \tag{1.4}$$

The point of this manoeuvre is that all that one needs to know about $\mathbf{x}$ in order to determine $f(\mathbf{y} \mid \mathbf{x})$ is the statistic $\mathbf{X}$. This provides a statistic which, in a precise sense, contains all the information about $\mathbf{y}$ which is provided by $\mathbf{x}$. In that sense, then, one can use $\mathbf{X}$ in place of $\mathbf{y}$.

The usefulness of this observation depends, of course, on whether or not the representation of (1.3) is possible in a large enough class of problems. One needs to know for what class of models, defined by $f(\mathbf{x} \mid \mathbf{y})$, is this factorization possible. This is a simpler question than it appears to be because other considerations, outlined in Section 3.1 below, mean that one can restrict attention to the case where the x's are conditionally independent (see Equation 1.5). This means that one only has to ask the question of the univariate distributions of the individual x_i's. Roughly speaking this means that one requires that $f(x_i \mid \mathbf{y})$ be a member of an exponential family for all i (for further details see Bartholomew & Knott, 1999). Fortunately this family, which includes the normal, binomial, multinomial and gamma distributions, is large enough for most practical purposes.

The relationship given by Equation 1.3 is referred to as the *sufficiency principle* because it is, essentially, a statement that $\mathbf{X}$ is sufficient for $\mathbf{y}$ in the Bayesian sense. It should be noted that, in saying this, all parameters in the model are treated as known.

1.3 The General Linear Latent Variable Model

1.3.1 Theory

The foregoing analysis has been very general, and deliberately so. In order to relate it to the various latent variable models in use the analysis must now be more specific.

In a typical problem $\mathbf{x}$ is a vector of dimension p say, where p is often large. Elements of $\mathbf{x}$ may be scores on test items or responses to questions in a survey, for example. The point of introducing the latent variables $\mathbf{y}$, is to explain the inter-dependencies among the x's. If this can be done with a small number, q, of y's a substantial reduction in dimensionality shall be achieved. One may also hope that the y's can be identified with more fundamental underlying variables like attitudes or abilities.

If the effect of the y's is to induce dependencies among the x's, then one would know that enough y's have been introduced if, when one conditions on them, the x's are independent. That is, one needs to introduce just sufficient y's to make

$$f(\mathbf{x} \mid \mathbf{y}) = \prod_{i=1}^{p} f_i(x_i \mid \mathbf{y}). \tag{1.5}$$

The problem is then to find distributions $\{f_i(x_i \mid \mathbf{y})\}$ such that the sufficiency property holds.

There are many ways in which this can be done but one such way produces a large enough class of models to meet most practical needs. Thus, consider distributions of the following form

$$f_i(x_i \mid \theta_i) = F_i(x_i)G(\theta_i)e^{\theta_i u_i(x_i)} \tag{1.6}$$

where θ_i is some function of $\mathbf{y}$. It is then easily verified that there exists a sufficient statistic

$$\mathbf{X} = \sum_{i=1}^{p} u_i(x_i). \tag{1.7}$$

The particular special case considered, henceforth known as *the general linear latent variable model* (GLLVM), supposes that

$$\theta_i = \alpha_{i0} + \alpha_{i1}y_1 + \alpha_{i2}y_2 + \ldots + \alpha_{iq}y_q \qquad (i = 1, 2, \ldots q). \tag{1.8}$$

This produces most of the standard models - and many more besides.

1.3.2 Examples

Two of the most important examples arise when the x's are (a) all binary or (b) all normal.

The binary case. If x_i is binary then, conditional upon $\mathbf{y}$, it is reasonable to assume that it has a Bernoulli distribution with $\Pr\{x_i = 1 \mid \mathbf{y}\} = \pi_i(\mathbf{y})$. This is a member of the exponential family (1.6) with

$$\theta_i = \text{logit}\pi_i(\mathbf{y}) = \alpha_{i0} + \alpha_{i1}y_1 + \alpha_{i2}y_2 + \cdots + \alpha_{iq}y_q \qquad (i = 1,\ 2,\ \ldots p) \quad (1.9)$$

This is a latent trait model in which q is usually taken to be 1, when it is known as the logit model.

The normal case. If x is normal with $x_i \mid \mathbf{y} \sim N(\mu_i,\ \sigma_i^2) \qquad (i = 1,\ 2,\ \ldots p)$ the parameter θ_i in (1.6) is

$$\theta_i = \mu_i/\sigma_i = \alpha_{i0} + \alpha_{i1}y_1 + \ldots + \alpha_{iq}y_q \qquad (i = 1, 2, \ldots p). \qquad (1.10)$$

Since the distribution depends on two parameters, μ_i and σ_i, one of them must be treated as a nuisance parameter. If this is chosen to be σ_i, one may write the model

$$x_i = \lambda_{i0} + \lambda_{i1}y_1 + \lambda_{i2}y_2 + \ldots \lambda_{iq}y_q + e_i \qquad (i = 1,\ 2,\ \ldots p) \qquad (1.11)$$

where $\lambda_{ij} = \alpha_{ij}\sigma_i$ $(j = 1,\ 2,\ \ldots q)$ and $e_i \sim N(0,\ \sigma_i^2)$ with $\{e_i\}$ independent of $\mathbf{y}$. This will be recognized as the standard representation of the linear normal factor model.

Other special cases can be found in Bartholomew and Knott (1999) including the latent class model which can be regarded as a special case of the general latent trait model given above.

It is interesting to note that for both of the examples given above

$$X_j = \sum_{i=1}^{p} \alpha_{ij}x_i \qquad (j = 1,\ 2,\ \ldots q).$$

Weighted sums of the manifest variables have long been used as indices for underlying latent variables on purely intuitive grounds. The foregoing theory provides a more fundamental rationale for this practice.

1.4 Contrasts Between the Old and New Approaches

1.4.1 Computation

Factor analysis was introduced at a time when computational resources were very limited by today's standards. The inversion of even small matrices was very time consuming on a hand calculator and beset by numerical instabilities. This not only made fitting the models very slow, but it had a distorting effect on the development of the subject. Great efforts were made to devise shortcuts and approximations for parameter estimation. The calculation of standard errors was almost beyond reach. The matter of rotation was as much an art as a science, and this contributed to the perception by some that factor analysis was little better than mumbo jumbo.

Things were little better when latent structure analysis came on the scene in the 1950's. Inefficient methods of fitting based on moments and such like took precedence simply because they were feasible. There was virtually nothing in common between the methods used for fitting the various models beyond their numerical complexity. As computers became more powerful towards the end of the 20th century, a degree of commonality became apparent in the unifying effect of maximum likelihood estimation, but this did not exploit the common structure revealed by the new approach. The possibility of a single algorithm for fitting all models derived from the new approach was pointed out by Bartholomew and Knott (1999, Section 7) and this has now been implemented by Moustaki (1999).

1.4.2 Disciplinary focus

Another distorting feature springs from the diverse disciplinary origins of the various models. Factor analysis was invented by a psychologist and largely developed by psychologists. Latent structure analysis was a product of sociology. This close tie with substantive problems had obvious advantages, principally that the problems tackled were those which are important rather than merely tractable. But it also had disadvantages. Many of the innovators lacked the technical tools necessary and did not always realize that some, at least, were already available in other fields. By focussing on the particular psychological hypothesis of a single general factor, Spearman failed to see the importance of multiple factor analysis. Lazarsfeld emphasized the difference between his own work on latent structure and factor analysis, which were unimportant, and minimized the similarities, which were fundamental. In such ways progress was slowed and professional statisticians, who did have something to offer, were debarred from entering the field. When they eventually did make a tentative entry in the shape of the first edition of Lawley and Maxwell's book in the 1960's, the contribution was not warmly welcomed by either side!

1.4.3 Types of variables

One rather surprising feature which delayed the unification of the subject on the lines set out here runs through the whole of statistics, but is particularly conspicuous in latent variable modelling. This is the distinction between continuous and categorical variables.

The development of statistical theory for continuous variables was much more rapid than for categorical variables. This doubtless owed much to the fact that Karl Pearson and Ronald Fisher were mainly interested in problems involving continuous variables and, once their bandwagon was rolling, that was where the theoreticians wanted to be. There were some points of contact as, for example, on correlation and association but there seems to have been little recognition that much of what could be done for continuous variables could, in principle, also be done for categorical variables or for mixtures of the two types. In part this was a notational matter. A perusal of Goodman's work on latent class analysis (e.g., Goodman, 1974), in which he uses a precise but forbidding notation, obscures rather than reveals the links with latent trait or factor analysis. Formulating the new approach in a sufficiently abstract form to include all types of variables, reveals the essential common structure and so makes matters simpler.

1.4.4 Probability modelling

A probability model is the foundation of statistical analysis. Faced with a new problem the statistician will determine the variables involved and express the relationships between them in probabilistic terms. There are, of course, standard models for common problems, so the work does not always have to be done *ab initio*. However, what is now almost a matter of routine is a relatively recent phenomenon and much of the development of latent variable models lies on the far side of the water-shed, which may be roughly dated to the 1950's. This was common to all branches of statistics but it can easily be illustrated by reference to factor analysis.

In approaching the subject today, one would naturally think in terms of probability distributions and ask what is the distribution of $\mathbf{x}$ given $\mathbf{y}$. Approaching it in this way one might write

$$\mathbf{x} = \mu + \mathbf{\Lambda y} + \mathbf{e} \tag{1.12}$$

or, equivalently,

$$\mathbf{x} \sim N(\mu + \mathbf{\Lambda y},\ \psi) \tag{1.13}$$

with appropriate further assumptions about independence and the distribution of $\mathbf{y}$. Starting from this, one can construct a likelihood function and from that, devise methods of estimation, testing goodness of fit and so on. In earlier times the starting point would have been the structure of the covariance (or correlation) matrix, $\mathbf{\Sigma}$, and the attempt to find a representation of the form

$$\mathbf{\Sigma} = \mathbf{\Lambda \Lambda}' + \psi. \tag{1.14}$$

In fact, this way of viewing the problem still survives as when (1.14) is referred to as a model.

The distinction between the old and new approaches lies in the fact that, whereas $\mathbf{\Sigma}$ is specific to factor analysis and has no obvious analogue in latent structure analysis, the probabilistic representation of (1.12) and (1.13) immediately generalizes as our formulation of the GLLVM shows.

1.5 Some Benefits of the New Approach

1.5.1 Factor scores

The so-called "factor scores problem" has a long and controversial history, which still has some life in it as Maraun (1996) and the ensuing discussion shows. The problem is how to locate a sample member in the **y**-space on the basis of its observed value of **x**. In the old approach to factor analysis, which treated (1.12) as a linear equation in mathematical (as opposed to random) variables, it was clear that there were insufficient equations p to determine the q y's because, altogether, there were $p+q$ unknowns (y's and e's). Hence the y's (factor scores) were said to be indeterminate.

Using the new approach, it is obvious that **y** is not uniquely determined by **x** but that knowledge of it is contained in the posterior distribution of **y** given **x**. From that distribution one can predict **y** using some measure of location of the posterior distribution, such as $E(\mathbf{y} \mid \mathbf{x})$. Oddly enough, this approach has always been used for the latent class model, where individuals are allocated to classes on the basis of the posterior probabilities of belonging to the various classes. The inconsistency of using one method for factor analysis and another for latent class analysis only becomes strikingly obvious when the two techniques are set within a common framework.

1.5.2 Reliability

The posterior distribution also tells something about the uncertainty attached to the factor scores. In practice, the dispersion of the posterior distribution can be disconcertingly large. This means that the factor scores are then poorly determined or, to use the technical term, unreliable. This poor determination of latent variables is a common phenomenon which manifests itself in other ways. For example, it has often been noted that latent class and latent trait models sometimes give equally good fits to the same data. A good example is given by Bartholomew (2000). A latent trait model, with one latent variable, was fitted to one of the classical data sets of educational testing - the Law School Admissions Test - with 5 items. A latent class model with two classes was also fitted to the same data and the results for the two models were hardly distinguishable. It thus appears that it is very difficult to distinguish empirically between a model in which the latent variable is distributed normally and one in which it consists of two probability masses.

A similar result has been demonstrated mathematically by Molenaar and von Eye (1994) for the factor analysis model and the latent profile model. The latter is one where the manifest variables are continuous but the latent variable categorical. They were able to show that, given any factor model, it was possible to find a latent profile model with the same covariance matrix, and conversely. Hence, whenever one model fits the data, the other will fit equally well as judged by the covariances. Once again, therefore, the latent distribution is poorly determined. These conclusions have important implications for linear structural relations models which seek to explore the relationships between latent variables. If very little can be said about the distribution of a latent variable, it is clear that the form of any relationship between them must also be very difficult to determine.

1.5.3 Variability

The calculation of standard errors of parameter estimates and measures of goodness of fit has been relatively neglected. In part this has been due to the heavy computations involved, even for finding asymptotic errors. However, it may also owe something to the strong disciplinary focus which was noted in the previous section. The criterion of "meaningfulness" has often been invoked as a justification for taking the fit of models at face value, even when the sample size is very small. The broad span of professional experience, which is brought to bear in making such judgements, is not to be disregarded, but it cannot replace an objective evaluation of the variability inherent in the method of fitting.

The treatment of latent variable models given in Bartholomew and Knott (1999) lays emphasis on the calculation of standard errors and goodness of fit. In addition to the standard asymptotic theory, which flows from the method of maximum likelihood, it is now feasible to use re-sampling methods like the bootstrap to study sampling variation. This is made the more necessary by the fact that asymptotic sampling theory is sometimes quite inadequate for sample sizes such as one finds in practice (e.g., de Menezes 1999). A further complication arises when a model with more than one latent variable is fitted. This arises because, in the GLLVM, orthogonal linear transformation of the y's leave the joint distribution of the x's unchanged. In factor analysis, this process is familiar as "rotation", but the same point applies to any member of the general family. It means, for example, that there is not one solution to the maximum likelihood equation but infinitely many, linked by linear transformations. Describing the sampling variability of a set of solutions, rather than a point within the set, is not straightforward. Further problems arise in testing goodness of fit. For example, with p binary variables, there are 2^p possible combinations which may occur. The obvious way of judging goodness of fit is to compare the observed and expected frequencies of these response patterns (or cell frequencies). However, if p is large, 2^p may be large compared with the sample size. In these circumstances many expected frequencies will be too small for the usual chi-squared tests to be valid. This sparsity, as it is called, requires new methods on which there is much current research.

1.6 Conclusion

Latent variables analysis is a powerful and useful tool which has languished too long in the shadowy regions on the borders of statistics. It is now taking its place in the main stream, stimulated in part by the recognition that it can be given a sound foundation within a traditional statistical framework. It can justly be claimed that the new approach clarifies, simplifies, and unifies the disparate developments spanning over a century.

References

Anderson, T. W. (1959). Some scaling models and estimation procedures in the latent class model. In U. Grenander (Ed.), *Probability and statistics* (pp. 9-38). New York: Wiley.

Arminger, G. & Küsters, U. (1988). Latent trait models with indicators of mixed measurement level. In R. Langeheine and J. Rost (Eds.), *Latent trait and latent class models*, New York: Plenum.

Bartholomew, D. J. (1980). Factor analysis of categorical data (with discussion). *Journal of the Royal Statistical Society*, 42, 293-321.

Bartholomew, D. J. (1984). The foundations of factor analysis. *Biometrika*, 71, 221-232.

Bartholomew, D. J. (2000). The measurement of standards. In *Educational standards* H. Goldstein and A. Heath (Eds.), for The British Academy, Oxford University Press.

Bartholomew, D. J. & Knott, M. (1999). *Latent variable models and factor analysis* (2nd ed). London, UK: Arnold.

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord and M. R. Novick (Eds), *Statistical theories of mental test scores*, Reading, MA: Addison-Wesley.

de Menezes, L. M. (1999). On fitting latent class models for binary data. *British Journal of Mathematical and Statistical Psychology*, 62, 149-168.

Fielding, A. (1977). Latent structure models. In C. A. O'Muircheartaigh and C.Payne (Eds.), *The analysis of survey data: Exploring data structures, Vol 1* (pp. 125-157). Chichester: Wiley.

Goodman, L. A. (1974). The analysis of systems of qualitative variables when some of the variables are unobservable. Part I-A modified latent structure approach. *American Journal of Sociology, 79*, 1179-1259.

Harvey, A. & Chung, C. -H. (2000). Estimating the underlying change in unemployment in the UK (with discussion). *Journal of the Royal Statistical Society,* 163, A, 303-339.

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

MacDonald, I. L. & Zucchini, W. (1997). *Hidden Markov and other models for discrete-valued time series.* London, UK: Chapman & Hall.

Maraun, M. D. (1996). Metaphor taken as math: indeterminacy in the factor analysis model. *Multivariate Behavioral Research*, 31, 517-538.

Molenaar, P. C. M. & von Eye, A. (1994). On the Arbitrary Nature of Latent Variables. In A. von Eye and C. Clogg (Eds.), *Latent variables analysis* (pp 226-242). Thousand Oaks, CA: Sage Publications.

Moustaki, I. (1999). A latent variable model for ordinal variables. *Applied Psychological Measurement*, 24, 211-223.

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psycholgy*, 16, 201-293.

Wiggins, L. M. (1973). *Panel analysis: Latent probability models for attitude and behavior processes.* Amsterdam: Elsevier.