

# Linear Versus Nonlinear Models in Item Response Theory

Roderick P. McDonald  
Macquarie University

A broad framework for examining the class of unidimensional and multidimensional models for item responses is provided by nonlinear factor analysis, with a classification of models as strictly linear, linear in their coefficients, or strictly non-

linear. These groups of models are compared and contrasted with respect to the associated problems of estimation, testing fit, and scoring an examinee. The invariance of item parameters is related to the congruence of common factors in linear theory.

The most widely known and commonly used models in latent trait theory (item response theory or IRT) are the almost interchangeable logistic and normal ogive models. In recent years it is these models, out of all the possible IRT models, that have been almost exclusively studied by theorists and applied in psychometric practice. Indeed, it may be that many users of psychometric techniques are unaware of the existence of other models within this theory. In much the same way, Lazarsfeld's (Lazarsfeld & Henry, 1968) *latent structure analysis*, an alternative label for latent trait theory, has tended to be identified by many writers and readers with just one of Lazarsfeld's models, the latent class model.

The object of this paper is to place the logistic/normal ogive models in a reasonably broad context of other models for the structural analysis of multivariate data (in the sense of McDonald, 1979a) and to use the comparisons and contrasts thus provided as aids in the evaluation of the item response models.

First, a general account will be given of the family of nonlinear common factor models, that is, of all models based on the principle of local independence. It will be pointed out that the logistic/normal ogive models are counterparts for dichotomous items and, indeed, nonlinear transformations of Spearman's general-factor model in common factor analysis. It therefore becomes reasonable to ask the same questions about the purposes, assumptions, and limitations of these models and to see what can be learned from a comparison of the methods for fitting them and for scoring the examinees. It will be shown that the principle of local independence provides the definition of *latent traits*, and that falsifiable models based on this principle are obtained by postulating a small enough number of latent traits (a low enough *dimensionality*) and a sufficiently restrictive function form for the relationship between the observed variables and the latent traits. A classification of nonlinear

---

APPLIED PSYCHOLOGICAL MEASUREMENT  
Vol. 6, No. 4, Fall 1982, pp. 379-396  
© Copyright 1982 Applied Psychological Measurement Inc.  
0416-6216/82/040379-18\$1.90

common factor models is based on the type of function linking the observed variables and the latent traits, which may be (1) linear in its coefficients and in the latent traits, (2) linear in its coefficients but nonlinear in the latent traits, (3) nonlinear in both the coefficients and the latent traits. It is pointed out that linear approximations may be employed to solve some problems of models, such as the logistic/normal ogive models, that are nonlinear both in the latent traits and in the item parameters.

Problems of estimation and of testing fit are examined on the basis of the comparison with linear models. This is not intended as a comprehensive review. It is pointed out that what is needed is both an overall statistical test of significance, enabling the rejection or nonrejection of the postulated model, and a criterion for regarding the model as a reasonable approximation to the data, even if the misfit is technically significant. Problems in estimation and testing of fit are then examined for the common factor model and for the logistic/normal ogive models in turn.

The question of "scoring an examinee" is then considered. A close parallel is exhibited between the maximum likelihood estimators of common factor scores in the Spearman model and the maximum likelihood estimators in the logistic model. Finally, the problem of matching results of factor analyses by congruence transformation is related to the questions of invariance of parameters in the complete latent space and of "scale-free" measurement in IRT.

#### The Family of Common Factor Models

The oldest and best understood of all the models for the structural analysis of multivariate data is Spearman's single common factor model. In Spearman's model, given  $n$  variables  $y_1, \dots, y_n$ , the existence of a single *common factor*, *latent trait*, or *latent variable*,  $\theta$ , is assumed such that the conditional mean of  $y_g$  for any fixed value  $\theta_a$  of  $\theta$ , is given by

$$\hat{y}_g = \mathcal{E}\{y_g | \theta = \theta_a\} = f_g \theta_a + m_g. \quad [1]$$

It is assumed also that for any fixed value  $\theta_a$  of  $\theta$ , the variables are mutually uncorrelated. That is,

$$\text{Cov}\{y_g, y_h | \theta = \theta_a\} = 0, \quad g \neq h = 1, \dots, n. \quad [2]$$

To put this another way, if the residual  $e_g$  of  $y_g$  about its regression on  $\theta$  is defined as  $e_g = y_g - f_g \theta_a - m_g$ , then the residuals are assumed to be mutually uncorrelated. By well-known theory, these assumptions imply the covariance structure

$$\text{Cov}\{y_g, y_h\} = f_g f_h, \quad g \neq h, \quad [3]$$

$$\text{Var}\{y_g\} = f_g^2 + \sigma_g^2, \quad [4]$$

where  $\sigma_g^2$  is the residual variance of  $y_g$ . It is reasonable to assert that a stronger assumption than Equation 2 is probably intended by all users of the model. That is, what is intended is the assumption that the variables are mutually statistically independent for any fixed value of  $\theta$ , and not merely uncorrelated. It would be surprising to find that they were uncorrelated but nonlinearly related, for example. We would then not be willing to accept the model. That is, in the Spearman model the principle of local independence in its general form given by Anderson (1959) and McDonald (1962) is assumed, but in applications the weak implication is tested that when the common factor is partialled out, the residual covariances are zero in the population from which the sample of examinees is drawn.

In the generally accepted interpretation of the Spearman model (and the more general common factor/latent trait models), the common factor is an abstractive attribute of the examinees which the

$n$  variables measure in common. Ideally, the attribute is defined beforehand and used to identify a behavior domain from which the variables to be analyzed are drawn. In practice, unfortunately, it is often named post facto as that which the variables that happen to have been analyzed appear to measure in common. The residual  $e_g$  of variable  $y_g$  is thought of as the sum of two components, one of which represents that which the variable measures specifically, as a stable aspect of examinee performance, while the other is an error of measurement, usually supposed unstable over replications of the measurement, if these were to be carried out (see McDonald, 1981, for a more extensive discussion). The importance of these rather obvious remarks in the present context lies in the fact that in the parallel theory for dichotomous items the distinction between stable item specifics and unstable item errors of measurement (which the items must have) is sometimes not explicitly made, and the status of an "ability" in IRT as no more and no less than a common factor can be consequently obscured.

**Generalizations of the Spearman Model**

*Multidimensional linear models.* There are two main directions in which the Spearman single common factor model can be generalized. One of these, usually identified with the name of Thurstone, is to go on from a unidimensional linear latent trait model to a multidimensional one, writing

$$\hat{y}_g = \mathcal{E} \left\{ y_g \mid \theta_1 = \theta_{1a}, \dots, \theta_k = \theta_{ka} \right\} = \sum_{p=1}^k f_{gp} \theta_{pa} + m_g \tag{5}$$

with

$$\text{Cov} \left\{ y_g, y_h \mid \theta_1 = \theta_{1a}, \dots, \theta_k = \theta_{ka} \right\} = 0, \quad g \neq h \tag{6}$$

and

$$e_g = y_g - \sum_{p=1}^k f_{gp} \theta_{pa} - m_g \tag{7}$$

in place of Equations 1, 2, and 3. The theory of the multiple factor model (the multiple latent linear model) now seems to be reasonably well understood. It is tempting to remark that it is possibly the only psychometric model that is well enough understood to be used in practice, yet its faults are so well known that some users avoid it in favor of other models whose corresponding faults have not yet been as thoroughly investigated.

Setting aside problems of rotation in the still popular exploratory use of this model, it can be said that, again, as in the Spearman case, the latent traits  $\theta_1, \dots, \theta_k$  are, in theory and in the practice of most investigators, treated as  $k$  abstractive attributes of the examinees which the  $n$  variables measure in common, and again, as before, the residual in Equation 7 consists of a stable specific component and an unstable error of measurement. There is nothing in the definition of the unidimensional or multidimensional latent linear model that directly precludes it from being applied to dichotomous items, and there is a considerable literature on the question of its applicability to such variables. (For some aspects of this see McDonald, 1965; McDonald & Ahlawat, 1974.)

*Unidimensional nonlinear models.* The second main direction in which the Spearman model can be generalized consists in the substitution of a nonlinear regression function for the linear regression function. In fact, a particularly strong motive for this arises in the case where the observed variables are dichotomous items. In such a case, if the item is given the binary score unity for whatever is deemed to be the positive response and zero for the corresponding negative response, it follows that the regression curve of the item score on the common factor becomes the conditional probability  $P\{y_g = 1 \mid \theta_1 = \theta_{1a}, \dots, \theta_k = \theta_{ka}\}$  of the positive response for given  $\theta_1, \dots, \theta_k$ . As a consequence, a linear

regression function (item characteristic curve; ICC) cannot then yield a rigorous mathematical treatment if the scale of  $\theta$  and the distribution of  $\theta$  are thought of as unbounded, since some values of  $\theta$  would then cause a function that represents a probability to take values greater than unity or less than zero, which is absurd. It is therefore desirable to choose item characteristic functions for dichotomous items in one or more dimensions which are bounded by zero and unity, and which have asymptotes. In the case of unidimensional models for dichotomous items measuring abilities, it is also appropriate to assume that the probability of passing each item is a monotone increasing function of the latent trait (scored in the direction of increasing ability). With suitably written items, appropriately scored, this can be true of other types of items, such as attitude and personality items, for example.

These considerations lead naturally to the choice of a cumulative distribution function to represent the ICC. Two cumulative distribution functions that have been used for this purpose are, of course, the normal ogive (Lawley, 1943; Lord, 1952) and the logistic function (Birnbaum, 1968). The normal ogive and logistic models are nonlinear counterparts and nonlinear transformations of the Spearman single factor model (see Equation 1), as can be seen on writing them respectively as

$$\hat{y}_g = \mathcal{E} \left\{ y_g | \theta = \theta_a \right\} = N \left( f_g \theta_a + m_g \right) \quad [8]$$

and

$$\hat{y}_g = \mathcal{E} \left\{ y_g | \theta = \theta_a \right\} = \Psi \left[ D \left( f_g \theta_a + m_g \right) \right] \quad [9]$$

where

$N(\cdot)$  is the normal cumulative distribution function;

$\Psi$  is the logistic cumulative distribution function; and

$D = 1.7$ , chosen to match the functions as well as possible.

Here, use will be made of

$$\hat{y}_g = P \left( f_g \theta_a + m_g \right), \quad [10]$$

where  $P(\cdot)$  is any cumulative density function, as a model that can be specialized to either of these. Because the models resulting from these transformations are nonlinear both in the latent trait  $\theta$  and the item parameters  $f_g, m_g$ , the combination of the assumed ICCs with the principle of local independence does not yield obvious counterparts of the mathematical treatment of the Spearman model.

For applications to multiple-choice cognitive items the assumption of a zero lower asymptote is unreasonable, since it implies that examinees can be found who have zero probability of obtaining the correct answer, even by "guessing." For such cases, a model has been introduced that has a third item parameter, the height of the lower asymptote, sometimes loosely described as a "guessing parameter." (A fourth parameter, the height of the upper asymptote, may also be considered. See McDonald, 1967a.) If the parameters  $f_g$  are set equal to a common value in Equation 9, one version of the Rasch model is obtained, in which there is just one parameter describing each item.

*A multidimensional nonlinear model.* In addition to multidimensional linear models and unidimensional nonlinear models, a very general common factor (or latent trait) model, incorporating both, may be written as

$$\hat{y}_g = \mathcal{E} \left\{ y_g | \theta_1 = \theta_{1a}, \dots, \theta_k = \theta_{ka} \right\} = \phi_g \left( \theta_{1a}, \dots, \theta_{ka} \right) \quad [11]$$

together with the assumption of the principle of local independence (McDonald, 1967a, 1967b). In the general formulations given by Anderson (1959) and McDonald (1962), this principle states that the members of a population of examinees (or other experimental units) can be characterized by latent traits  $\theta_1, \dots, \theta_k$  such that, in any subpopulation in which every examinee has the same set of values of

these, the measures of those examinees on variables  $y_1, \dots, y_n$  are mutually statistically independent. The statement of the principle itself constitutes the definition of latent traits. That is,  $\theta_1, \dots, \theta_k$  are latent traits if and only if they are quantities characterizing examinees such that, in a subpopulation in which the values of the traits are fixed, the scores of the examinees are mutually statistically independent. This is a purely mathematical definition of latent traits, although it does correspond reasonably well to the notion that latent traits are quantities that the items measure in common, since by definition they explain all statistical dependencies among them. The important points are (1) that words such as "latent," "unobservable," or "underlying" variables do not provide or even add to the definition of latent traits; (2) the principle of local independence is not as such a falsifiable assumption; and (3) the principle of local independence cannot be rejected without changing the concept denoted by the English language label "latent traits."

The truth of these assertions can be seen most clearly in the case of the classical common factor model. If  $k$  common factors do not explain the correlations of the tests, the hypothesis of  $k$  common factors is rejected in favor of a larger number that indeed serves to explain the correlations. The defining property of common factors, that they explain the correlations by the (weak) principle of local independence, is not rejected. (See McDonald, 1981, for a more extensive discussion.) As is there pointed out, the principle of local independence in the strong form just stated is not commonly used, but instead the weaker principle is substituted that in a subpopulation in which the common factors are fixed, the variables are uncorrelated. The position taken here is that this is merely a matter of the convenience of having a weaker, testable consequence of the strong principle, which is the form it is really desirable to assume, and that there is therefore no distinction to be made between common factors and latent traits. These terms are, indeed, being used interchangeably in this paper.

To obtain falsifiable models from Equation 11 a sufficiently small dimensionality,  $k$ , is specified in the model, and a sufficiently simple form for the regression functions  $\phi_g(\theta_{1g}, \dots, \theta_{kg})$  is chosen. The Spearman case (Equation 1), the multiple factor case (Equation 5), and the cumulative distribution function models (Equations 8 and 9) are the best-known special cases of Equation 11. McDonald (1967a, 1967b) discussed a class of regression functions that are linear in the coefficients (the item parameters) but nonlinear in the latent traits, being of the form

$$\hat{y}_g = \sum_{p=1}^k f_{gp} \phi_p(\theta_1, \dots, \theta_k) \quad [12]$$

He showed that these models yield the same covariance structure as the multiple linear latent trait model represented by Equation 5. (This equivalence gives rise to the possibility of "difficulty factors" in the application of the latent linear model to dichotomous items. See McDonald, 1965; McDonald & Ahlwat, 1974). The model of Equation 12 can be distinguished from the model of Equation 5 on the basis of an examination of the distribution of estimates of factor scores in the latent space, using the linear model. The theory provided two-stage least squares estimation for models in which the functions  $\phi_p(\cdot)$  are taken to be polynomials, for example, the single-factor cubic model

$$\hat{y}_g = f_{g3} \theta^3 + f_{g2} \theta^2 + f_{g1} \theta + m_g \quad [13]$$

One-stage least squares and maximum likelihood ratio estimators have recently been derived (McDonald, 1979b) and successfully programmed (Etezadi-Amoli, 1981; Etezadi-Amoli & McDonald, in prep.).

The point of these remarks is that it is useful to conceptualize the existence of a family of latent trait models, of which very few have been seriously investigated. The family includes (1) models that are strictly linear (linear in both their coefficients and the latent traits, such as Equations 1 and 5); (2)

models that are linear in their coefficients but not in the latent traits, such as Equations 12 and 13; and (3) models that are strictly nonlinear, such as the cumulative distribution function models of Equations 8 and 9. Generally, but not always, models of type 1 are suitable for measured variables ("quantitative" test scores) while models of type 3 are suitable for dichotomous and polychotomous items. A link between these is provided by models of type 2. McDonald (1967a) showed by harmonic analysis that the normal ogive model can be approximated as closely as desired by a polynomial series,

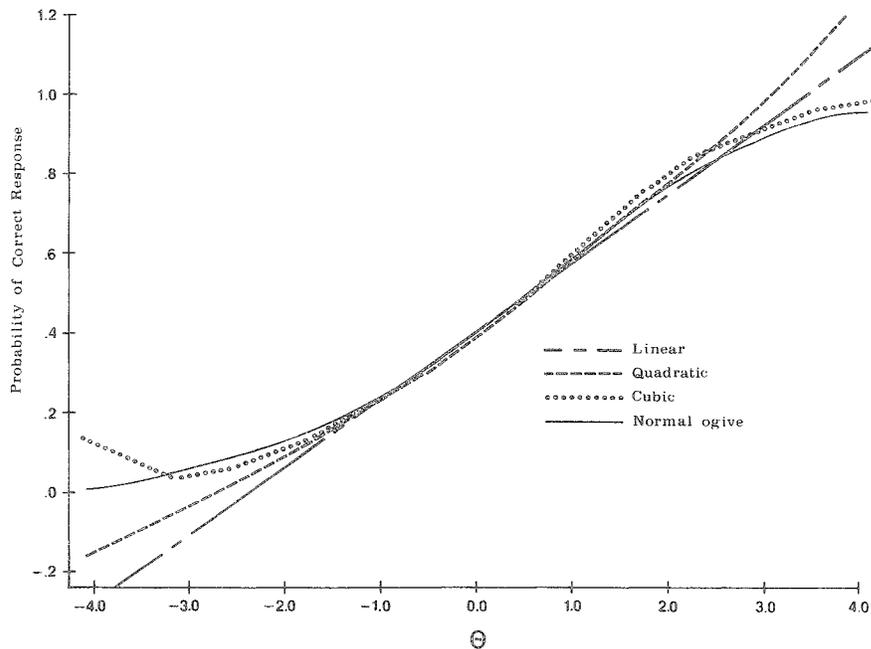
$$\hat{y}_g = \alpha_{g0} + \alpha_{g1}\theta + \alpha_{g2}\theta^2 + \dots \quad [14]$$

whose coefficients  $\alpha_{gp}$  are functions of the item parameters  $f_g, m_g$  in Equation 8, chosen so that a finite segment of its terms gives a least squares best fit to the normal ogive. Taking into account the distribution of the latent trait, the linear model actually does, in practice, provide a remarkably good approximation to the normal ogive, and terms beyond the cubic can generally be neglected. Figure 1 shows a normal ogive with the approximating linear and cubic ICCs. Weighted by a normal distribution of  $\theta$ , with mean zero and variance unity, the curves lie reasonably close together.

#### Estimation and the Testing of Fit

It is a remarkable achievement in psychological measurement if, by extremely careful selection of tests or items, an investigator succeeds in putting together a large battery of them which exhibit reasonable fit to the Spearman single factor model. It is true that the statistical theory given by Lawley

Figure 1  
A Normal Ogive with the Approximating  
Linear and Cubic Item Characteristic Curves



(1940) and implemented very effectively by Jöreskog (1967) and by others since, enables using Neyman-Pearson decision theory and rejecting or not rejecting the hypothesis of one common factor (or of any prescribed number of them) in favor of a larger number of factors (given multinormal data). Yet this hypothesis, like all restrictive statistical hypotheses, is surely false and will be proved so by a significant chi-square if only the sample size is made sufficiently large.

The more important question, which does not fit into the framework of statistical decision theory at all, is whether a data set is reasonably approximated by the single-factor model (or, more generally, by a prescribed number of factors). Even if, at some sample size, the misfit to the single-factor model is, technically, statistically significant at the chosen Type I error rate, the approximation to the model could be regarded as reasonable if the residual covariances are sufficiently small and distributed in such a way as not to allow the fitting of a satisfactory two-factor model.<sup>1</sup>

From this point of view the statistical test of significance has the limited but useful function of protecting against the overfactoring, especially in small samples, that seems a characteristic fault of cruder methods of factor analysis. That is, it provides some protection against the actual Type I error of fitting two factors when only one is statistically justified. (Notice that the number of latent traits is decided by how well the principle of local independence is satisfied in the sample when that number of latent traits is assumed. The hypothesized number of latent traits is not retained while the principle of local independence is rejected. In the context of IRT this fact is perhaps not so well understood.)

From the Hessian matrix of the likelihood function (the matrix of second derivatives of the function with respect to the parameters of the model), it is possible also to compute the standard errors of estimate of the factor loadings and residual variances and, though this is seldom done in practice, to compute the correlations between these errors. If the study is carefully designed to ensure that the model is well identified, these correlations should be low and the standard errors can be safely enough treated as though the errors were independent. If the correlations are high, the model is not well identified, and Heywood cases can occur. (See McDonald & Krane, 1977, 1979, for a technical account of this problem.) Parallel results for strictly nonlinear models, such as the cumulative distribution function models, have not yet been developed.

Almost all methods for common factor analysis are based upon a random regressors model. That is, it can be assumed that the latent traits  $\theta_1, \dots, \theta_k$ , as well as the variables  $y_1, \dots, y_n$ , are random variables. With this assumption the covariance structure in Equation 3 is fit to the sample covariance matrix (usually as a correlation matrix), measuring the misfit of the model by some function of the residual covariances, which in the Spearman case are given by

$$a_{gh} - \hat{f}_g \hat{f}_h, \quad g \neq h; \quad a_{gg} - \hat{f}_g^2 - \hat{\sigma}_g^2 \quad [15]$$

where  $a_{gh}$  is the sample covariance of  $y_g$  and  $y_h$ , and  $\hat{f}_g, \hat{\sigma}_g^2$  are estimates of  $f_g, \sigma_g^2$  in Equation 4.

Following work by Lawley (1942), McDonald (1979b) has shown that the linear common factor model (and certain nonlinear models that are linear in their coefficients) can be treated as a fixed regressors model, simultaneously estimating the test parameters  $f_g$  and  $\sigma_g^2$  and the examinee scores  $\theta_{ga}$ . Neyman and Scott (1948) have discussed statistical models in which there are both structural parameters, such as  $f_g, \sigma_g^2$  in this case, which are independent of sample size, and incidental parameters, such as  $\theta_{ga}$  in this case, whose number increases with sample size. They showed that in such cases maximum likelihood estimates of the structural parameters may not be consistent. McDonald's (1979b) results show that the estimates of the test parameters ( $f_g, m_g$ , and  $\sigma_g^2$  in the Spearman case) are

<sup>1</sup>Some computer programs for factor analysis unfortunately do not print out residual covariances. The use of these programs cannot be recommended, as they do not permit a decision whether the analysis bears any relationship to the data.

the same as in the random regressors model and are therefore consistent, while the estimates of  $\theta_{ga}$  are inconsistent, being given by the Kestelman equations for "indeterminate" factor scores (Guttman, 1955; Kestelman, 1952).

Fitting an IRT model such as the cumulative distribution function models to dichotomous items again presents a choice between a random regressors model, based on an  $(n + 1)$ -variable distribution of random  $y_1, \dots, y_n, \theta$ , and a fixed regressors model in which the  $\theta_a$  for  $N$  examinees are thought of as  $N$  fixed unknown values, not as random variables. The "randomness" in the latter model has a very curious conceptual basis, since replications of the study (impossible to realize in practice) must be imagined in which sets of  $N$  further examinees are found, having the same  $\theta_a$  values as the given set, and arrayed in the same order. In practice, the fixed regressors model may be used as a way to estimate the item parameters, with the examinee scores behaving as nuisance parameters which are estimated only because they have not been eliminated. In fact, it seems there should rarely be interest in knowing the latent trait values of examinees in the "calibration" sample to which the model is initially fit, but rather estimates of item parameters are required from the calibration sample in order to score future examinees (perhaps by using all  $n$  items, perhaps by using subsets as in tailored/adaptive testing and item banking procedures).

#### Fixed-Regressors Treatments

Methods for fitting the cumulative distribution function models of Equations 8 and 9 as fixed regressors models have been described by Kolakowski and Bock (1970) and by Lord (1974), and, in the special case commonly identified as the Rasch model in which the parameters  $f_a$  are taken to be a common constant, by Wright and Stone (1979), at least. Methods for fitting them as random regressors models have been described by Lord (1952), Bock and Lieberman (1970), Christofferson (1975), Muthén (1978), and McDonald (1980, in press).

The fixed regressors model does not seem to have provided a test of significance, such as a chi-square criterion with known degrees of freedom, against a less restrictive alternative, either the existence of two or more latent traits, or, just possibly, a less restrictive form of the item characteristic function. As in the case of the common factor model, such a test is of limited usefulness, but it would serve as protection against the Type I error of falsely rejecting a restrictive hypothesis and unjustifiably accepting a more general alternative.

More importantly, although this is not a difficult problem, existing computer programs for fitting the fixed regressors model do not yield a set of residual covariances whose magnitude (and possibly arrangement in the residual covariance matrix) would enable the formation of a rational judgment as to whether the data are well enough approximated by the simple unidimensional model. Given estimates of the item and examinee parameters, the residuals, and hence their covariances, may easily be computed. In the case of the Rasch model, Wright and Stone (1979) recommend the computation of certain statistics based on the residuals, but a mathematical account of the connection between these and the fit of the model does not seem to have been given, and it is difficult to see how the magnitude of the (standardized or unstandardized) residuals (as opposed to the covariances of residuals from distinct items) could have any direct bearing on the fit of the model. Thus, with existing methods, it seems impossible to make any decision based on the analysis as to whether or not the model adequately fits the data, or, indeed, shows even a crude resemblance to its structure.

By analogy with McDonald's (1979b) model, and perhaps invoking a principle of mathematical continuity from the approximating polynomial model to the cumulative distribution function models (see Figure 1), it might be conjectured that the estimates of the item parameters should prove consistent but the estimates of the examinee scores should not (for a given, fixed set of items). Indeed, these

may be indeterminate on curved subspaces of the latent space just as McDonald's (1979b) result shows that score estimates in the fixed regressors linear latent trait model are indeterminate on linear subspaces of the latent space, except in a behavior domain of infinitely many tests. In the case of the Rasch model, the estimates of item parameters and examinee parameters are easily shown to be jointly consistent, as the number of items as well as the number of examinees increases indefinitely.

### Random-Regressors Treatments

In the random regressors model the usual assumption is made that the latent trait has a normal distribution. This assumption yields, in effect, estimates of the item parameters based on the Spearman factor analysis of the matrix of item tetrachoric correlations (Lord, 1952), and, more generally, multiple factor analysis of such matrices (Christofferson, 1975; Muthén, 1978). McDonald (1980, in press) has used the harmonic analysis of the normal ogive given in McDonald (1967a) to fit the normal ogive model by substituting the best-fitting cubic model. A variant on this developed by Fraser (Program NOHARM; *Normal Ogive Harmonic Analysis Robust Method*)<sup>2</sup> has proved in monte carlo studies (McDonald & Fraser, in prep.) to give estimates closely agreeing with those from Muthén's FADIV program, as Hattie (1981) has demonstrated. (Both agree closely with results from the Spearman factor analysis of tetrachoric coefficients.) The NOHARM program has proved reasonably robust against violations of the normal distribution assumption. It is considerably faster than FADIV and seems suitable for very large data sets. It yields the usual residual covariances, as in the common factor model, since it is essentially representing the normal ogive model by an approximating model that is linear in its coefficients. Hence, it allows a judgment of the adequacy of the approximation of the model to the data. Since Muthén (1978) has employed generalized least squares estimation, he has obtained an asymptotic test of significance. Muthén's method can therefore be strongly recommended for small data sets.

The general point here is that the random regressors treatments of the model, in principle, yield an evaluation of fit and effectively contain within themselves a test of unidimensionality (see McDonald, 1981). Although this useful property is bought by what may seem the strong assumption that the latent trait has a normal distribution, it seems likely that these methods will be robust against violations of this assumption.

Little seems to have been done on the questions of identifiability and the degree of mutual correlation of errors of estimate of the parameters in these models. Attempts to use the cubic approximation of Figure 1 to estimate the three-parameter version of the normal ogive model, in which a lower asymptote above zero is introduced to allow for the effects of guessing in multiple-choice cognitive items, have not given good estimates. A possible explanation is that the errors of estimate are extremely highly correlated. The results from this procedure suggest that the three parameters in this model are jointly almost underidentified and not separately estimable. Research is needed to establish whether the result holds independently of the method of estimation.

### Estimating the Score of an Examinee

In both the linear latent trait models and the strictly nonlinear models the structural parameters of the model, treated as a random regressors model, may first be fit to a preferably large calibration sample. For any examinee in the population that was thus sampled, and which the model therefore

<sup>2</sup>Program NOHARM is available from C. Fraser, Centre for Behavioural Studies, University of New England, Armidale NSW 2351 Australia.

can be taken to describe, the value(s) of the examinee's latent trait(s) may then be estimated by the method of maximum likelihood.

In the unidimensional cumulative distribution function models the likelihood equation is given by

$$\sum_{g=1}^n \left[ \frac{\partial P}{\partial \theta} / P_g (1 - P_g) \right] [y_g - P_g] = 0 \quad [16]$$

where  $P_g = P(f_g \theta + m_g)$ . In the Spearman model for normally distributed variables the likelihood equation is easily shown to be

$$\sum_{g=1}^n \left[ f_g / \sigma_g^2 \right] [y_g - (f_g \theta + m_g)] = 0. \quad [17]$$

Both Equation 16 and Equation 17 may be written as

$$\sum_{g=1}^n \left[ \frac{\partial \hat{y}_g}{\partial \theta} / \text{Var}\{y_g | \theta\} \right] [y_g - \hat{y}_g] = 0, \quad [18]$$

or as

$$\sum_{g=1}^n w_g e_g = 0 \quad [19]$$

where

$$w_g = \frac{\partial \hat{y}_g}{\partial \theta} / \text{Var}\{y_g | \theta\}. \quad [20]$$

That is, in both cases  $\theta$  is chosen such that the weighted sum of the residuals is zero, where the weights consist of the slope of the regression of each  $y_g$  on  $\theta$  divided by the conditional variance.

In the linear model with normal variables, both terms in the weight are independent of  $\theta$ , one because the model is linear, the other because the residuals are homoscedastic. Consequently, Equation 17 may easily be solved in closed form to yield the familiar weighted least squares estimate due to Bartlett (see McDonald & Burr, 1967),

$$\tilde{\theta} = \left[ \begin{array}{cc} n & \sum_{g=1}^n \frac{f_g^2}{\sigma_g^2} \\ \sum_{g=1}^n \frac{f_g}{\sigma_g^2} & \sum_{g=1}^n \frac{f_g^2}{\sigma_g^2} \end{array} \right]^{-1} \sum_{g=1}^n \frac{f_g}{\sigma_g^2} (y_g - m_g). \quad [21]$$

In contrast, in strictly nonlinear models for dichotomous items, both terms in the weight are functions of  $\theta$ , and Equation 16 cannot be solved in closed form. In the case of the logistic model of Equation 9, it is easy to see that  $w_g = Df_g$  independent of  $\theta$ , so Equation 16 then becomes

$$\sum_{g=1}^n f_g \Psi \left[ D(f_g \theta + m_g) \right] = \sum_{g=1}^n f_g y_g, \quad [22]$$

which is a way of expressing the fact that the weighted item sum on the right of the equation is a sufficient statistic for  $\theta$ . In the special case of the Rasch model, in which the coefficients  $f_g$  have a common value, it follows that the unit-weighted sum  $\sum_{g=1}^n y_g$  is a sufficient statistic for  $\theta$ . However, it is primarily if what is desired is fitting the fixed regressors version of the model in the calibration sample and knowing the latent trait values of the examinees in that sample (at least for those of them who do not

obtain all zero or all unit scores on the items) that the one-parameter model has an advantage over those models which allow, realistically, for the fact that items measuring a given trait will not, in general, measure it equally well.

Factorial Congruence and  
the Invariance of Item Parameters

One of the main reasons for the current upsurge of interest in IRT is the promise it holds out of providing "scale-free" measurement through the invariance of item parameters under selection of the populations on which they are determined. Again, on this topic, the linear latent trait model, as treated in classical common factor theory, provides guidance in the clarification of concepts.

Item Parameter Invariance

Lord and Novick (1968, pp. 359-360) define a *complete latent space* as a set of  $k$  latent traits  $\theta_1, \dots, \theta_k$ , such that the conditional distribution of  $y_1, \dots, y_n$  for fixed  $\theta_1, \dots, \theta_k$  is invariant from one subpopulation to another in the entire population of interest, no matter how selected. In particular, the complete latent space is such that the item characteristic function, which is the mean of that conditional distribution, is invariant from one subpopulation to another. Hence, also the parameters of the item characteristic function are invariant, except for the effects of choosing the joint metric of  $\theta_1, \dots, \theta_k$ , that is, choosing origin and scale, and pairwise correlations between them. That is, quite generally, in the complete latent space the item parameters are determined except for the effect of an arbitrary translation and dilation of scale, and an oblique rotation of axes, as in the linear common factor model.

The selection of subpopulations from the entire population of interest can always be conceptualized as the choice of values of some additional variables  $s_1, \dots, s_m$ , which might merely be dummy variables labeling each subpopulation. By the two equivalent definitions of statistical independence given in texts on probability theory, it is known that

$$\begin{aligned}
 &P\{y_1, \dots, y_n \mid \theta_1, \dots, \theta_k; s_1, \dots, s_m\} \\
 &= P\{y_1, \dots, y_n \mid \theta_1, \dots, \theta_k\}
 \end{aligned}
 \tag{23}$$

if and only if

$$\begin{aligned}
 &P\{y_1, \dots, y_n; s_1, \dots, s_m \mid \theta_1, \dots, \theta_k\} \\
 &= P\{y_1, \dots, y_n \mid \theta_1, \dots, \theta_k\} P\{s_1, \dots, s_m \mid \theta_1, \dots, \theta_k\},
 \end{aligned}
 \tag{24}$$

where  $P\{\cdot \mid \cdot\}$  represents a conditional probability density. That is, the conditional distribution of  $y_1, \dots, y_n$  is invariant under selection on the basis of  $s_1, \dots, s_m$ , if and only if variables in the set  $\{y_1, \dots, y_n\}$  are statistically independent of variables in the set  $\{s_1, \dots, s_m\}$  for fixed  $\theta_1, \dots, \theta_k$ . If these equivalent conditions are not satisfied, in principle further latent traits may be found such that they are satisfied for some number  $K (< n)$  of traits.

To examine the implications of this, let it be supposed that several investigators carry out a factor analytic study using the same  $n$  tests  $y_1, \dots, y_n$ . Investigator A studies Population 1, to which the single-factor model applies, with factor loadings  $f_{11}, \dots, f_{1n}$ . That is, in this population

$$\hat{y}_g = f_{1g} \theta + m_{1g}, \quad g = 1, \dots, n. \quad [25]$$

Using a very large sample, A correctly fits the model. Investigator B studies Population 2, to which the single-factor model also applies, but with a distinct set of factor loadings  $f_{21}, \dots, f_{2n}$ , and of test means,  $m_{21}, \dots, m_{2n}$ . That is,

$$\hat{y}_g = f_{2g} \theta + m_{2g}, \quad g = 1, \dots, n. \quad [26]$$

Using a very large sample B correctly fits the model. Investigator C studies Population 1-2, which is a mixture in equal proportions of Population 1 and Population 2. (Unequal proportions would complicate the discussion without changing its essence.) Using a very large sample, C fits the single-factor model. D notices that the misfit in C's results is possibly serious, and re-analyzes them, preferring a two-factor model.

By the straightforward algebra of linear theory, it is easy to verify that the simplest linear latent trait model describing Population 1-2 is, indeed, except for a rotational indeterminacy,

$$\hat{y}_g = \frac{1}{\sqrt{2}} f_{1g} \theta_1 + \frac{1}{\sqrt{2}} f_{2g} \theta_2 + \frac{1}{2} (m_{1g} + m_{2g}). \quad [27]$$

The constants  $m_{1g}$ ,  $m_{2g}$ , and  $\frac{1}{2}(m_{1g} + m_{2g})$  are, in these models, equated to the unconditional means of  $y_1, \dots, y_n$ , in each population and thus are eliminated from the factor analysis. D should therefore obtain results that are in a sense consistent both with those of A and those of B, since D's matrix of factor loadings can be scaled and rotated (if D does not unthinkingly employ simple structure) so as to contain  $f_{11}, \dots, f_{1n}$  and  $f_{21}, \dots, f_{2n}$  in its two columns. C will obtain a set of single-factor loadings that resemble single least squares approximations to the pairs of numbers  $[f_{1g}, f_{2g}]$  that C really needs, and these will be inconsistent with the parameters obtained by A, B and D, while somewhat resembling all of them, thus perhaps giving evidence of invariance, perhaps giving evidence of the contrary.

Investigator M, studying A, B, C, and D's results, might ask for C's test scores (usually not published, unfortunately), compute estimates of  $\theta_1, \theta_2$  in Equation 27, and study their joint distribution. While Equation 27 gives the simplest linear item characteristic function describing the composite population, the density function of the latent traits associated with it is curiously degenerate. The probability of both  $\theta_1$  and  $\theta_2$  being nonzero is zero, and all the density is concentrated on two lines in the latent space, representing a mixture in two-space of the two one-dimensional models describing the subpopulations. Alternatively, the model for the composite population can be expressed as

$$\hat{y}_g = \left[ \theta_2^* f_{1g} + (1 - \theta_2^*) f_{2g} \right] \theta_1^* + \left[ \theta_2^* m_{1g} + (1 - \theta_2^*) m_{2g} \right] \quad [28]$$

where the latent trait  $\theta_1^*$  has any reasonable distribution, while the latent trait  $\theta_2^*$  is a discrete random variable taking values unity (for Population 1) or zero (for Population 2) and thereby reducing to the two models of Equations 25 and 26 for the subpopulations. Equation 28 is the factor interaction model, which is nonlinear in the latent traits  $\theta_1^*, \theta_2^*$  (since it contains their product), though linear in its coefficients (see McDonald, 1967c.) It is a moderator variable model, in which the discrete latent trait  $\theta_2^*$  acts to "moderate", i.e., change the regressions of  $y_1, \dots, y_n$  on  $\theta_1^*$ . The methods described by McDonald (1967c) would serve to discriminate the correct model Equation 28 from a model of the form Equation 27 in which the latent traits have an unrestricted distribution in the latent space.

The generalization of these remarks to the case of a cumulative distribution function model as in Equation 10 is straightforward, care being needed over one point. If

$$\hat{y}_g = P(f_{1g} \theta + m_{1g}) \quad [29]$$

in Population 1, and

$$\hat{y}_g = P(f_{2g}\theta + m_{2g}) \tag{30}$$

in Population 2, then necessarily in Population 1-2,

$$\hat{y}_g = P \left\{ \left[ \theta_2^* f_{1g} + (1 - \theta_2^*) f_{2g} \right] \theta_1^* + \left[ \theta_2^* f_{1g} + (1 - \theta_2^*) m_{2g} \right] \right\}, \tag{31}$$

the nonlinear counterpart of Equation 28.

However, the item characteristic function of Equation 28 does not have the “natural” counterpart of Equation 24 as its equivalent. That is, in general, the item characteristic function, as in Equation 27,

$$\hat{y}_g = P \left\{ \frac{1}{\sqrt{2}} f_{1g} \theta_1 + \frac{1}{\sqrt{2}} f_{2g} \theta_2 + \frac{1}{2} (m_{1g} + m_{2g}) \right\}, \tag{32}$$

with  $\theta_1 \neq 0$  if  $\theta_2 = 0$  and conversely is not equivalent to Equation 31 and does not contain Equations 29 and 30 as specializations in the subpopulations. This is because the location (“difficulty”) parameters  $m_{1g}$  or  $m_{2g}$  do not enter the model in a linear fashion and are not eliminated from it as item means, as in the simple linear model. Hence, the difficulty parameter  $1/2(m_{1g} + m_{2g})$  in Equation 32 is not related to  $m_{1g}$  or  $m_{2g}$  in Equations 29 and 30 by the addition or subtraction of a constant common to the items, as would at least be required for quasi-invariance of the item characteristic function.

Some general conclusions and conjectures may now be offered. By definition, as Lord and Novick (1968, p. 359) have stated, the complete latent space is a set of latent traits such that the conditional distribution of  $y_1, \dots, y_n$  is invariant over all subpopulations “of interest” in the population, and in particular it is such that the item characteristic functions and their parameters are invariant (on choosing a common  $k$ -dimensional metric for the latent traits). It is possible, however, that some readers have regarded this tautology as proving more than it does. It therefore seems desirable to point out that it is quite consistent with the following possibilities:

1. For a model of fixed dimensionality, and in particular for a unidimensional model such as is used in practice, the item characteristic function may not be invariant from one subpopulation to another (examples are Equations 25 versus 26 and 29 versus 30).
2. A unidimensional model may fit each subpopulation but not the entire population (as in both the linear case and the nonlinear case just considered).
3. The best fitting unidimensional model for the entire population may not yield parameters consistent with any of its subpopulations.
4. The item characteristic function for the entire population that is invariant over the subpopulations may be a moderator variable model representing the union of the distinct item characteristic functions that apply to distinct subpopulations (e.g., Equation 28, containing Equations 25 and 26, and Equation 31, containing Equations 29 and 30). The model more likely to be considered for such a population (e.g., Equation 32) need not yield invariance of item parameters, even if it is chosen to have the same dimensionality as the complete latent space.

At a more practical level, it must be noted that at the present there do not seem to be adequate methods for estimating both the item parameters and the examinee scores in models of the type

$$\hat{y}_g = P \left\{ f_{1g} \theta_1 + \dots + f_{kg} \theta_k + m_g \right\}, \tag{33}$$

in which the latent traits are combined according to the simple linear combination rule (see McDonald, 1979a), except possibly in the unidimensional case. Also, unless the investigator uses a quasi-factor-analytic method, yielding a residual covariance matrix, he/she may have no measure of misfit and may falsely accept a unidimensional model. It therefore seems, up to the present, that there is an inability (1) to determine the true dimensionality of the complete latent space, (2) to fit item characteristic functions in that space (say, of the type in Equation 33), and (3) to detect cases in which subpopulations need distinct item characteristic functions that should be combined according to a moderator variable model. All of these problems are solvable in principle if, using the methods of exploratory nonlinear factor analysis, linear approximations or models that are linear in their coefficients are used.

Broadly, then, traditional common factor theory indicates that in linear latent trait models, and hence in more general nonlinear models, for test scores or for item responses, behavior need not pattern the same way in distinct populations or in a composite of them, and that the parameters of the regression curves in such populations need not bear any simple relationship to each other.

Remarks so far in this section might be considered to support a pessimistic view about the use of latent trait theory, including (in particular) common factor analysis, to obtain invariant item parameters. This is not at all the intention. Equations 23 and 24 state, as already mentioned, that the conditional distributions of  $y_1, \dots, y_n$  given  $\theta_1, \dots, \theta_k$  (and, hence, the item characteristic functions) are invariant under selection on the basis of  $s_1, \dots, s_m$ , if and only if variables in the set  $\{y_1, \dots, y_n\}$  are statistically independent of each other and of variables in the set  $\{s_1, \dots, s_m\}$ , for fixed  $\theta_1, \dots, \theta_k$ . The stronger condition that  $s_1, \dots, s_m$  and  $y_1, \dots, y_n$  jointly satisfy the principle of local independence for fixed  $\theta_1, \dots, \theta_k$  is sufficient but not necessary for invariance. The weaker condition that  $y_1, \dots, y_n$  are mutually statistically independent for fixed  $\theta_1, \dots, \theta_k$  is necessary but not sufficient for invariance. In the special case of the linear common factor model, the necessary and sufficient condition is that  $y_1, \dots, y_n$  and  $s_1, \dots, s_m$  jointly satisfy the interbattery factor model (McDonald, 1969; Tucker, 1958) with the interbattery factors explaining the covariances of  $y_1, \dots, y_n$ , but not necessarily explaining the covariances of the selection variables  $s_1, \dots, s_m$ .

Instead of regarding it as axiomatic that item characteristic functions will be invariant (except for metric transformations), this might be recognized as a strong assumption, and yet it is one that should be commonly satisfied in practice. This supposes that care is taken not to select populations on the basis of the item scores, or any function of them, such as total test score, or of any tests, e.g., a "parallel form," that might by construction be related to the item specifics or item errors of measurement.

The conclusion is, quite simply, that the hypothesis of invariance, over populations, of the item characteristic functions in any given item response model, whether unidimensional or of sufficiently great dimensionality to fit the data, is a very natural and compelling hypothesis, though falsifiable. Before abandoning it, strong evidence of the alternative hypothesis, that the behaviors yield a different structure in different populations, may reasonably be demanded. Evidence such as that of Slinde and Linn (1979), that invariance seems to hold in a case that might be regarded as "typical," is of value. Counter-instances might also be actively sought by choosing sex-related or culture-related items that could be predicted to yield different structures across sexes or cultures. In general, it may reasonably be hoped that invariance of item characteristic functions (except for the choice of joint metric in  $k$ -dimensional latent space) will be obtained in item response models, as in traditional factor analysis.

#### Measuring Congruence

Under the hypothesis of invariance, the methods developed for achieving congruence of factor

analytic studies in distinct populations (e.g., Mulaik, 1972, chap. 14) apply to any latent trait model in which, as in Equation 33, the latent traits combine according to a linear combination rule, whether or not the model is strictly nonlinear. Consider the unidimensional case. Suppose the model of Equation 10 is applied to subpopulations, yielding

$$\hat{y}_{gs} = P \left\{ f_{gs} \theta_s + m_{gs} \right\}, \quad \begin{matrix} s = 1, \dots, t, \\ g = 1, \dots, n. \end{matrix} \quad [34]$$

If the item characteristic function is invariant except for metric, then each  $\theta_s$  is a scale transformation

$$\theta_s = k_s \theta + \mu_s \quad [35]$$

of the one latent trait  $\theta$ , with

$$\hat{y}_g = P \left\{ f_g \theta + m_g \right\} = P_g, \text{ say.} \quad [36]$$

Also, substitution of Equation 35 in Equation 36

$$\hat{y}_{gs} = P \left\{ k_s f_{gs} \theta + f_{gs} \mu_s + m_{gs} \right\} = P_{gs}, \text{ say,} \quad [37]$$

gives

$$k_s f_{gs} = f_g \quad [38]$$

and

$$f_{gs} \mu_s + m_{gs} = m_g, \quad \begin{matrix} s = 1, \dots, t \\ g = 1, \dots, n. \end{matrix} \quad [39]$$

(Note that Equation 38 represents Cattell's well-known principle of parallel proportional profiles.

Equations 38 and 39 are not consistent systems in general, when sample estimates of the item parameters are substituted, if only because of sampling errors. The best approximate solutions to them might then be chosen by the method of least squares, or by other devices to secure a match, such as have been used in common factor analysis under the label "Procrustes methods"—in some cases treating one set of parameters as a target and transforming the other sets to match it, or perhaps transforming all sets to a common best-fitting scale.

Apart from the question of transforming the parameters to achieve congruence, there is also the question of a suitable index to measure congruence. From the form of the relationship in Equation 38, it is clear that Burt's (1948) index of factorial congruence,

$$\tau_c = \frac{\sum_{g=1}^n f_{gs} f_{gv}}{\left( \begin{matrix} n & n \\ \sum_{g=1}^n f_{gs}^2 & \sum_{g=1}^n f_{gv}^2 \end{matrix} \right)^{1/2}} \quad [40]$$

would be a better measure of agreement between samples  $s$  and  $v$  than the commonly used correlation coefficient, since  $\tau_c$  equals unity if and only if the sets of numbers  $f_{gs}$  and  $f_{gv}$  are strictly proportional. From the form of the relationship in Equation 39 it would seem that no single index would serve to measure the agreement of  $m_{gs}$  and  $m_{gv}$  from samples  $s$  and  $v$ . If the slope parameters  $f_{gs}$ ,  $f_{gv}$  are considered to be equal to a common constant, as in the Rasch model, then the location parameters  $m_{gs}$ ,

$m_{gs}$  should differ by a common constant, and failure of invariance of these parameters could be measured by the variance of the difference  $m_{gs} - m_{gv}$ , perhaps converted to a dimensionless quantity such as

$$\tau_m = 1 - \frac{\text{Var} \{m_{gs} - m_{gv}\}}{\frac{1}{2} [\text{Var} \{m_{gs}\} + \text{Var} \{m_{gv}\}]} \quad [41]$$

The more general problem of matching location parameters in the presence of differences in slope parameters seems to require further research.

### Conclusions

The importance of item response theory based on latent traits defined by the principle of local independence has in the last few years received much wider recognition than it had during the main period of its development. This is because its practical utility as a basis for the optimal design of conventional tests and of tailored/adaptive tests has now become generally understood. There is a well-recognized need for good estimates of item parameters, and a hardly less well-recognized need for rational criteria for determining whether a unidimensional model gives a satisfactory account of the data.

From experience with the better understood linear common factor model, it has been suggested above that item calibration is better performed by the application of the random regressors model (Bock & Lieberman, 1970; Christofferson, 1975; Muthén, 1978; McDonald, 1980, in press) to a large calibration sample than by use of the fixed regressors model. The main objection to the use of the random regressors model is that it seems to require the strong assumption that the latent trait has a normal distribution. Research is needed on random regressors models with a more general distribution of the latent trait.

It is clear that the two-parameter model can be reasonably estimated in samples of commonly attainable size, possibly with nonzero values of the lower asymptote supposed known or separately determined, but there is room for doubt whether the three-parameter model can be reasonably estimated at commonly attained sample sizes. If, as seems likely, the parameters are not well identified in the latter model, the separate estimation of the lower asymptote based on the lowest fractile, of suitable width, with respect to total test score, may turn out to be the best option available, crude though it seems. Further work is needed on this question.

It cannot be emphasized too strongly that from estimates of the item parameters and, possibly, of the examinee parameters, the matrix of residual covariances of pairs of distinct items should be computed. A loss function such as the mean square of these covariances serves as a direct, rationally based criterion for the adequacy with which the model approximates the data, including the hypothesis of unidimensionality. Other criteria for checking unidimensionality tend to lack theoretical justification and do not perform well in monte carlo studies (Hattie, 1981). The disposition of large residual covariances in the matrix can be used to identify misfitting groups of items.

### References

- Anderson, T. W. Some scaling models and estimation procedures in the latent class model. In O. Grenander (Ed.), *Probability and statistics* (The Harold Cramer Volume). New York: Wiley, 1959.
- Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick, *Statistical theories of mental test scores*. Reading MA: Addison-Wesley, 1968.

- Bock, R. D., & Lieberman, M. Fitting a response model for  $m$  dichotomously scored items. *Psychometrika*, 1970, 35, 179-197.
- Burt, C. The factorial study of temperamental traits. *British Journal of Psychology, Statistical Section*, 1948, 1, 178-203.
- Christofferson, A. Factor analysis of dichotomized variables. *Psychometrika*, 1975, 40, 5-32.
- Etezadi-Amoli, J. *A general polynomial model for nonlinear factor analysis*. Unpublished doctoral dissertation. University of Toronto, 1981.
- Etezadi-Amoli, J., & McDonald, R. P. *A second generation nonlinear factor analysis*, in preparation.
- Guttman, L. The determinacy of factor score matrices with implications for five other basic problems of common-factor theory. *British Journal of Statistical Psychology*, 1955, 8, 65-81.
- Hattie, J. A. *Decision criteria for determining unidimensionality*. Unpublished doctoral dissertation, University of Toronto, 1981.
- Jöreskog, K. G. Some contributions to maximum likelihood factor analysis. *Psychometrika*, 1967, 32, 443-482.
- Kestelman, H. The fundamental equation of factor analysis. *British Journal of Psychology, Statistical Section*, 1952, 5, 1-6.
- Kolakowski, D., & Bock, R. D. *A Fortran-IV program for maximum likelihood item analysis and test scoring: Normal ogive model* (Research Report RM No. 12). University of Chicago, Department of Education, Statistical Laboratory, September 1970.
- Lawley, D. N. The estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society of Edinburgh*, 1940, 60, 64-82.
- Lawley, D. N. Further investigations in factor estimation. *Proceedings of the Royal Society of Edinburgh*, 1942, 61, 176-185.
- Lawley, D. N. On problems connected with item selection and test construction. *Proceedings of the Royal Society of Edinburgh*, 1943, 61, 273-287.
- Lazarsfeld, P. F., & Henry, N. W. *Latent structure analysis*. New York: Houghton-Mifflin, 1968.
- Lord, F. M. A theory of test scores. *Psychometric Monograph*, 1972, No. 7.
- Lord, F. M. Estimation of latent ability and item parameters when there are omitted responses. *Psychometrika*, 1974, 39, 247-264.
- Lord, F. M., & Novick, M. R. *Statistical theories of mental test scores*. Reading MA: Addison-Wesley, 1968.
- McDonald, R. P. A note on the derivation of the general latent class model. *Psychometrika*, 1962, 27, 203-206.
- McDonald, R. P. Difficulty factors and nonlinear factor analysis. *British Journal of Mathematical and Statistical Psychology*, 1965, 18, 11-23.
- McDonald, R. P. Nonlinear factor analysis. *Psychometric Monograph*, 1967, No. 15. (a)
- McDonald, R. P. Numerical methods for polynomial models in nonlinear factor analysis. *Psychometrika*, 1967, 32, 77-112. (b)
- McDonald, R. P. Factor interaction in nonlinear factor analysis. *British Journal of Mathematical and Statistical Psychology*, 1967, 20, 209-215. (c)
- McDonald, R. P. A generalized common factor analysis based on residual covariance matrices of prescribed structure. *British Journal of Mathematical and Statistical Psychology*, 1969, 22, 149-163.
- McDonald, R. P. The structural analysis of multivariate data: A sketch of a general theory. *Multivariate Behavioral Research*, 1979, 14, 21-38. (a)
- McDonald, R. P. The simultaneous estimation of factor loadings and scores. *British Journal of Mathematical and Statistical Psychology*, 1979, 32, 212-228. (b)
- McDonald, R. P. A simple comprehensive model for the analysis of covariance structures: Some remarks on applications. *British Journal of Mathematical and Statistical Psychology*, 1980, 33, 161-183.
- McDonald, R. P. The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology*, 1981, 34, 100-117.
- McDonald, R. P. Fitting latent trait models. In D. Spearritt (Ed.), *The improvement of measurement in education and psychology*. Australian Council for Educational Research Invitational Seminar, in press.
- McDonald, R. P., & Ahlawat, K. S. Difficulty factors in binary data. *British Journal of Mathematical and Statistical Psychology*, 1974, 27, 82-99.
- McDonald, R. P., & Burr, E. J. A comparison of four methods of constructing factor scores. *Psychometrika*, 1967, 32, 381-401.
- McDonald, R. P., & Fraser, C. Fitting the normal ogive model by harmonic analysis: A Monte Carlo study, in preparation.
- McDonald, R. P., & Krane, W. R. A note on identifiability and degrees of freedom in the asymptotic likelihood ratio test. *British Journal of Mathematical and Statistical Psychology*, 1977, 30, 198-203.
- McDonald, R. P., & Krane, W. R. A Monte Carlo study of local identifiability and degrees of freedom in the asymptotic likelihood ratio test. *British Journal of Mathematical and Statistical Psychology*, 1979, 32, 121-132.
- Mulaik, S. A. *The foundations of factor analysis*. New York: McGraw-Hill, 1972.

- Muthén, B. Contributions to factor analysis of dichotomous variables. *Psychometrika*, 1978, 43, 551-560.
- Neyman, J., & Scott, E. L. Consistent estimates based on partially consistent observations. *Econometrika*, 1948, 16, 1-32.
- Slinde, J. A., & Linn, R. L. The Rasch model, objective measurement, equating, and robustness. *Applied Psychological Measurement*, 1979, 3, 437-452.
- Tucker, L. R. An inter-battery method of factor analysis. *Psychometrika*, 1958, 23, 111-136.
- Wright, B. D., & Stone, M. H. *Best test design*. Chicago: Mesa Press, 1979.

**Author's Address**

Send requests for reprints or further information to Roderick P. McDonald, School of Education, Macquarie University, North Ryde, N.S.W. 2113, Australia.